



**HAL**  
open science

## Methods to detect spatial biases in tracking studies caused by differential representativeness of individuals, populations and time

Virginia Morera-Pujol, Paulo Catry, Maria Magalhães, Clara Péron, José Manuel Reyes-González, José Pedro Granadeiro, Teresa Militão, Maria Dias, Daniel Oro, Giacomo Dell’Omo, et al.

### ► To cite this version:

Virginia Morera-Pujol, Paulo Catry, Maria Magalhães, Clara Péron, José Manuel Reyes-González, et al.. Methods to detect spatial biases in tracking studies caused by differential representativeness of individuals, populations and time. *Diversity and Distributions*, 2023, 29 (1), pp.19-38. 10.1111/ddi.13642 . hal-03895401

**HAL Id: hal-03895401**

**<https://hal.science/hal-03895401v1>**

Submitted on 2 Jan 2025








**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Methods to detect spatial biases in tracking studies caused by differential representativeness of individuals, populations and time

Virginia Morera-Pujol<sup>1,2</sup>  | Paulo Catry<sup>3</sup> | Maria Magalhães<sup>4</sup> | Clara Péron<sup>5</sup>  | José Manuel Reyes-González<sup>1,2</sup> | José Pedro Granadeiro<sup>6</sup>  | Teresa Militão<sup>1,2</sup>  | Maria P. Dias<sup>7</sup> | Daniel Oro<sup>8,9</sup> | Giacomo Dell'Omo<sup>10</sup> | Martina Müller<sup>11</sup> | Vitor H. Paiva<sup>12</sup> | Benjamin Metzger<sup>13</sup> | Verónica Neves<sup>14</sup> | Joan Navarro<sup>15</sup> | Georgios Karris<sup>16</sup> | Stavros Xirouchakis<sup>17</sup>  | Jacopo G. Cecere<sup>18</sup> | Antonio Zamora-López<sup>19,20</sup> | Manuela G. Forero<sup>21</sup> | Ridha Ouni<sup>22</sup> | Mohamed Salah Romdhane<sup>23</sup> | Fernanda De Felipe<sup>1,2</sup> | Zuzana Zajková<sup>1,2,8</sup> | Marta Cruz-Flores<sup>1,2</sup>  | David Grémillet<sup>24,25</sup> | Jacob González-Solís<sup>1,2</sup> | Raúl Ramos<sup>1,2</sup> 

<sup>1</sup>Departament de Biologia Evolutiva, Ecologia i Ciències Ambientals, Facultat de Biologia, Universitat de Barcelona (UB), Barcelona, Spain

<sup>2</sup>Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona (UB), Barcelona, Spain

<sup>3</sup>MARE - Marine and Environmental Sciences Centre, ISPA-Instituto Universitário, Lisbon, Portugal

<sup>4</sup>Regional Secretariat for the Sea, Science and Technology, Regional Directorate for Sea Affairs (DRAM), Horta, Portugal

<sup>5</sup>Laboratoire de Biologie des Organismes et Ecosystèmes Aquatiques (UMR BOREA) MNHN, CNRS, IRD, SU, UCN, UA, Paris, France

<sup>6</sup>Departamento de Biologia Animal, CESAM, Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal

<sup>7</sup>Centre for Ecology, Evolution and Environmental Changes (cE3c) & CHANGE (Global Change and Sustainability Institute). Departamento de Biologia Animal, Faculdade de Ciências da Universidade de Lisboa, Lisboa, Portugal

<sup>8</sup>Centre d'Estudis Avançats de Blanes (CSIC), Blanes, Spain

<sup>9</sup>IMEDEA (CSIC-UIB), Esporles, Spain

<sup>10</sup>Ornis Italica, Rome, Italy

<sup>11</sup>Department of Natural Resources Science, University of Rhode Island, Kingston, Rhode Island, USA

<sup>12</sup>Department of Life Sciences, MARE - Marine and Environmental Sciences Centre/ARNET - Aquatic Research Network, University of Coimbra, Coimbra, Portugal

<sup>13</sup>BirdLife Malta, Ta' Xbiex, Malta

<sup>14</sup>Institute of Marine Sciences - Okeanos, University of the Azores, Horta, Portugal

<sup>15</sup>Institut de Ciències del Mar, CSIC, Barcelona, Spain

<sup>16</sup>Department of Environment, Faculty of Environment, Ionian University, Zakynthos, Greece

<sup>17</sup>Natural History Museum of Crete, University Campus (Knossos). School of Sciences & Engineering, University of Crete, Crete, Greece

<sup>18</sup>Area per l'Avifauna Migratrice (BIO-AVM), Istituto Superiore per la Protezione e la Ricerca Ambientale (ISPRA), Ozzano Emilia, Italy

<sup>19</sup>Southeast Naturalists Association (ANSE), Murcia, Spain

<sup>20</sup>Department of Zoology and Physical Anthropology, University of Murcia, Espinardo Campus, Murcia, Spain

<sup>21</sup>Estación Biológica de Doñana CSIC, Sevilla, Spain

<sup>22</sup>Faculté des Sciences Mathématiques, Physiques et Naturelles de Tunis (FST), Université de Tunis El Manar, Tunis, Tunisia

<sup>23</sup>Université de Carthage Institut National Agronomique de Tunisie, Tunis, Tunisia

<sup>24</sup>Centre d'Ecologie Fonctionnelle et Evolutive (CEFE), UMR 5175, CNRS, EPHE, IRD, Université La Rochelle, Montpellier, France

<sup>25</sup>Percy Fitzpatrick Institute of African Ornithology, NRF-DST Centre of Excellence, University of Cape Town, Rondebosch, South Africa

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Diversity and Distributions* published by John Wiley & Sons Ltd.

**Correspondence**

Virginia Morera-Pujol, School of Biology and Environmental Sciences, University College Dublin, Belfield, Dublin 4, Ireland.

Email: [morera.virginia@gmail.com](mailto:morera.virginia@gmail.com)

**Funding information**

European Commission, Grant/Award Number: LIFE10 NAT/MT090 and LIFE11 NAT/IT/000093; European Strategy Forum on Research Infrastructures, Grant/Award Number: ESFRI LifeWatch Project; Fundação para a Ciência e a Tecnologia, Grant/Award Number: IF/00502/2013/CP1186/CT0003, MARE-UID/MAR/04292/2019, PTDC/BIA-ANM/3743/2014, PTDC/MAR-PRO/0929/2014, UID/AMB/50017/2019 and UIDP/50017/2020+UIDB/50017/2020; H2020 European Research Council, Grant/Award Number: 618841; Hellenic Bird Ringing Centre; Ministerio de Ciencia e Innovación, Grant/Award Number: CGL2009-11278/BOS, CGL2013-42203-R, CGL2013-42585-P, CGL2016-78530-R and CGL2017-85210-P; MSDEC; Office Français de la Biodiversité, Grant/Award Number: PACOMM, Natura 2000 en mer; Organismo Autónomo Parques Nacionales, Grant/Award Number: 1248/2014

Editor: Nigel E Hussey

**Abstract**

**Aim:** Over the last decades, the study of movement through tracking data has grown exceeding the expectations of movement ecologists. This has posed new challenges, specifically when using individual tracking data to infer higher-level distributions (e.g. population and species). Sources of variability such as individual site fidelity (ISF), environmental stochasticity over time, and space-use variability across species ranges must be considered, and their effects identified and corrected, to produce accurate estimates of spatial distribution using tracking data.

**Innovation:** We developed R functions to detect the effect of these sources of variability in the distribution of animal groups when inferred from individual tracking data. These procedures can be adapted for their use in most tracking datasets and tracking techniques. We demonstrated our procedures with simulated datasets and showed their applicability on a real-world dataset containing 1346 year-round migratory trips from 805 individuals of three closely related seabird species breeding in 34 colonies in the Mediterranean Sea and the Atlantic Ocean, spanning 10 years. We detected an effect of ISF in one of the colonies, but no effect of the environmental stochasticity on the distribution of birds for any of the species. We also identified among-colony variability in nonbreeding space use for one species, with significant effects of population size and longitude.

**Main conclusions:** This work provides a useful, much-needed tool for researchers using animal tracking data to model species distributions or establish conservation measures. This methodology may be applied in studies using individual tracking data to accurately infer the distribution of a population or species and support the delineation of important areas for conservation based on tracking data. This step, designed to precede any analysis, has become increasingly relevant with the proliferation of studies using large tracking datasets that has accompanied the globalization process in science driving collaborations and tracking data sharing initiatives.

**KEYWORDS**

animal movement, environmental stochasticity, metapopulation study, site fidelity, species distribution

**1 | INTRODUCTION**

Although movement is a widespread characteristic in the animal kingdom, the study of movement ecology has been hampered by the scarcity of, and difficulty in obtaining, individual tracking data (Dingle, 1996). Since the 90s, the technological advances in animal tracking have facilitated the acquisition of large amounts of individual movement data with increasing precision (Wilmers et al., 2015). This boosted the development of long-term (Klaassen et al., 2014), metapopulation (Ferrerás, 2001) and even ecosystem-wide (Courbin et al., 2018) tracking studies. The miniaturization and increasing precision of tracking devices has overtaken the development of tools to prepare and analyse tracking data (Gupte et al., 2022) and has revealed sources of variability in the estimation of higher-level

ecological groups' spatial distributions (e.g. colony, population and species) from individual movement data that are yet to be properly addressed (Gutowsky et al., 2015).

Among the most important sources of variability when using methods to scale up from individual tracks to the space use of higher-level ecological groups, we can find the following: the presence of individual site fidelity (ISF; Spiegel et al., 2017), the variation in space use according to changing environmental conditions at a range of temporal scales (Paiva et al., 2013) and the heterogeneous use of space by distant populations (Frederiksen et al., 2012; Matthiopoulos, 2003).

Firstly, fidelity to a geographic area is a well-documented phenomenon, present in a wide range of animal species (Switzer, 1993), and can be related to breeding or foraging behaviour, and to social

interactions (Giuggioli & Bartumeus, 2012; Torney et al., 2018). An individual shows fidelity to a site when, based on previous experience, it returns to the same area where it had bred successfully or found a favourable environment for foraging (Schmidt, 2004). Although ISF usually refers to the individuals returning to the same breeding area (Schumm et al., 2021; Shimada et al., 2021), fidelity also occurs independently of the breeding period, to foraging areas or nonbreeding areas (Léandri-Breton et al., 2021; Robillard et al., 2018). Regardless, its effect must be considered when individual tracking data are used to infer the distribution of a population, as it can bias results towards areas preferred by the animals with better data coverage (Giuggioli & Bartumeus, 2012; Lascelles et al., 2012).

Secondly, temporal changes in environmental conditions also play an important role in populations' distributions. Despite being based on individual decisions, an animal's use of space is linked to the optimal use of resources, and therefore, affected by environmental dynamism (Wolf & Trillmich, 2007). For instance, herds of nomadic herbivores move following peaks in productivity of grasslands (Aikens et al., 2017), migratory bird species change their distributions along the year or specific route characteristics and nonbreeding areas of migratory birds are linked to changing environmental conditions (Dias et al., 2011).

Lastly, failing to consider the entire distribution of a species when it spans a heterogeneous environment leads to underestimating its space use. Individuals from different parts of this distribution will be exposed to different environmental conditions and ecological pressures and will thus exhibit a differential use of space (Jeltsch et al., 2013). In fact, it has been shown that the size of the home ranges calculated using tracking data can be correlated with the size of the study area (Nekolny et al., 2017). Thus, maximizing not only the number of animals tracked, but also the geographical area where they are tracked from, seems necessary to obtain precise space-use estimates (Börger et al., 2006). In addition, in species with spatially structured distributions, or in migratory species with nonoverlapping nonbreeding distributions among populations, tracking individuals from only a few breeding populations could lead to an erroneous estimation of their nonbreeding areas (Webster et al., 2002).

Behind all these sources of variability there are specific ecological drivers, mentioned in previous paragraphs which, although of great interest, do not represent the goal of the current study. Here, we focus on presenting a method to detect (1) how individuals having different spatial preferences and uneven sampling intensities can affect the higher-level distribution when pooling all individual data without weighting or correcting; (2) how different sampling efforts across time can bias the results, if spatial preferences of the study species vary along time; (3) how different sampling efforts across space can bias the results, if there is a spatial structure in the space use of the target species. The set of functions we developed here, once applied to the tracking data prior to their use in any analyses answering ecological questions, can help to detect the presence of these three sources of variability. Correcting for the detected sources of variability, and then using the corrected data in the

ensuing analyses, will guarantee more accurate distribution maps, protected area delimitations and space-use studies.

