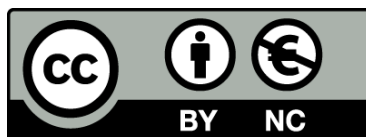




UNIVERSITAT DE
BARCELONA

Multi-colony approaches to study migratory and foraging strategies in pelagic seabirds

Virginia Morera Pujol



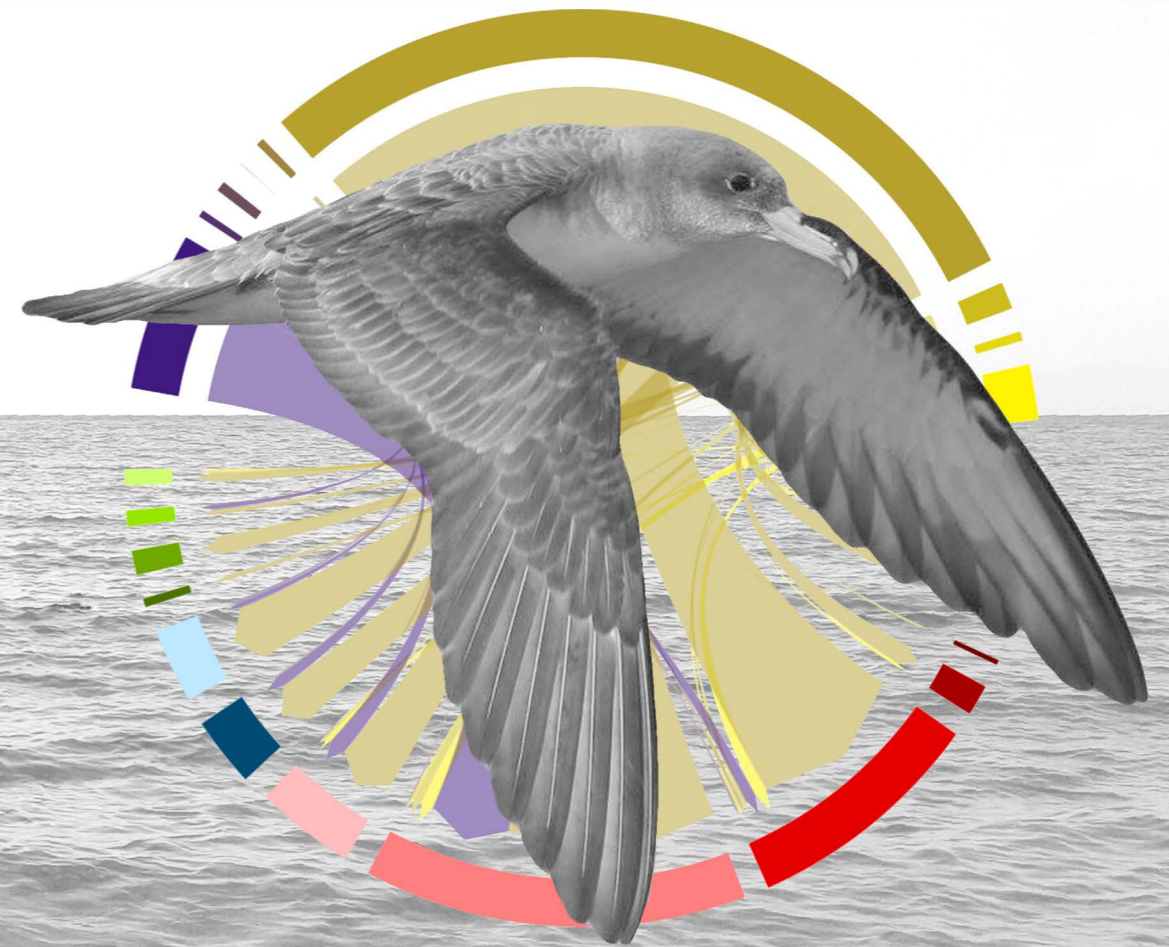
Aquesta tesi doctoral està subjecta a la llicència **Reconeixement- NoComercial 4.0. Espanya de Creative Commons.**

Esta tesis doctoral está sujeta a la licencia **Reconocimiento - NoComercial 4.0. España de Creative Commons.**

This doctoral thesis is licensed under the **Creative Commons Attribution-NonCommercial 4.0. Spain License.**

Multi-colony approaches to study migratory and foraging strategies in pelagic seabirds

Virginia Morera Pujol



September 2019



UNIVERSITAT DE
BARCELONA

Facultat de Biologia

Departament de Biologia Animal, Ecologia i Ciències Ambientals

Programa de Doctorat en Biodiversitat (HDK04)

**Multi-colony approaches to study migratory
and foraging strategies in pelagic seabirds**

*Una aproximació multi-colònia a l'estudi de les
estratègies migratòries i de recerca d'aliment en aus
pelàgiques*

Memòria presentada per

Virginia Morera Pujol

per optar al grau de Doctora per la Universitat de Barcelona,

Barcelona, Setembre de 2019

La doctoranda,

El director de la tesi,

El director i tutor de la tesi,

Acknowledgments

Algú (no recordo qui) em va dir en algun moment d'aquest any boig que fes els agraïments amb temps, que després a última hora tindria les neurones fregides i no sabia a qui he d'agrair què. Tant de bo li hagués fet cas, però fer les coses amb temps no serà mai el meu estil, així que sense ordre ni prioritat, aquí van els meus agraïments:

El Raül i el Jacob han corregut aquest final de maratón amb mi i això, en si, ja és una fita. Més enllà d'aquest tram final, no sé ni com comptar, molt menys agrair, les estones de “Raül, baixem a buscar un cafè” que s'han convertit en hores i hores de discussió i resolució de problemes, i sense les quals aquesta tesi no estaria ni la meitat de bé (i ni parlar de la maleïda consistència!). Ni la de vegades que un “Jacob, jo això no ho sé fer” ha rebut un “però segur que te'n surts” per resposta, i ves per on, la majoria tenies raó! Aquesta confiança cega, de vegades tan desesperant, m'ha empès fins a on som ara, gràcies!

Teresa y Jose me enseñaron todo lo que necesitaba saber para arrancar en esta carrera de locos, y me siguen enseñando cada día cosas nuevas. Sin ellos, no hubiese partido con la ventaja que me ha permitido llegar hasta aquí. Con ellos y con Zuzana, Fernanda, Marta, Laura y Leia hemos remado en el mismo barco, haciendo el camino mucho más agradable. Siempre hay algo que aprender de tus compañeros de viaje, y yo he aprendido