Marine top predators are a particularly useful group to study the effect of these sources of variability in tracking data since they show geographically widespread distributions and diverse movement and migratory patterns (Yurkowski et al., 2018). In addition, (1) they are often faithful to foraging and wintering areas (Chapman et al., 2009; Léandri-Breton et al., 2021; Wege et al., 2016); (2) their movements and distributions are affected by the dynamism of the marine environment (Albert et al., 2021; Patterson et al., 2021); and (3) they may display space-partitioning among populations (Wakefield et al., 2013). Lastly, they have, in recent years, been subject to a plethora of tracking studies, and although some start to acknowledge the need to account for these and other biases (Fauchald et al., 2021; O'Toole et al., 2020), a unified framework to detect these three main sources of bias would be beneficial particularly for multiyear and multicolony studies.

Among marine top predators, *Calonectris* shearwaters are medium-sized Procellariiforms that engage on year-round, long-distance and often trans-equatorial migrations (González-Solís et al., 2007) and show remarkable philopatry to the natal colony (Thibault, 1994). Three of the four extant species breed on the Mediterranean and Atlantic coasts, and their nonbreeding distributions are composed of discrete pelagic areas over the Atlantic and Indian Oceans (Dias et al., 2011; González-Solís et al., 2007). All these characteristics make them a suitable study group to investigate the effects of the sources of variability listed above.

In this work, we aim to provide a set of tools to understand the effects that (a) ISF, (b) environmental stochasticity over time and (c) the geographic extent of sampling effort, have on the distributions of mobile species tracked using animal-borne devices, regardless of tracking method, or habitat and characteristics of the study species, and to demonstrate their applicability to a real set of tracking data. To do so, we first tested the performance of the method with artificial datasets simulating different degrees of ISF, as well as different sampling regimes and efforts, and then demonstrated its applicability to real data by collating, for the first time, a dataset of 1346 year-round trips from 805 individuals of three *Calonectris* shearwater species breeding in 34 colonies. This constitutes a robust and diverse dataset and provides a relevant example to demonstrate the applicability of the method. Finally, we provide examples of the direct application of these methods by (1) calculating the nonbreeding distribution of each of the species and (2) evaluating the representativeness of each colony on the entire species based on their geographic location and centrality to the species breeding distribution.

## 2 | METHODS

We describe here the procedures developed to detect instances in which the methodology used to obtain population- or species-level space-use measures from individual tracking data can introduce biases. In particular, we aim to test the effects of ISF and temporal and

spatial variation in a tracking dataset, applicable to most sets of animal tracking data, as long as it can be split into discrete bouts such as foraging trips, migratory cycles, or even days, weeks or years (bouts, hereafter). We first demonstrate the efficiency of the procedures in simulated datasets and then illustrate the use of these methods using the nonbreeding distributions of *Calonectris* shearwaters tracked with Global Location Sensors (GLS, Wilson et al., 1992). In our worked example, the bouts were clearly defined as a migratory trip. In other types of data and species, bouts might be less evident and need to be carefully decided based on the research question and spatial ecology of the study species.

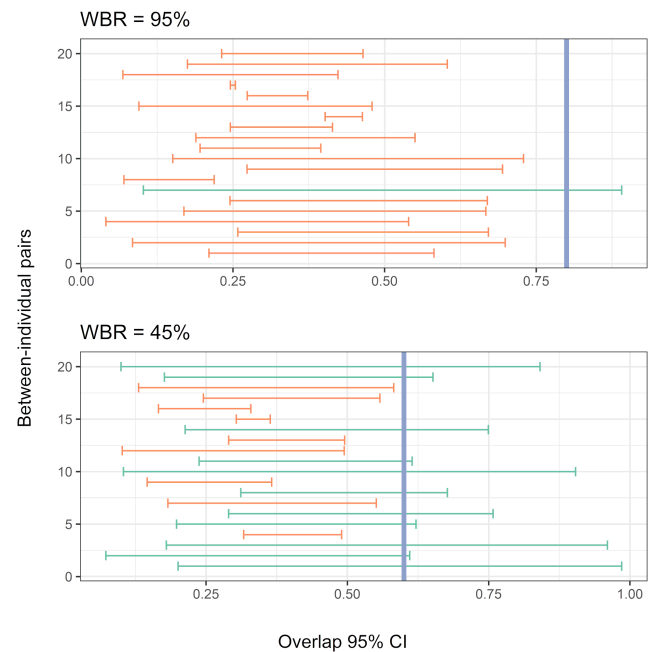
## 2.1 | Testing for ISF

To detect the bias caused by individuals with larger sample sizes preferring certain areas, we propose a method that consists of obtaining an estimate of space use for each bout and then comparing the similarities of bouts from the same individual to those of bouts from different individuals. If ISF exists within the sample being evaluated, space use of bouts from the same individual should be, on average, more similar among them than that of bouts of different individuals of the same sample. We estimate the space use of each individual with the autocorrelated kernel density estimate method (AKDE; Fleming & Calabrese, 2017) provided by the package “ctmm” (Fleming & Calabrese, 2021). This method models the movement of the animal, from the relocations obtained with the tracking device, as a continuous-time stochastic process, which separates the discrete observation process (a relocation every few minutes or hours depending on the device) from the continuous-time process which is the movement of the animal (the animal has continued to move between relocations, even if our observation process does not have information on where the animal has been in that time). From these models, an unbiased measure of space use can be estimated that considers the existing autocorrelation in tracking data with high spatiotemporal resolution (Fleming et al., 2015).

We developed a function, *indEffectTest*, in the R programming language, that computes, for every possible pair of bouts regardless of the individual that performs them, the similarity (i.e. spatial overlap) between their utilization distributions (UD). To assess this similarity, we used the Bhattacharyya affinity (BA), which is based on the comparison of two UD ( $p_1$  and  $p_2$ , evaluated at locations determined by  $x$  and  $y$ ) to their joint distribution, and ranges from 0 (completely disjunct UD) to 1 (identical UD):

$$BA = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \sqrt{p_1(x, y)p_2(x, y)} dx dy$$

In the BA version provided in the overlap function from the “ctmm” package, the calculation of the BA is corrected for small sample bias (Winner et al., 2018) which, in addition to the modelling of autocorrelation in the AKDE, provides a fairly unbiased measure of overlap, and a distribution to draw confidence intervals (CI) from,



**FIGURE 1** Example results of the individual site fidelity (ISF) function. On the top panel, a simulated dataset with high ISF, with a WBR of 95%, and on the bottom panel a simulated dataset without ISF, with a WBR of 45%. The vertical blue line represents the median within-individual overlap, while the horizontal lines represent the 95% CI of each between-individual overlap. The CI represented in orange does not contain the median within-individual overlap, while those represented in green do. High proportion of orange intervals indicates high levels of ISF.

thanks to the propagation of error from the uncertainties generated around the UD estimates. The mean and 95% CI of pairwise overlaps between all bouts are stored in an array of square matrices. The *indEffectTest* function provides the density plots of the within-individual and between-individual median overlaps to visualize differences in their distribution. In addition, it calculates the median of all within-individual median overlaps as a point estimate of within-individual overlap and provides the number of times in which this point estimate falls above each of the between-individual overlap CI (as a proportion of the total number of between-individual pairings). This value (within-between ratio, WBR hereafter) is a population-level measure of how similar the within-individual and between-individual overlaps are in that population's sample. Thus, if a large proportion of the between-individual overlap CI contains the within-individual overlap median, this would indicate that the within-individual and between-individual overlaps are similar, suggesting no effect of the ISF in the spatial distribution at a population level. Conversely, if the median of the within-individual overlaps falls above the majority of the between-individual CI, this would indicate higher overlaps within individuals than between individuals, a clear indication of the presence of ISF (Figure 1).

To demonstrate the function's ability to detect the effect of ISF, we performed practical examples by generating datasets that simulated 10 different degrees of fidelity (one dataset for each scenario).

To simulate ISF, we generated the bouts over a resistance surface with a “hole” of null resistance inside. We set high resistance outside the “hole” to simulate maximum levels of fidelity and to simulate sequentially decreasing levels of fidelity, we sequentially decreased the value of resistance.

We performed additional tests on simulated datasets to analyse the performance of the function under different sample sizes for each individual, as well as for different sampling rates. The code for simulating the trips and running the different tests can be found at the GitHub repository specified in Appendix S1.

## 2.2 | Testing for temporal variability

The aim of this test was to explore the existence of changes in distribution over time and not to understand why or how these changes occur. In the example with empirical data, we tested variability among years, but the user may select any other temporal unit. When testing differences in nonbreeding distributions of migratory species, for instance, or differences in foraging distributions of central place foragers such as colonial animals, the temporal unit is obvious: nonbreeding season and foraging trip, respectively. For species with nomadic movement or with less evident temporal units, the selection of this unit is not as straightforward. The user is advised to carefully consider the biological meaning of defining a unit, the aim of the analysis and the hypotheses being tested to select an appropriate time unit to split the data. In general, larger temporal units will allow larger within-group variabilities, and thus less possibility of finding significant differences when compared to between-group variability.

The existence of temporal variability can be detected with an approach that mimics the method used to test the effect of ISF (see Section 2.1): the aim is to check whether variability in space use within a unit of time is the same as between units. The function is the same as above, but instead of the individual, the temporal unit is the grouping variable. Thus, we ignore the individual identity for this and compare overlaps of trips within and between temporal units. The “between-group” values are overlaps of all pairs of trips from the different temporal units (regardless of the individual identity), while the “within-group” values are overlaps of all pairs of trips from the same temporal unit. Thus, a large WBR indicates that the overlap values within temporal units are similar to the overlap values between temporal units, which suggests a lack of effect of the time in the spatial distribution at a population level.

In the case of individuals tracked repeatedly each year, and then for consecutive years, the ISF effect and year effect can be confounded: in case there is an effect of the year (i.e. distributions shift from year to year), testing the effect of ISF using the dataset containing all years of tracking might cause an underestimation of within-individual overlap, thus underestimating in turn the WBR for the individual effect test. In this case, we suggest testing first for the temporal effect and, if the test shows no differences among years, calculating the effect of ISF for the entire dataset. In case a temporal

effect is detected, we recommend testing the effect of ISF year by year instead of for all the data pooled together.

## 2.3 | Testing for spatial variability

We propose a method to test the representativeness of each spatial unit with respect to the entire species (i.e. how well the distribution of a unit represents the distribution of the entire species) in spatially structured populations with large spatial ranges. In our example, since the study species is colonial, we used breeding colonies (from now on *colonies*) as spatial units. However, depending on the species studied, this choice might not be obvious. The user should decide then, based on the spatial ecology and distribution of the study species, and the spatial scope of their research question, if spatial variability is an issue and at which scale it should be tested.

This method works in two steps schematically described in Figure 2a–g. The first step is aimed at obtaining unbiased species distributions and can be accomplished with the *simulateDistribution* function (Appendix S2: Table S1). Empirical tracking data collected at different colonies have biases related to different sampling efforts and population sizes in all sampled colonies (i.e. the larger colonies will not necessarily be the ones with a larger sample). To correct these biases, we first simulate distributions for each colony by pooling all trips of each colony together and calculating colony-level space use (Figure 1a,b). In this instance, we are going to estimate space use using independent and identically distributed (IID) kernel density estimation (KDE) provided by the functions in package “adehabitatHR” (Calenge, 2006), since the AKDE method uses individual movement models that are not applicable to population-level space-use estimates. The user should be aware that, for tracking data with high temporal resolution and for small sample sizes, this method, which does not consider autocorrelation of tracking data, can underestimate space use (Fleming et al., 2015). The effects that the choice of smoothing parameter has on the resulting distribution should also be considered, a parameter that can be controlled by the scale argument in the *simulateDistribution* function (Lascelles et al., 2016). Colony-level space use is represented by the obtained KDE surface, and a number of locations that is proportional to the colony size can be generated with a distribution proportional to the density surface (Figure 2c). Once we have locations simulating each population's distribution, in a number proportional to each population's size, we pool the simulated locations from all populations to generate a simulated distribution that mimics that of the entire species, but unbiased by different sample sizes of each population, and faithful to the real population sizes (Figure 2c,d).

The drawback of this process is that it requires a good knowledge of the population distribution to generate a simulated location dataset against which researchers can compare their real location data. In our worked example, we have used our own data, since the dataset includes a good representation of the entire breeding distribution of the species. If such a comprehensive dataset is not available, users can skip the *simulateDistribution* step, and simply

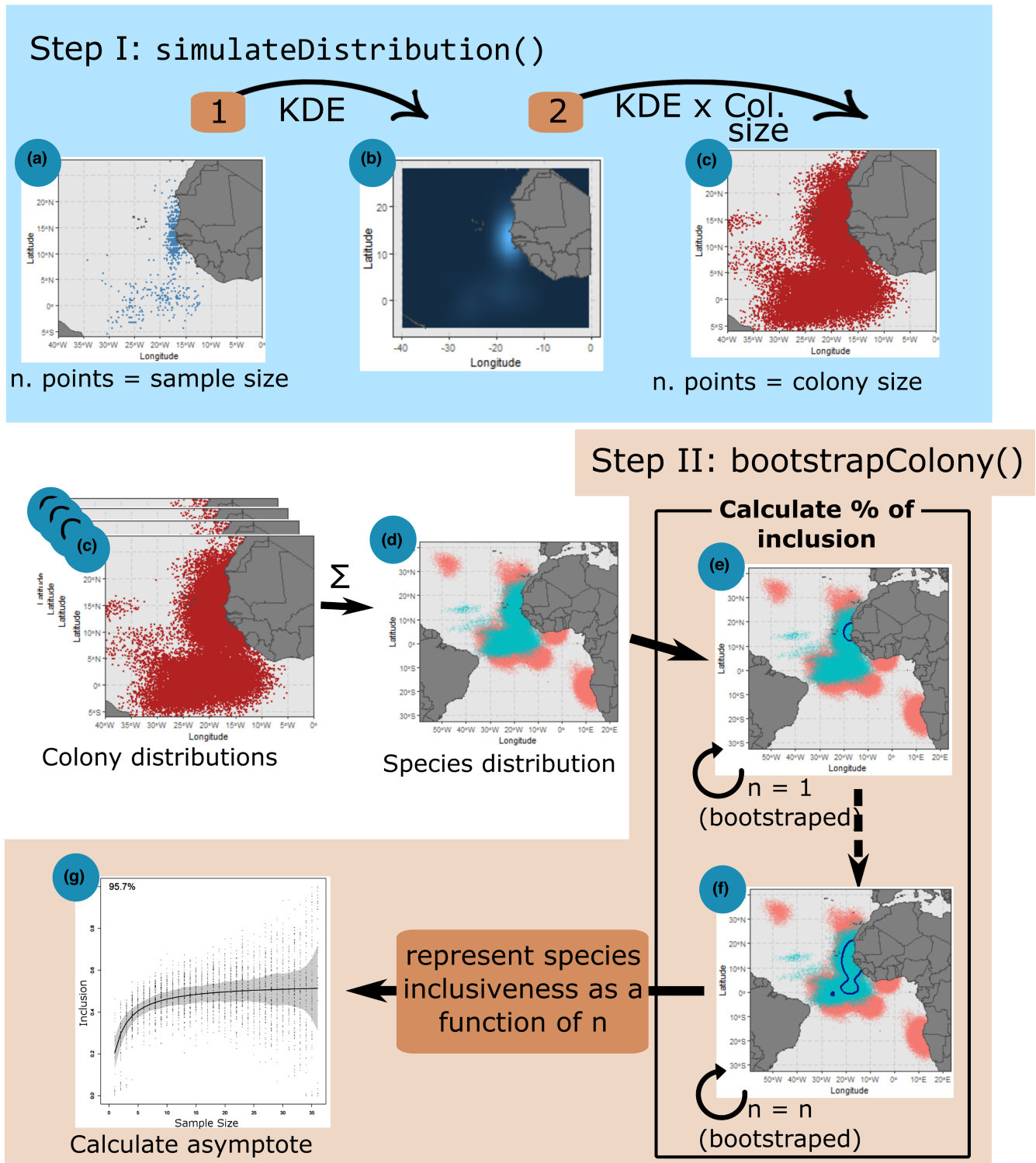


FIGURE 2 Schematic representation of the two-step process developed to test the effect the spatial variability in the use of space.

generate random points within a known distribution of the species (e.g. those provided by the BirdLife database for birds, or the International Union for Conservation of Nature, IUCN, website for species with a red list assessment), thus starting the process at step II in Figure 2. Although this second procedure simplifies the process, it has the drawback that the relative abundances within different regions of the distribution will not be correctly represented, and thus,

the representativeness of each subpopulation will not be accurately represented.

Once we have a set of spatial locations representing the distribution of the species, we calculate how well each colony represents it, based on inclusion with the *bootstrapColony* function (Figure 2e,f). For each colony, it iteratively subsamples individual tracks from the full colony sample and calculates a pooled UD

from which the desired outline polygon (usually a core area, 50% UD) is obtained. The inclusion value is calculated as the number of locations from the simulated species distribution (created on the previous step) included in this subsample core area. This inclusion value is estimated sequentially from  $n = 1$  trip until  $n = \text{total of trips}$  and for a user-defined number of iterations at each step. If the subsample is correctly representing the population distribution, the inclusion value should approximate the quantile for the UD (in this case 50%). As the number of randomly selected trips increases, the inclusion value increases as well, until it stabilizes at a point (an asymptote) when adding more trips to the dataset no longer increases the inclusion value. To calculate the asymptote, a nonlinear least squares model ( $nls$  function in R) is fitted, with the formula:

$$\text{Inclusion mean} \sim \frac{a \times \text{Sample size}}{1 + b \times \text{Sample size}}$$

where  $a$  and  $b$  are given start values of 1 and 0.1, respectively, and Inclusion mean is the mean inclusion value at each sample size. Through this model, we calculate the value for the Inclusion mean at which continuing to increase sample size will not increase inclusion further (i.e. the asymptote; Figure 2g). We regard this value as the maximum inclusion value (MIV) possible for that colony. Because MIV is calculated based on 50% UD, they range between 0 and 0.5 (i.e. even if a sample represents the entire species perfectly, its 50% UD will only include around 50% of the simulated locations). However, values higher than the UD percentage (in this case 50%) can occasionally occur, as the UD is calculated based on the locations of the colony, but the inclusion is calculated from the simulated locations of the entire species. In addition, users must exercise caution when comparing the representativity of colonies with vastly different sample sizes, as the biases introduced by the violation of the IID assumption and the bootstrapping of the KDE have a larger effect in small sample sizes. The function then multiplies the MIV obtained by 100, producing the value of “species inclusiveness,” defined as the percentage of points of the entire species that falls within the 50% UD of the sampled colony.

## 2.4 | Empirical application

To demonstrate the use of these functions, we put together a dataset containing year-round trips from Cory's (*Calonectris borealis*), Scopoli's (*C. diomedea*) and Cape Verde (*C. edwardsii*) shearwaters obtained between the summers of 2007 and 2016, from 34 breeding colonies (Table 1). Locations were obtained from GLS (Wilson et al., 1992), which register ambient light and derive longitude from the time of twilight and latitude from the length of the light period, obtaining low spatial accuracy positions with a temporal resolution of one or two positions per day (Halpin et al., 2021). The dataset contains individuals tracked for up to eight consecutive years, which allows us to also test the effect of ISF. Details of GLS deployment for each of the colonies can be found in Appendix S2: Table S2.

### 2.4.1 | Data preparation

The twilight events were calculated from the raw light measurements (obtained from the GLS) and visually inspected and adjusted when necessary. The locations were obtained using either Intiproc® (Migrate Technology, 2012) or Biotrack® (BiotrackLtd) software, or the “GeoLight” package in R (Lisovski & Hahn, 2012). We discarded position data from 20 days before and after each equinox, as latitudes cannot be correctly inferred from day length during these periods (Ekstrom, 2004), and applied a quadratic speed filter following McConnell et al. (1992) to remove other highly inaccurate locations. Phenological states (migrating, breeding, wintering or staging) were assigned using custom-made R routines and confirmed by visual inspection: the first two consecutive days of directional commuting movement were assigned as the beginning of migration, and the first two consecutive days without a fixed direction were considered as the beginning of a stationary period (either breeding, wintering or staging). Only the nonbreeding stationary periods (i.e. wintering and staging) were selected for the analyses.

### 2.4.2 | Individual site fidelity

To test the effect of ISF on nonbreeding distributions, we selected, for each species, all the colonies where individuals had been tracked more than once (Appendix S2: Table S3). We ran the *indEffectTest* function for each colony, using individual as the grouping variable, to obtain a WBR value based on debiased estimates of within- and between-individual overlaps. High WBR values are indicative of an effect of the ISF in the studied colony.

### 2.4.3 | Temporal differences

Since in our dataset each of the trips corresponded to a year, and to avoid confounding interannual differences with ISF, we used a subset of the original data containing only one trip per individual, randomly selected but ensuring a similar sample size for each year (through a stepwise custom-made R function that randomly selected trips per year, without selecting two trips from the same individual, but weighted the selection at every step to ensure all years had the same number of trips, whenever possible). With the resulting dataset (Appendix S2: Table S4), we ran the *indEffectTest* test for each colony, using year as the grouping variable.

### 2.4.4 | Spatial differences

Before proceeding with the analysis, we calculated the representativity of our sample at each of the colonies, using the *repAssess* function from the package “track2kba” (Beal et al., 2020; Beal, Oppel, et al., 2021). This function performs a representativity test to estimate how many trips would be necessary to accurately represent



TABLE 1 Number of trips obtained each year at the 13 colonies of Cory's shearwaters, 19 colonies of Scopoli's shearwaters and two colonies of Cape Verde shearwaters

Species	Breeding population	Breeding colony	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	Total	
Cory's shearwater	East Azores	Vila			15	9	30	8	3		4	3	72	
		Faial			8	5	10							23
		Graciosa			8	7	10							25
	Central Azores	Pico					15							15
		Corvo			6	2	2							10
	West Azores	Berlenga	10					14					13	37
		Sisargas							4	2			2	2
	Alboran Sea	Chafarinas			5							5		16
		Terreros										1	1	2
	Canary and Savage Islands	Selvagem Grande	30	28	40	50	43	28						219
		Montaña Clara						18	12	15		11		56
		Timanfaya												
		Veneguera		27	28	27	44	61	31	18	18	44	29	309
Palomas								4	3			2	16	
Balearic Islands	Chafarinas			3							4		16	
	Cala Morell										10	13	23	
	Na Foradada											6	6	
	Na Pobra											9	9	
French coast	Pantaleu				28	26	24	24	6	9	10		103	
	Porquerolles						7						7	
	Riou						27						27	
Corsica and Sardinia	Frioul						8						8	
	Giraglia						24						24	
	Lavezzi						18						18	
	Filfla							5	7				12	
Sicilian Channel	Malta					1				3		9	13	
	Gozo					1			1	2			4	
	Linosa				22	27	11	11	7	17	6		90	
Middle Adriatic	Zembra											13	26	
	Tremiti							7	1				8	
	Strofades				3	2			2	2			9	
	Paximada										8	8	16	

TABLE 1 (Continued)

Species	Breeding population	Breeding colony	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	Total
Cape Verde shearwater	Cabo Verde												
		Curral Velho	13	8	11	9	10	5	3	2	10		71
		Raso	13	9	5	2							29
		Total	66	72	129	164	221	268	76	75	144	131	1346

Note: The bottom row contains the total number of trips per year and the end column the total number of trips per colony.

the distribution of the entire colony, and how far away the current sample size is from that number. This approach is akin to that of species discovery curves and calculates a value of representativity that ranges between 0 and 100.

After calculating the colonies representativeness, we selected only those with a representativeness >80% (Table 2). From those colonies, we simulated the distribution of each species using the *simulatedDistribution* function, with a multiplication factor of 1 (i.e. one location at sea was generated for every individual in the colony). We obtained a set of simulated locations for each species (Table 2), which was used in the following step. Finally, we ran the *bootstrap-Colony* function (see Section 2.3 of this section and Appendix S2: Table S1), which calculated, for each colony, a value of species inclusiveness, a measure of how well space use by individuals of that colony represents space use by the entire species.

#### 2.4.5 | Examples of applicability of the developed methods

To further demonstrate the applicability of these methods, we used them in two different types of analyses, which exemplify possible uses. First, we calculated the nonbreeding distribution of each species, and to ensure these distributions were unbiased by different population and sample sizes from each colony, and due to different colony sizes, we used the following procedure: we first selected the colonies with a sample size larger than five individuals and a local representativeness >80%. Then, for each of those colonies, we selected the full dataset (if ISF has not been detected) or one track per individual (if ISF has been detected), and with the resulting dataset we followed the workflow of the package "track2kba" (Beal et al., 2020) to obtain rasterized abundance maps proportional to the colony sizes. We obtained 21 raster layers, seven for Cory's shearwater colonies, 12 for Scopoli's shearwater colonies and two for Cape Verde shearwater colonies. Finally, for each species, we added the colony-level raster layers to produce a more accurate, unbiased representation of the distribution of the species.

Second, we used generalized linear models to model the obtained colony inclusiveness value, that is, how well they represent species nonbreeding distribution in relation to several variables that could be of interest to disentangle ecological from methodological drivers of differences in inclusiveness: geographic location, centrality in the breeding distribution and colony and sample size. We ran these models for Cory's and Scopoli's shearwaters (data from Cape Verde were excluded since we had data from only two colonies). The sampling unit for these models were the sampled colonies, and to use as response variables, we calculated two values: (1) the distance of each colony to the geographic centroid of the species' spatial distribution and (2) the distance of each colony to the centre of the species' spatial distribution accounting for population sizes of each colony. For the distance to the geographic centroid (1), we calculated a KDE of all known breeding locations

**TABLE 2** Sample size, within-between range (WBR) for the individual site fidelity (ISF) (in %) for the colonies with repeated samples, WBR for the year effect (in %) for the colonies sampled more than 1 year, local representativeness (in %), number of simulated locations and species inclusiveness (in %) for each colony.

Species	Population	Colony	Sample size	WBR (ISF)	WBR (year)	Local representativeness	n simulated locations	Species inclusiveness
Cory's shearwater	East Azores	Vila	71	30.1	15.8	92.3	800	52.2
		Faial	22	19.2	13.2	89.2	11,426	37.0
	Central Azores	Graciosa	12	36.0	7.3	68.6		
		Pico	14	-	-	71.8		
	West Azores	Corvo	9	20.5	-	68.1		
		Berlenga	36	-	6.2	95.6	620	44.1
	Iberian coast	Sisargas	2	-	-			
		Chafarinas	10	11.3	8.2	80.6		
	Alboran Sea	Terreros	2	-				
		Selvagem Grande	211	31.9	6.9	98.6	59,080	47.8
		Montaña Clara	55	28.1	12.2	95.1	2000	40.4
		Timanfaya	14	-		84.1	700	27.4
		Veneguera	301	29.5	21.5	98.9	4000	35.1
	Canary and Savage Islands	Palomas	25	17.8	0.0	91.8	134	38.4
Chafarinas		5	9.7	2.5	56.1			
Balearic Islands	Cala Morell	22	28.2	0.0	91.4	2000	42.0	
	Na Foradada	5	-	-	38.5			
	Na Pobra	8	-	-	61.8			
French coast	Pantaleu	101	24.0	0.0	97.6	400	47.9	
	Porquerolles	6	-	-	55.2			
	Riou	26	-	-	95.0	560	28.7	
Corsica and Sardinia	Frioul	7	-	-	82.3	140	39.7	
	Giraglia	23	-	-	96.6	76	24.8	
	Lavezzi	17	-	-	93.7	600	22.6	
Sicilian Channel	Fiffla	9	53.1	0.0	74.2	3100	31.1	
	Malta	12	-	0.0	85.0			
	Gozo	3	-	-				
	Linosa	89	20.7	0.0	97.5	10,000	47.6	
Zembra	25	19.5	0.0	93.9	227,440	53.4		

TABLE 2 (Continued)

Species	Population	Colony	Sample size	WBR (ISF)	WBR (year)	Local representativeness	n simulated locations	Species inclusiveness
	Middle Adriatic	Tremiti	7	14.8	-	74.6		
	Ionian Sea	Strofades	8	0.0	0.0	89.1	11,000	17.6
	Aegean Sea	Paximada	15	17.8	0.0	96.7	1420	29.9
Cape Verde shearwater	East Cabo Verde	Curral Velho	70	18.3	0.0	99.4	66	49.4
	West Cabo Verde	Raso	28	8.7	0.0	97.9	13,000	50.8

Note: Colonies with a sample size <5 or local representativeness <80.0% (in italics) were considered nonrepresentative, and species inclusiveness was not calculated.

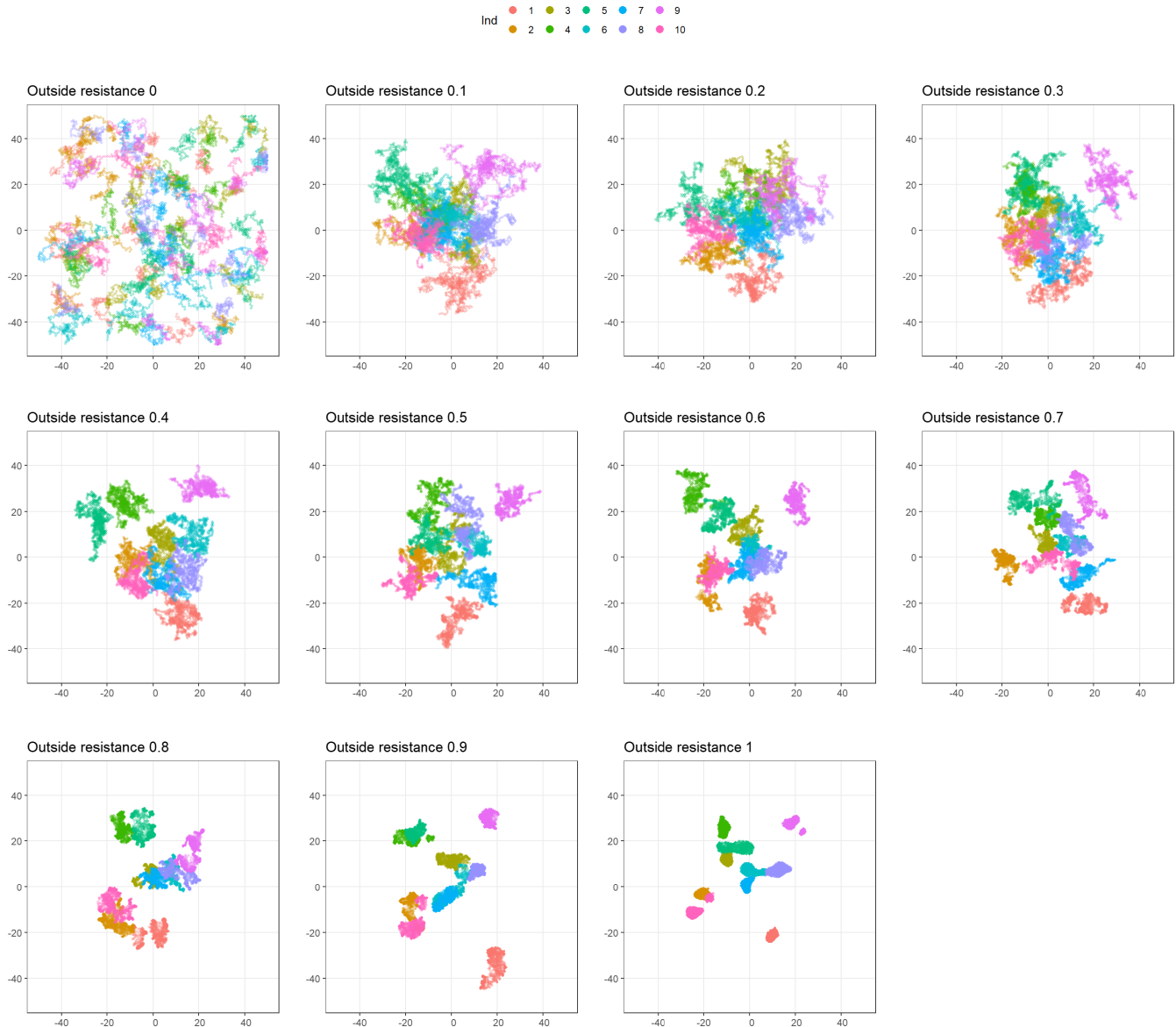
for each species (Appendix S2: Table S5), obtained the centroid of the 95% contour of that kernel and calculated the straight-line distance of each colony to that centroid. To calculate the centre of the species accounting for population sizes (2), we repeated the same process but weighing each colony by their population size and calculating the KDE of those weighed locations. The centroid of that distribution was taken as the species centre of mass (*species centre*, hereafter), and we calculated the straight-line distance of each colony to this species centre. These two distances were included as predictors in the model, as a measure of how centric each colony is within the species breeding range, both geographically and taking into account breeding population distribution. For Scopoli's shearwaters, we modelled the inclusiveness value against colony latitude and longitude (and their interaction), distance to the species centre, geographic centre, sample size and colony size. Since sample size and colony size had a few extreme values, these variables were log-transformed before being entered into the model. For Cory's shearwaters, we could not include all variables in the full model since the number of colonies was smaller than the number of predictors (we are only using the colonies with representativeness >80%), so we ran the full model without the interaction between latitude and longitude, and without the distance to the geographic centre (since it was strongly correlated with the distance to the species centre). For both species, we used Variance Inflated Factor (VIF; Dormann et al., 2013) values of the predictor variables and diagnostic plots to sequentially remove variables in order to obtain the most parsimonious model. To compensate for the small sample size in these models, particularly for Cory's shearwaters, we used a robust method to estimate standard errors, provided by the package "jtools" (Long, 2019), which is more stable to the effects of highly leveraged points than the ordinary least squares (OLS) method provided by the *lm* function (Cribari-Neto et al., 2007).

All analyses were performed in R (R Core Team, 2021) versions from 3.4.4 to 4.0.2. All the developed functions and their arguments are defined and explained in the supplementary material (Appendix S2: Table S1). The code for all the functions and performance tests can be found at [https://github.com/VirginiaMorera/Tracking\\_data\\_analysis](https://github.com/VirginiaMorera/Tracking_data_analysis)

### 3 | RESULTS

#### 3.1 | Results of ISF test on the simulated dataset

We ran the ISF test in data simulating a situation where the tracked animals were not colonial (e.g. a residential species where all trips from the same individual had a common origin but trips from different individuals had different origins). Within that paradigm, we tested the function in 10 simulated datasets of decreasing site fidelity (Figure 3). At null resistance (i.e. simulating the absence of ISF), WBR value was 4.8%, increasing up to 85.8% at maximum resistance (Table 3).



**FIGURE 3** Simulated datasets created to test the ISF function. The datasets consist of 10 trips of 100 positions each for each of the 10 animals simulated (in different colours). From left to right and top to bottom, ISF has been increased by increasing the resistance of the surface outside the *allowed* area (see Section 2) from 0 to 1.

## 3.2 | Results of the empirical application

### 3.2.1 | ISF effect

Values of WBR (with high values indicative of an effect of ISF) were on average higher for Cory's shearwater than for the other two species, ranging between 11.3% (Chafarinas) and 36.0% (La Graciosa) for Cory's shearwater, between 0.0% (Strofades) and 53.1% (Filfla) for Scopoli's shearwater and between 8.7% (Raso) and 18.3% (Curral Velho) for Cape Verde shearwater (Table 2).

### 3.2.2 | Interannual differences in distribution

Values of WBR were much lower in testing for interannual differences (evidencing no differences among years) than in the test for ISF

and again higher for Cory's shearwaters than for the other two species. Values ranged between 6.2% (Berlenga) and 21.5% (Veneguera) for Cory's shearwaters, and the only nonzero value for the other two species was for the Scopoli's shearwaters of Chafarinas (2.5%).

### 3.2.3 | Spatial differences in species representativeness

Before the analyses, we excluded three of the 34 sampled colonies, which had less than five trips. From the 31 remaining colonies, we selected one trip per individual for Flifla (Scopoli's shearwater) since a WBR higher than 50% suggested an effect of ISF. With the resulting dataset, we obtained values of local representativeness >80% in 22 of them (Table 2). With these, we proceeded to test their species inclusiveness.

We simulated nonbreeding distributions from each of those breeding colonies using the *simulateDistribution* function and obtained a dataset that contained more than 78,000 locations for Cory's shearwaters, more than 250,000 for Scopoli's shearwaters and more than 10,000 for Cape Verde shearwaters (Table 2).

Values of species inclusiveness for Cory's shearwater colonies ranged between 27.4% (Timanfaya) and 52.2% (Vila); for Scopoli's shearwater colonies from 17.6% (Strofades) to 53.4% (Zembra); and for Cape Verde shearwaters from 49.4% (Curral Velho) to 50.8% (Raso; Table 2).

TABLE 3 Results of the ISF test on simulated datasets

Within-between ratio (WBR, %)	
Resistance	Noncolonial
0.0	4.8
0.1	6.3
0.2	4.3
0.3	8.3
0.4	12.2
0.5	12.9
0.6	23.4
0.7	21.6
0.8	30.7
0.9	54.7
1.0	85.8

Note: For resistances outside the *allowed* polygon from 0 to 1, simulating ISF from low to high (see Section 2), we calculated how many within-individual overlap values fell below, inside and above the 95% CI of the between-individual overlap, for datasets simulating both a noncolonial and a colonial species.

TABLE 4 Main parameters from the models obtained from the selected linear models testing representativeness for Cory's and Scopoli's shearwaters.

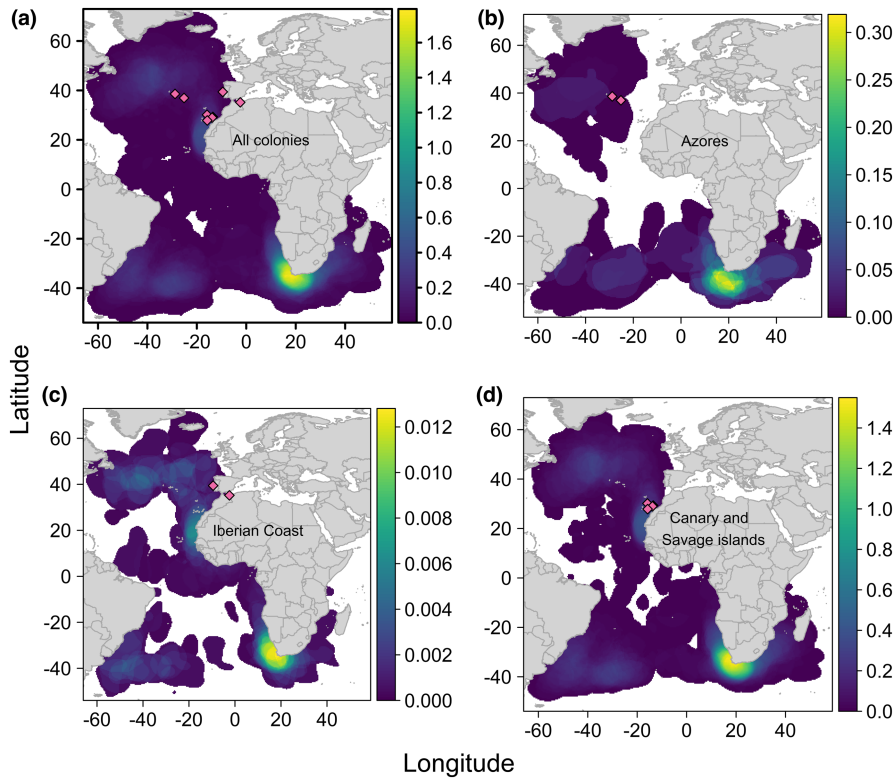
Species	Effect	Estimate	p Value	95% CI
Cory's shearwater	<b>Intercept</b>	<b>41.27</b>	<b>.006</b>	<b>27.63, 54.91</b>
	Latitude	5.82	.199	-7.41, 19.06
	Distance to population centre	2.67	.643	-18.59, 23.94
	Population size	-2.36	.669	-22.79, 18.08
	Sample size	7.47	.261	-13.20, 28.05
	Adjusted $R^2$	.57		
Scopoli's shearwater	<b>Intercept</b>	<b>36.02</b>	<b>.000</b>	<b>31.33, 40.70</b>
	<b>Longitude</b>	<b>-7.93</b>	<b>.039</b>	<b>-15.35, -0.52</b>
	Distance to population centre	3.43	.259	-3.08, 9.94
	<b>Population size</b>	<b>7.87</b>	<b>.004</b>	<b>3.34, 12.40</b>
	Sample size	4.13	.061	-0.23, 8.50
	Adjusted $R^2$	.51		

Note: Continuous predictors are mean-centred and scaled by 1 standard deviation. Effects with a  $p$  value  $<.05$  and/or a CI not overlapping 0 are considered significant and highlighted in bold. A parameter of the goodness of fit (adjusted  $R^2$ ) is also shown.

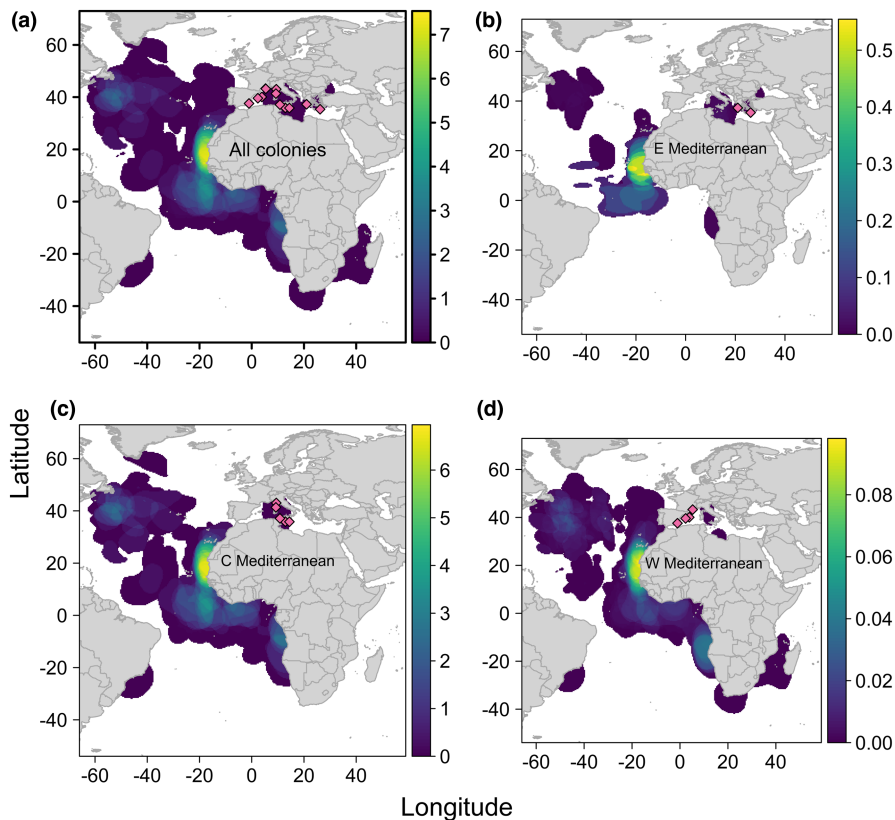
For Cory's shearwaters, after running the full model we found none of the predictors had a significant effect on the species inclusiveness value. We performed stepwise selection on the full model and found that none of the models improved the intercept-only model, with all effects being nonsignificant despite an  $R^2$  of .57 (Table 4). After running the full linear model for the Scopoli's shearwater data, we found high VIF values for distance to geographic centre and distance to population centre. We removed distance to geographic centre and latitude, until we obtained a model with low VIF values indicating no correlation among variables. From the final model, longitude had a negative effect on species inclusiveness and population size had a positive effect on it, while distance to the population centre and sample size had no significant effect (Table 4; Appendix S2: Figure S1).

### 3.2.4 | Nonbreeding distributions of three species of *Calonectris* shearwaters

After selecting one track per individual for the colony of Filfla, for which we found an effect of ISF, and since tests for temporal variability were not able to detect any effect in our data, we used data from all years together to plot the nonbreeding distribution of each species. We also plotted, for each species, the nonbreeding distribution of each of the main breeding areas, to compare them to the distribution of the entire species. For Cory's shearwaters globally, the main nonbreeding area was off the coast of South Africa, at the confluence between the Agulhas and Benguela currents and the Agulhas current retroflexion (Figure 4a; Appendix S2: Figure S2). In single-area distributions, however, we observed that the main nonbreeding area for animals breeding in the Azores was the Agulhas retroflexion, with secondary areas in the North and South central Atlantic (Figure 4b); for colonies in the Atlantic Iberian coast, the main nonbreeding area was the confluence between the Agulhas



**FIGURE 4** Nonbreeding distributions of Cory's shearwaters corrected for sample effort and weighted by population size. Using data from all representative colonies pooled together to represent the species' nonbreeding distribution (a), and using data only from the representative colonies in (b) Azores, (c) the Iberian coast and (d) Canary and Savage islands. Pink diamonds in (a) show the location of all sampled colonies. In (b–d), pink diamonds show representative colonies for each population. Scales show the number of individuals per  $0.1 \times 0.1^\circ$  cell.

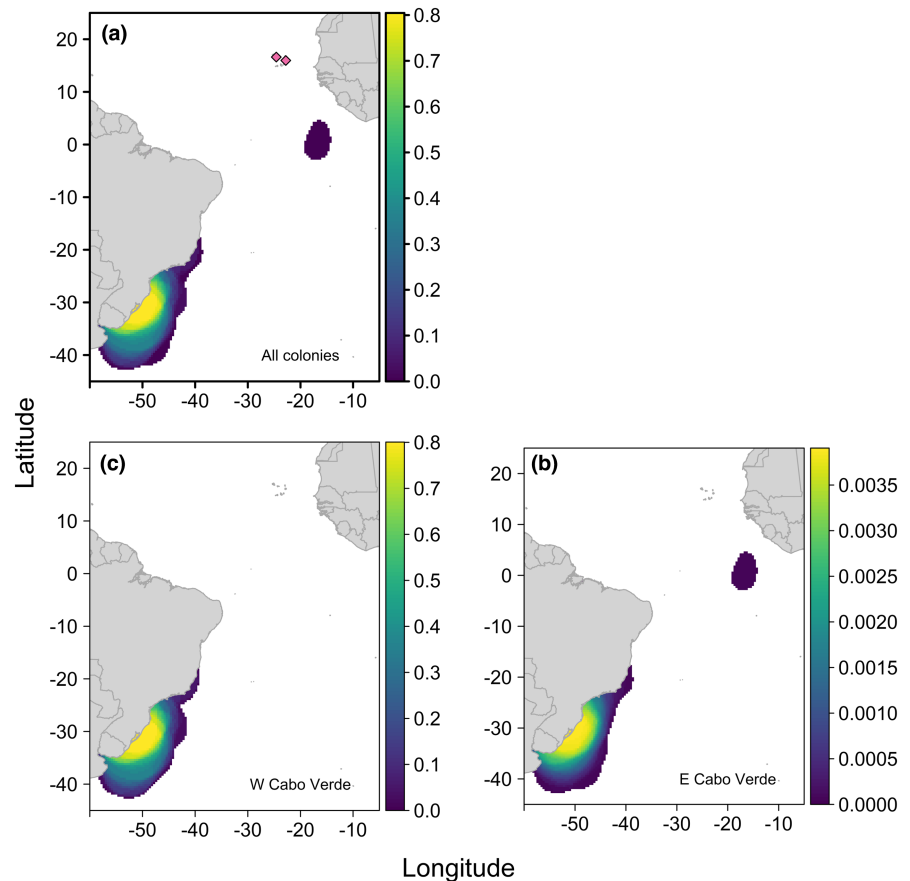


**FIGURE 5** Nonbreeding distributions of Scopoli's shearwaters corrected for sample effort and weighted by population size. Using data from all representative colonies pooled together to represent the species' nonbreeding distribution (a), and using data only from the representative colonies in (b) the Eastern Mediterranean, (c) Central Mediterranean and (d) the Western Mediterranean. Pink diamonds in (a) show the location of all sampled colonies. In (b–d), pink diamonds show representative colonies for each population. Scales show the number of individuals per  $0.1 \times 0.1^\circ$  cell.

and Benguela currents (Figure 4c); and for animals from the Canary and Savages Islands, the main nonbreeding area was the southern Benguela current, with the Canary current, off the coast of Western Sahara and Senegal, as a secondary area (Figure 4d).

For Scopoli's shearwaters, the main nonbreeding area was the Canary current (Figure 5a), but the distribution of nonbreeding birds changed if we considered only animals from the eastern, central or western colonies, with animals from the westernmost colonies using

**FIGURE 6** Nonbreeding distributions of Cape Verde shearwaters corrected for sample effort and weighted by population size. Using data from all representative colonies pooled together to represent the species' nonbreeding distribution (a), and using data only from the representative colonies in (b) eastern Cabo Verde and (c) western Cabo Verde. Pink diamonds in (a) show the location of all sampled colonies. In (b) and (c), pink diamonds show representative colonies for each population. Scales show the number of individuals per  $0.1 \times 0.1^\circ$  cell.



more southern nonbreeding areas, and animals from the eastern Mediterranean colonies spending the nonbreeding period mainly in the Northern hemisphere (Figure 5b–d). The only nonbreeding area of Cape Verde shearwaters was off the southern coast of Brazil and Uruguay (Figure 6).

## 4 | DISCUSSION

In this study, we provide a workflow to test the effect of three major sources of variability in spatial studies at the species level (ISF, temporal variability and spatial variability), potentially applicable to a wide variety of tracking datasets. In a context of increasingly sophisticated, miniaturized, and inexpensive animal tracking devices, studies using individual tracking data are proliferating, with ever larger extents both spatially and temporally. This emphasizes the need to account for the biases described here when studying space use at a population or species level, that is, before putting all tracks together towards generating a single measure of space use that does not consider potential differences in individual behaviour.

The ISF test is useful for datasets with uneven amounts of data for each individual. It detects instances when trips from the same individual are more similar than trips from different individuals, which causes the areas preferred by the more represented individual to be overrepresented in the distribution at a population level, and it can help designing future tracking studies to avoid this

effect. The solution will largely depend on the type of data, size of the dataset, the type of analysis to develop and the aims of each particular study. In general, researchers should aim to (a) subset the data so the sampling effort is evenly distributed among all individuals, for example, sequentially removing one random track from the repeatedly sampled individuals and testing for the effect of ISF again until the effect is no longer detected or (b) apply different weights to each track, with larger weights applied to trips of underrepresented individuals.

The test for interannual differences gives a measure of how spatial distribution varies among years at the population level. We would expect this test to find a significant effect of temporal variability in adaptable species that live in dynamic but predictable environments. On the contrary, species less able to track environmental conditions, or those that live in relatively constant environments, would have the same distributions year after year, resulting in nonsignificant effects of temporal variability on the distributions. Practically, what to do with a dataset that exhibits temporal variability will depend on the type of analysis planned. As an example, if the goal is to model species distributions based on environmental variables (a species distribution model), the effect of year should be considered by either modelling the tracks of each year separately and with the corresponding values for the environmental variables, or using a spatiotemporal model that will account for temporal as well as spatial autocorrelation (e.g. Martínez-Minaya et al., 2018). Conversely, if no temporal variability has been detected, a possible



course of action could be to average the values of the environmental variables during the whole tracking period, to capture general trends rather than detailed conditions, and use it in a species distribution model with all tracked positions to obtain general large-scale environmental preferences of the species.

The spatial variability test calculates how well a population represents the entire species distribution. In spatially structured populations, we would expect the inclusiveness value to be low for any of the sub-populations tested. Conversely, in unstructured populations, or in migratory species with very low migratory connectivity (i.e. that mix in common nonbreeding areas; Marra et al., 2019), we would expect the inclusiveness to be relatively high and similar for the majority of populations. Researchers aiming to represent the space use or environmental preferences of a species should take care in not using individual populations to represent the whole species when the representativeness is low. Instead, the different population sizes should be considered when combining data from all tracked populations, so the weight of each population in the species distribution is proportional to its size (Beal, Dias, et al., 2021; Beal, Oppel, et al., 2021; Carneiro et al., 2020; Davies et al., 2021).

In our simulated datasets, the *indEffectTest* was able to detect the effect of ISF as intended, as the WBR value increased with increasing simulated ISF. These datasets were simulated as noncolonial but central place forager individuals (i.e. each individual had a different starting position, but all trips from one individual started in the same location). These particular settings are just an example of the conditions under which tracking data may be collected, but users can use the code provided (see Section 2) and modify it to but simulate trips with characteristics similar to their real data (duration, number of positions and step length), in order to know the ability of the function to detect ISF in a dataset of those particular characteristics.

Applied to our GLS dataset of nonbreeding distributions of Cory's, Scopoli's and Cape Verde shearwaters, the proposed method detected an effect of the ISF in the distributions of two out of the 31 colonies tested. It did not detect an effect of temporal variability for any of the colonies, and there was spatial variability in species among the 21 colonies that were representative enough to enter the analysis. For most of the sampled colonies, the WBR value was high. For the colony of Filfla (Scopoli's shearwater), where the function detected a WBR of 53.1% (i.e. the median within-individual overlap is above the majority of between-individual overlap values) indicating an effect of ISF in the population-level space use, the correct course of action to take was to select one single trip per individual. For the rest of the colonies, no effect of the ISF was found, and we did not take any action on them.

The general lack of ISF found in our shearwater dataset agrees with published evidence suggesting a lack of individual consistency in the use of foraging areas during breeding and in several migratory and nonbreeding parameters for both Cory's and Scopoli's shearwaters (Courbin et al., 2018; Dias et al., 2011; Müller et al., 2014; Zango et al., 2019). At a population level, there are studies reporting

interannual consistency in nonbreeding distribution and migration phenology of seabirds (Bogdanova et al., 2014; Legrand et al., 2016), and our results agreed with these findings, showing that within-individual overlaps were mainly inside the 95% CI of between-individual overlaps for the three species.

In our example with these three phylogenetically close species that share nonbreeding areas, we found that the Cape Verde shearwater, the species with the most restricted breeding and nonbreeding distributions, had consistently higher inclusiveness values, while Cory's and Scopoli's shearwaters, with widespread breeding distributions, and several discrete nonbreeding areas had more diverse values of inclusiveness, as high as around 50% (maximum inclusiveness) but as low as 17.6% in the case of Strofades (one of the easternmost colonies of Scopoli's shearwaters). As expected, these highly spatially structured populations exhibited low values of species inclusiveness in each of their populations, that is, they poorly represented the distribution of the entire species, which has implications for study design and interpretation of results of single-population studies.

When trying to understand which factors affected the inclusiveness values, we found that for Cory's shearwaters, neither population or sample size, nor any geographical or population structure (represented in the distances to the geographic and population centre of the species) had an effect in such values. Since in six out of the 13 colonies we had sampled the sample size was not enough to achieve an 80% of representativeness, only seven colonies were used for the inclusiveness analysis. This small sample, and the fact that we did not have a good representation of the species to simulate the distribution, might have affected our ability to detect the effects of the geographic structure in the inclusiveness values. For Scopoli's shearwaters, however, we found that inclusiveness was positively affected by sample size and population size. These results indicate that larger colonies would better represent the entire distribution of the species, which particularly in the case of Scopoli's shearwaters seems logical since its largest colony, Zembra, represents about 80% of the entire species population, and, thus, getting a representative sample from that colony means obtaining a representative sample of the majority of the species. The model was not able to predict large differences in inclusiveness with increasing values of sample size. The relationship with longitude, although not significant, might suggest a west-to-east decline in inclusiveness values. Birds breeding in the easternmost colonies of the Mediterranean used the northernmost nonbreeding areas preferentially (Ramos, 2019), which made their use of the southernmost nonbreeding locations less likely. Thus, they were not good in representing the entire nonbreeding distribution of the species, although necessary for the complete knowledge of the wintering grounds of the species. As with the results detailed above, the representativeness of each colony or population can offer insights into their migratory connectivity: in species with low migratory connectivity, all populations will present higher values of representativeness, while in species with high connectivity, each population will have a low level of representativeness.

The findings obtained from our shearwater dataset are a clear example of the implications of this work for conservation studies. When the aim is to identify areas where the study species is more abundant, it would be more convenient to concentrate sampling efforts in colonies or breeding areas with larger populations. However, when the aim is to define the entire range of a given species, that is, any region in which an individual from the study species can be found, a sampling strategy including spread colonies or areas throughout the range of the species would be more appropriate. In addition, these methods could help decide when the data already collected are representative enough to extract conclusions relevant for conservation, thus avoiding the collection of more superfluous data and making the most of the data already collected. This work will serve as an example of the importance of taking these issues into consideration, as well as providing the tools to do so.

Collaborative datasets, metapopulation studies and their applications to predicting distributions of animals from unsampled areas (Péron et al., 2018; Wakefield et al., 2017) and to informing conservation and management policies (Hays et al., 2019) are becoming common-place in the current scientific context, although they do not always account for the sources of variability explored here (but see Beal, Dias, et al., 2021). The proposed method provides an objective protocol for the detection of three of the main sources of variability (ISF, interannual variability and spatial variability). This is a first step in the building of more accurate spatial and spatiotemporal models of movement and species distribution that can be used by movement ecologists working with a wide diversity of tracking data types.

## ACKNOWLEDGEMENTS

We would like to thank all students, volunteers and technicians who helped us at several stages of fieldwork. We are grateful to the following institutions for permission and logistical support in conducting fieldwork in the respective countries: Direção Nacional do Ambiente (Cabo Verde); Cabildo de Lanzarote, Cabildo de Gran Canaria, Timanfaya National Park, Gobierno de Canarias, Xunta de Galicia, Cabrera National Park, Govern Balear, Consejería de Agua, Agricultura y Medio Ambiente de la Región de Murcia, Consejería de Medio Ambiente y Ordenación Territorial de la Junta de Andalucía and Organismo Nacional de Parques Nacionales (Spain); and Direcção Regional do Ambiente dos Açores, Instituto das Florestas e da Conservação da Natureza (IFCN—Madeira), Instituto da Conservação da Natureza e Florestas (ICNF), Reserva Natural das Berlengas and Parque Natural da Graciosa (Portugal), Parc National de Port-Cros, Conservatoire d'Espaces Naturels de Provence-Alpes-Côte d'Azur, Réserve Naturelle Nationale de l'archipel de Riou, Parc Maritime des îles du Frioul, Réserve Naturelle des Bouches de Bonifacio and Direction Départementale de la Protection des Populations (France); Regione Siciliana, Assessorato Agricoltura e Foreste, the Wild Bird Regulation Unit and the Environment and Resources Authority (Malta), Management Agency of the National Marine Park of Zakynthos, Ministry of Environment (Greece). We also thank Steffen Oppel and Martin Beal for helpful advice regarding the methods section of the manuscript.

## FUNDING INFORMATION

We thank the following institutions for funding: EU H2020 programme through grant 634495; Seventh Framework Programme (Research Executive Agency) through Marie Curie Career Integration Grant 618841 (FP7-PEOPLE-2013-CIG); ESFRI LifeWatch Project; LIFE programme of the European Commission through projects LIFE10 NAT/MT090 and LIFE11 NAT/IT/000093; Ministerio de Ciencia e Innovación/Ministerio de Economía y Competitividad (Spain) through projects CGL2009-11278/BOS, CGL2013-42585-P, CGL2013-42203-R, CGL2016-78530-R and CGL2017-85210-P; Organismo Autónomo de Parques Nacionales (Spain) through project 1248/2014; Fundação para a Ciência e a Tecnologia (MCTES, Portugal) through projects MARE-UID/MAR/04292/2019; IF/00502/2013/CP1186/CT0003, PTDC/BIA-ANM/3743/2014, PTDC/MAR-PRO/0929/2014, UID/AMB/50017/2019 and UIDP/50017/2020+UIDB/50017/2020 (to CESAM); Office Français de la Biodiversité (France), through the Programme PACOMM, Natura2000 en mer; Hellenic Bird Ringing Centre; MSDEC (Malta). VMP was supported by pre-doctoral contract BES-2014-068025 of the Spanish Ministerio de Industria, Economía y Competitividad; MM by grant SFRH/BPD/47047/2008 from the Portuguese Foundation for Science and Technology; JMRG by Ph.D. grant AP2009-2163 from the Spanish Ministerio de Educación; GDO and MMü by Ornithologica and by the Regione Siciliana and Assessorato Risorse Agricole e Alimentari through a grant to the Ringing Unit of Palermo; VHP by grant SFRH/BPD/85024/2012 from the Portuguese Foundation for Science and Technology; VN by grant SFRH/BPD/88914/2012 from the Portuguese Foundation for Science and Technology; and JN by the Spanish National Programme Ramón y Cajal (RYC-2015-17809); GK and SX were partially funded by the Operational Program "Environment and Sustainable Development" (EPPERAA) of the National Strategic Reference Framework (NSRF) 2007-2013, co-financed by the ERDF and Greek EDP; FdF by a Ph.D. grant from the Coordination for the Improvement of Higher Education Personnel (CAPES—Brazilian government agency; Bex Process 1307/13-4); ZZ by a PhD grant from the University of Barcelona (APIF/2012); MCF by a PhD grant from the University of Barcelona; and RR by post-doctoral contracts of the PLEAMAR programme from MINECO and Fundación Biodiversidad (2017/2349), and Ministerio de Ciencia, innovación y Universidades (RYC-2017-22055). This publication is part of the project I+D+i/PID2020-117155GB-I00, funded by MCIN/AEI/10.13039/501100011033.

## CONFLICT OF INTEREST

The authors declare no conflicts of interest.

## DATA AVAILABILITY STATEMENT

The novel code developed for this manuscript, and to recreate the simulated datasets used to demonstrate the functions, can be found at [https://github.com/VirginiaMorera/Tracking\\_data\\_analysis](https://github.com/VirginiaMorera/Tracking_data_analysis). The animal tracking data used to demonstrate the real-life application of the functions developed are archived in the University of Barcelona data repository under DOI <https://doi.org/10.34810/data232>

## ORCID

Virginia Morera-Pujol  <https://orcid.org/0000-0001-6500-5548>

Clara Péron  <https://orcid.org/0000-0002-6529-4578>

José Pedro Granadeiro  <https://orcid.org/0000-0002-7207-3474>

Teresa Militão  <https://orcid.org/0000-0002-2862-1592>

Stavros Xirouchakis  <https://orcid.org/0000-0003-1979-4990>

Marta Cruz-Flores  <https://orcid.org/0000-0001-9905-4727>

Raül Ramos  <https://orcid.org/0000-0002-0551-8605>

## REFERENCES

- Aikens, E. O., Kauffman, M. J., Merkle, J. A., Dwinnell, S. P. H., Fralick, G. L., & Monteith, K. L. (2017). The greenscape shapes surfing of resource waves in a large migratory herbivore. *Ecology Letters*, 20(6), 741–750. <https://doi.org/10.1111/ele.12772>
- Albert, C., Bråthen, V. S., Descamps, S., Anker-Nilssen, T., Cherenkov, A., Christensen-Dalsgaard, S., Danielsen, J., Erikstad, K. E., Gavrilo, M., Hanssen, S. A., Helgason, H. H., Jónsson, J. E., Kolbeinsson, Y., Krasnov, Y., Langset, M., Lorentzen, E., Olsen, B., Reiertsen, T. K., Strøm, H., ... Moe, B. J. (2021). Inter-annual variation in winter distribution affects individual seabird contamination with mercury. *Marine Ecology Progress Series*, 676, 243–254.
- Beal, M., Dias, M. P., Phillips, R. A., Opper, S., Hazin, C., Pearmain, E. J., Adams, J., Anderson, D. J., Antolos, M., Arata, J. A., Arcos, J. M., JPY, A., Awkerman, J., Bell, E., Bell, M., Carey, M., Carle, R., Clay, T. A., Cleeland, J., ... Catry, P. (2021a). Global political responsibility for the conservation of albatrosses and large petrels. *Science Advances*, 7(10), eabd7225. <https://doi.org/10.1126/sciadv.abd7225>
- Beal, M., Opper, S., Handley, J., Pearmain, E. J., Morera-Pujol, V., Carneiro, A. P. B., Davies, T. E., Phillips, R. A., Taylor, P. R., Miller, M. G. R., Franco, A. M. A., Catry, I., Patrício, A. R., Regalla, A., Staniland, I., Boyd, C., Catry, P., & Dias, M. P. (2021b). track2KBA: An R package for identifying important sites for biodiversity from tracking data. *Methods in Ecology and Evolution*, 2021, 1–7. <https://doi.org/10.1111/2041-210X.13713>
- Beal, M., Opper, S., Handley, J., Pearmain, L., Morera-Pujol, V., Miller, M., Taylor, P., Lascelles, B., & Dias, M. P. (2020). BirdLifeInternational/track2kba: First release (version 0.5.0). Zenodo. <http://doi.org/10.5281/zenodo.3823902>. <https://github.com/BirdLifeInternational/track2kba>
- Bogdanova, M. I., Wanless, S., Harris, M. P., Lindström, J., Butler, A., Newell, M. A., Sato, K., Watanuki, Y., Parsons, M., & Daunt, F. (2014). Among-year and within population variation in foraging distribution of European shags *Phalacrocorax aristotelis* over two decades: Implications for marine spatial planning. *Biological Conservation*, 170, 292–299. <https://doi.org/10.1016/j.biocon.2013.12.025>
- Börger, L., Franconi, N., de Michele, G., Gantz, A., Meschi, F., Manica, A., Lovari, S., & Coulson, T. (2006). Effects of sampling regime on the mean and variance of home range size estimates. *Journal of Animal Ecology*, 75(6), 1393–1405. <https://doi.org/10.1111/j.1365-2656.2006.01164.x>
- Calenge, C. (2006). The package “adehabitat” for the R software: A tool for the analysis of space and habitat use by animals. *Ecological Modelling*, 197(3–4), 516–519. <https://doi.org/10.1016/j.ecolmod.2006.03.017>
- Carneiro, A. P. B., Pearmain, E. J., Opper, S., Clay, T. A., Phillips, R. A., Bonnet-Lebrun, A.-S., Wanless, R. M., Abraham, E., Richard, Y., Rice, J., Handley, J., Davies, T. E., Dille, B. J., Ryan, P. G., Small, C., Arata, J., Arnould, J. P. Y., Bell, E., Bugoni, L., ... Dias, M. P. (2020). A framework for mapping the distribution of seabirds by integrating tracking, demography and phenology. *Journal of Applied Ecology*, 57(3), 514–525. <https://doi.org/10.1111/1365-2664.13568>
- Chapman, D. D., Babcock, E. A., Gruber, S. H., Dibattista, J. D., Franks, B. R., Kessel, S. A., Guttridge, T., Pikitch, E. K., & Feldheim, K. A. (2009). Long-term natal site-fidelity by immature lemon sharks (*Negaprion brevirostris*) at a subtropical Island. *Molecular Ecology*, 18(16), 3500–3507. <https://doi.org/10.1111/j.1365-294X.2009.04289.x>
- Courbin, N., Besnard, A., Péron, C., Sarau, C., Fort, J., Perret, S., Tornos, J., & Grémillet, D. (2018). Short-term prey field lability constrains individual specialisation in resource selection and foraging site fidelity in a marine predator. *Ecology Letters*, 21(7), 1043–1054. <https://doi.org/10.1111/ele.12970>
- Cribari-Neto, F., Souza, T. C., & Vasconcellos, K. L. P. (2007). Inference under heteroskedasticity and leveraged data. *Communications in Statistics - Theory and Methods*, 36(10), 1877–1888. <https://doi.org/10.1080/03610920601126589>
- Davies, T. E., Carneiro, A. P. B., Tarzia, M., Wakefield, E., Hennicke, J. C., Frederiksen, M., Hansen, E. S., Campos, B., Hazin, C., Lascelles, B., Anker-Nilssen, T., Arnardóttir, H., Barrett, R. T., Biscoito, M., Bollache, L., Boulinier, T., Catry, P., Ceia, F. R., Chastel, O., ... Dias, M. P. (2021). Multispecies tracking reveals a major seabird hotspot in the North Atlantic. *Conservation Letters*, 14(5), e12824. <https://doi.org/10.1111/conl.12824>
- Dias, M. P., Granadeiro, J. P., Phillips, R. A., Alonso, H., & Catry, P. (2011). Breaking the routine: Individual Cory's shearwaters shift winter destinations between hemispheres and across ocean basins. *Proceedings of the Royal Society B*, 278(1713), 1786–1793. <https://doi.org/10.1098/rspb.2010.2114>
- Dingle, H. (1996). *Migration: The biology of life on the move*. Oxford University Press.
- Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., García Marquéz, J. R., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., & Lautenbach, S. (2012). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36(1), 27–46. <https://doi.org/10.1111/j.1600-0587.2012.07348.x>
- Ekstrom, P. A. (2004). An advance in geolocation by light. *Memoirs of National Institute of Polar Research*, 58, 210–226.
- Fauchald, P., Tarroux, A., Amélineau, F., Bråthen, V. S., Descamps, S., Ekker, M., Helgason, H. H., Johansen, M. K., Merkel, B., Moe, B., Åström, J., Anker-Nilssen, T., Bjørnstad, O., Chastel, O., Christensen-Dalsgaard, S., Danielsen, J., Daunt, F., Dehnhard, N., Erikstad, K. E., ... Strøm, H. (2021). Year-round distribution of Northeast Atlantic seabird populations: Applications for population management and marine spatial planning. *Marine Ecology Progress Series*, 676, 255–276. <https://doi.org/10.3354/meps13854>
- Ferreras, P. (2001). Landscape structure and asymmetrical inter-patch connectivity in a metapopulation of the endangered Iberian lynx. *Biological Conservation*, 100(1), 125–136. [https://doi.org/10.1016/S0006-3207\(00\)00213-5](https://doi.org/10.1016/S0006-3207(00)00213-5)
- Fleming, C. H., & Calabrese, J. M. (2017). A new kernel density estimator for accurate home-range and species-range area estimation. *Methods in Ecology and Evolution*, 8(5), 571–579. <https://doi.org/10.1111/2041-210X.12673>
- Fleming, C. H., & Calabrese, J. M. (2021). ctmm: Continuous time movement modelling. <https://cran.r-project.org/package=ctmm>
- Fleming, C. H., Fagan, W. F., Mueller, T., Olson, K. A., Leimgruber, P., & Calabrese, J. M. (2015). Rigorous home range estimation with movement data: A new autocorrelated kernel density estimator. *Ecology*, 96(5), 1182–1188. <https://doi.org/10.1890/14-2010.1>
- Frederiksen, M., Moe, B., Daunt, F., Phillips, R. A., Barrett, R. T., Bogdanova, M. I., Boulinier, T., Chardine, J. W., Chastel, O., Chivers, L. S., Christensen-Dalsgaard, S., Clément-Chastel, C., Colhoun, K., Freeman, R., Gaston, A. J., González-Solís, J., Goutte, A., Grémillet, D., Guilford, T., ... Anker-Nilssen, T. (2012). Multicolony tracking reveals the winter distribution of a pelagic seabird on an ocean

- basin scale. *Diversity and Distributions*, 18(6), 530–542. <https://doi.org/10.1111/j.1472-4642.2011.00864.x>
- Giuggioli, L., & Bartumeus, F. (2012). Linking animal movement to site fidelity. *Journal of Mathematical Biology*, 64(4), 647–656. <https://doi.org/10.1007/s00285-011-0431-7>
- González-Solís, J., Croxall, J. P., Oro, D., & Ruiz, X. (2007). Trans-equatorial migration and mixing in the wintering areas of a pelagic seabird. *Frontiers in Ecology and the Environment*, 5(6), 297–301. [https://doi.org/10.1890/1540-9295\(2007\)5\[297:TMAMIT\]2.0.CO;2](https://doi.org/10.1890/1540-9295(2007)5[297:TMAMIT]2.0.CO;2)
- Gupte, P. R., Beardsworth, C. E., Spiegel, O., Lourie, E., Toledo, S., Nathan, R., & Bijlvelde, A. I. (2022). A guide to pre-processing high-throughput animal tracking data. *Journal of Animal Ecology*, 91(2), 287–307. <https://doi.org/10.1111/1365-2656.13610>
- Gutow, S. E., Leonard, M. L., Conners, M. G., Shaffer, S. A., & Jonsen, I. D. (2015). Individual-level variation and higher-level interpretations of space use in wide-ranging species: An albatross case study of sampling effects. *Frontiers in Marine Science*, 2(November), 93. <https://doi.org/10.3389/fmars.2015.00093>
- Halpin, L. R., Ross, J. D., Ramos, R., Mott, R., Carlile, N., Golding, N., Reyes-González, J. M., Militão, T., De Felipe, F., Zajková, Z., Cruz-Flores, M., Saldanha, S., Morera-Pujol, V., Navarro-Herrero, L., Zango, L., González-Solís, J., & Clarke, R. H. (2021). Double-tagging scores of seabirds reveals that light-level geolocator accuracy is limited by species idiosyncrasies and equatorial solar profiles. *Methods in Ecology and Evolution*, 12, 2243–2255. <https://doi.org/10.1111/2041-210X.13698>
- Hays, G. C., Bailey, H., Bograd, S. J., Bowen, W. D., Campagna, C., Carmichael, R. H., Casale, P., Chiaradia, A., Costa, D. P., Cuevas, E., Nico de Bruyn, P. J., Dias, M. P., Duarte, C. M., Dunn, D. C., Dutton, P. H., Esteban, N., Friedlaender, A., Goetz, K. T., Godley, B. J., ... Sequeira, A. M. M. (2019). Translating marine animal tracking data into conservation policy and management. *Trends in Ecology & Evolution*, 34(5), 459–473. <https://doi.org/10.1016/j.tree.2019.01.009>
- Jeltsch, F., Bonte, D., Pe'er, G., Reineking, B., Leimgruber, P., Balkenhol, N., Schröder, B., Buchmann, C. M., Mueller, T., Blaum, N., Zurell, D., Böhning-Gaese, K., Wiegand, T., Eccard, J. A., Hofer, H., Reeg, J., Eggers, U., & Bauer, S. (2013). Integrating movement ecology with biodiversity research – Exploring new avenues to address spatio-temporal biodiversity dynamics. *Movement Ecology*, 1(1), 6. <https://doi.org/10.1186/2051-3933-1-6>
- Klaassen, R. H. G., Hake, M., Strandberg, R., Koks, B. J., Trierweiler, C., Exo, K.-M. M., Bairlein, F., & Alerstam, T. (2014). When and where does mortality occur in migratory birds? Direct evidence from long-term satellite tracking of raptors. *Journal of Animal Ecology*, 83(1), 176–184. <https://doi.org/10.1111/1365-2656.12135>
- Lascelles, B. G., Langham, G. M., Ronconi, R. A., & Reid, J. B. (2012). From hotspots to site protection: Identifying marine protected areas for seabirds around the globe. *Biological Conservation*, 156, 5–14. <https://doi.org/10.1016/j.biocon.2011.12.008>
- Lascelles, B. G., Taylor, P. R., Miller, M. G. R., Dias, M. P., Opper, S., Torres, L., Hedd, A., Le Corre, M., Phillips, R. A., Shaffer, S. A., Weimerskirch, H., & Small, C. (2016). Applying global criteria to tracking data to define important areas for marine conservation. *Diversity and Distributions*, 22(4), 422–431. <https://doi.org/10.1111/ddi.12411>
- Léandri-Breton, D.-J., Tarroux, A., Elliott, K. H., Legagneux, P., Angelier, F., Blévin, P., Brâthen, V. S., Fauchald, P., Goutte, A., Jouanneau, W., Tartu, S., Moe, B., & Chastel, O. (2021). Long-term tracking of an Arctic-breeding seabird indicates high fidelity to pelagic wintering areas. *Marine Ecology Progress Series*, 676, 205–218. <https://doi.org/10.3354/meps13798>
- Legrand, B., Benneveau, A., Jaeger, A., Pinet, P., Potin, G., Jaquemet, S., & Le Corre, M. (2016). Current wintering habitat of an endemic seabird of Réunion Island, Barau's petrel *Pterodroma baraui*, and predicted changes induced by global warming. *Marine Ecology Progress Series*, 550, 235–248. <https://doi.org/10.3354/meps11710>
- Lisovski, S., & Hahn, S. (2012). GeoLight – Processing and analysing light-based geolocator data in R. *Methods in Ecology and Evolution*, 3(6), 1055–1059. <https://doi.org/10.1111/j.2041-210X.2012.00248.x>
- Long, J. A. (2019). jtools: Analysis and presentation of social scientific data. <https://cran.r-project.org/package=jtools>
- Marra, P. P., Cohen, E. B., Harrison, A.-L., Studds, C. E., & Webster, M. S. (2019). Migratory connectivity. In K. R. Crooks & M. A. Sanjayan (Eds.), *Encyclopedia of animal behavior* (2nd ed., pp. 643–654). Elsevier.
- Martínez-Minaya, J., Cameletti, M., Conesa, D., & Pennino, M. G. (2018). Species distribution modeling: A statistical review with focus in spatio-temporal issues. *Stochastic Environmental Research and Risk Assessment*, 32(11), 3227–3244. <https://doi.org/10.1007/s00477-018-1548-7>
- Matthiopoulos, J. (2003). The use of space by animals as a function of accessibility and preference. *Ecological Modelling*, 159(2–3), 239–268. [https://doi.org/10.1016/S0304-3800\(02\)00293-4](https://doi.org/10.1016/S0304-3800(02)00293-4)
- McConnell, B. J., Chambers, C., & Fedak, M. A. (1992). Foraging ecology of southern elephant seals in relation to the bathymetry and productivity of the Southern Ocean. *Antarctic Science*, 4, 393–398.
- Müller, M., Massa, B., Phillips, R. A., & Dell'Omo, G. (2014). Individual consistency and sex differences in migration strategies of Scopoli's shearwaters *Calonectris diomedea* despite year differences. *Current Zoology*, 60(5), 631–641.
- Nekolny, S. R., Denny, M., Biedenbach, G., Howells, E. M., Mazzoil, M., Durden, W. N., Moreland, L., David Lambert, J., & Gibson, Q. A. (2017). Effects of study area size on home range estimates of common bottlenose dolphins *Tursiops truncatus*. *Current Zoology*, 63(6), 693–701. <https://doi.org/10.1093/cz/zox049>
- O'Toole, M., Queiroz, N., Humphries, N. E., Sims, D. W., Sequeira, A. M. M., & Freckleton, R. (2020). Quantifying effects of tracking data bias on species distribution models. *Methods in Ecology and Evolution*, 12(1), 170–181. <https://doi.org/10.1111/2041-210X.13507>
- Paiva, V. H., Geraldes, P., Marques, V., Rodríguez, R., Garthe, S., & Ramos, J. A. (2013). Effects of environmental variability on different trophic levels of the North Atlantic food web. *Marine Ecology Progress Series*, 477(August), 15–28. <https://doi.org/10.3354/meps10180>
- Patterson, A., Gilchrist, H. G., Gaston, A., & Elliott, K. H. (2021). Northwest range shifts and shorter wintering period of an Arctic seabird in response to four decades of changing ocean climate. *Marine Ecology Progress Series*, 679, 163–179. <https://doi.org/10.3354/meps13890>
- Péron, C., Authier, M., & Grémillet, D. (2018). Testing the transferability of track-based habitat models for sound marine spatial planning. *Diversity and Distributions*, 24, 1772–1787. <https://doi.org/10.1111/ddi.12832>
- R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Ramos, R. (2019). Crossing the pillars of Hercules: Understanding trans-oceanic migrations of seabirds throughout their breeding range. *Ecology and Evolution*, 9(8), 4760–4771. <https://doi.org/10.1002/ece3.5079>
- Robillard, A., Gauthier, G., Therrien, J. F., & Bêty, J. (2018). Wintering space use and site fidelity in a nomadic species, the snowy owl. *Journal of Avian Biology*, 49(5), 1–13. <https://doi.org/10.1111/jav.01707>
- Schmidt, K. A. (2004). Site fidelity in temporally correlated environments enhances population persistence. *Ecology Letters*, 7(3), 176–184. <https://doi.org/10.1111/j.1461-0248.2003.00565.x>
- Schumm, Y. R., Metzger, B., Neuling, E., Austad, M., Galea, N., Barbara, N., & Quillfeldt, P. (2021). Year-round spatial distribution and migration phenology of a rapidly declining trans-Saharan migrant—Evidence of winter movements and breeding site fidelity in European turtle

- doves. *Behavioral Ecology and Sociobiology*, 75(11), 1–16. <https://doi.org/10.1007/s00265-021-03082-5>
- Shimada, T., Duarte, C. M., Al-Suwailem, A. M., Tanabe, L. K., & Meekan, M. G. (2021). Satellite tracking reveals nesting patterns, site fidelity, and potential impacts of warming on major green turtle rookeries in the Red Sea. *Frontiers in Marine Science*, 8. <https://doi.org/10.3389/fmars.2021.633814>
- Spiegel, O., Leu, S. T., Bull, C. M., & Sih, A. (2017). What's your move? Movement as a link between personality and spatial dynamics in animal populations. *Ecology Letters*, 20(1), 3–18. <https://doi.org/10.1111/ele.12708>
- Switzer, P. V. (1993). Site fidelity in predictable and unpredictable habitats. *Evolutionary Ecology*, 7(6), 533–555. <https://doi.org/10.1007/BF01237820>
- Thibault, J.-C. (1994). Nest-site tenacity and mate fidelity in relation to breeding success in Cory's shearwater *Calonectris diomedea*. *Bird Study*, 41(1), 25–28. <https://doi.org/10.1080/00063659409477193>
- Torney, C. J., Lamont, M., Debell, L., Angohiatok, R. J., Leclerc, L. M., & Berdahl, A. M. (2018). Inferring the rules of social interaction in migrating caribou. *Philosophical Transactions of the Royal Society of London B*, 373(1746), 20170385. <https://doi.org/10.1098/rstb.2017.0385>
- Wakefield, E. D., Bodey, T. W., Bearhop, S., Blackburn, J., Colhoun, K., Davies, R., Dwyer, R. G., Green, J. A., Grémillet, D., Jackson, A. L., Jessopp, M. J., Kane, A., Langston, R. H., Lesacroë, A., Murray, S., le Nuz, M., Patrick, S. C., Péron, C., Soanes, L. M., ... Hamer, K. C. (2013). Space partitioning without territoriality in gannets. *Science*, 341, 68–70.
- Wakefield, E. D., Owen, E., Baer, J., Carroll, M. J., Daunt, F., Dodd, S. G., Green, J. A., Guilford, T., Mavor, R. A., Miller, P. I., Newell, M. A., Newton, S. F., Robertson, G. S., Shoji, A., Soanes, L. M., Votier, S. C., Wanless, S., & Bolton, M. (2017). Breeding density, fine-scale tracking, and large-scale modeling reveal the regional distribution of four seabird species. *Ecological Applications*, 27(7), 2074–2091. <https://doi.org/10.1002/eap.1591>
- Webster, M. S., Marra, P. P., Haig, S. M., Bensch, S., & Holmes, R. T. (2002). Links between worlds: Unraveling migratory connectivity. *Trends in Ecology and Evolution*, 17, 76–83.
- Wege, M., Tosh, C. A., de Bruyn, P. J. N., & Bester, M. N. (2016). Cross-seasonal foraging site fidelity of subantarctic fur seals: Implications for marine conservation areas. *Marine Ecology Progress Series*, 554, 225–239. <https://doi.org/10.3354/meps11798>
- Wilmers, C. C., Nickel, B. A., Bryce, C. M., Smith, J. A., Wheat, R. E., & Yovovich, V. (2015). The golden age of bio-logging: How animal-borne sensors are advancing the frontiers of ecology. *Ecology*, 96(7), 1741–1753. <https://doi.org/10.1890/14-1401.1>
- Wilson, R. P., Ducamp, J. J., Rees, W. G., Culik, B. M., & Niekamp, K. (1992). Estimation of location: Global coverage using light intensity. In I. G. Priede & S. M. Swift (Eds.), *Wildlife telemetry: Remote monitoring and tracking of animals* (pp. 131–134). Ellis Horwood.
- Winner, K., Noonan, M. J., Fleming, C. H., Olson, K. A., Mueller, T., Sheldon, D., & Calabrese, J. M. (2018). Statistical inference for home range overlap. *Methods in Ecology and Evolution*, 2018(January), 1679–1691. <https://doi.org/10.1111/2041-210X.13027>
- Wolf, J. B. W., & Trillmich, F. (2007). Beyond habitat requirements: Individual fine-scale site fidelity in a colony of the Galapagos Sea lion (*Zalophus wollebaeki*) creates conditions for social structuring. *Oecologia*, 152(3), 553–567. <https://doi.org/10.1007/s00442-007-0665-7>
- Yurkowski, D. J., Auger-Méthé, M., Mallory, M. L., Wong, S. N. P., Gilchrist, G., Derocher, A. E., Richardson, E., Lunn, N. J., Hussey, N. E., Marcoux, M., Togunov, R. R., Fisk, A. T., Harwood, L. A., Dietz, R., Rosing-Asvid, A., Born, E. W., Mosbech, A., Fort, J., Grémillet, D., ... Ferguson, S. H. (2018). Abundance and species diversity hotspots of tracked marine predators across the North American Arctic. *Diversity and Distributions*, 25, 328–345. <https://doi.org/10.1111/ddi.12860>
- Zango, L., Reyes-González, J. M., Militão, T., Zajková, Z., Álvarez-alonso, E., Ramos, R., & González-Solís, J. (2019). Year-round individual specialization in the feeding ecology of a long-lived seabird. *Scientific Reports*, 9(1), 1–12. <https://doi.org/10.1038/s41598-019-48214-0>

## BIOSKETCH

**Virginia Morera-Pujol** is a postdoctoral researcher at University College Dublin, broadly interested in the field of spatial ecology. Her research interests include the analysis of movement data and their use to infer populations' distributions and species distribution models and are centred around the development and improvement of methods for the modelling of species distributions. This study was part of her Ph.D. dissertation.

Author contributions: V.M.P., J.G.S. and R.R. conceived the ideas; V.M.P., J.G.S. and R.R. designed the methodology; V.M.P., C.P., P.C., M.M., J.M.R.G., J.P.G., T.M., M.D., D.O., G.D.O., M.Mü., V.H.P., B.M., V.N., J.N., G.K., S.X., J.G.C., A.Z.L., M.G.F., R.O., M.S.R., F.d.F., Z.Z., M.C.F.D.G., J.G.S. and R.R. collected and/or provided data; V.M.P., J.M.R.G., T.M. and F.d.F. analysed data; V.M.P. led the writing of the manuscript; J.G.S. and R.R. contributed to the writing process; P.C., M.M., J.M.R.G., J.P.G., T.M., M.D., V.H.P., V.N., J.N., G.K., S.X., A.Z.L., F.d.F., Z.Z., M.C.F. and D.G. reviewed and contributed to the final draft.

## SUPPORTING INFORMATION

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

**How to cite this article:** Morera-Pujol, V., Catry, P., Magalhães, M., Péron, C., Reyes-González, J. M., Granadeiro, J. P., Militão, T., Dias, M. P., Oro, D., Dell'Omo, G., Müller, M., Paiva, V. H., Metzger, B., Neves, V., Navarro, J., Karris, G., Xirouchakis, S., Cecere, J. G., Zamora-López, A. ... Ramos, R. (2023). Methods to detect spatial biases in tracking studies caused by differential representativeness of individuals, populations and time. *Diversity and Distributions*, 29, 19–38. <https://doi.org/10.1111/ddi.13642>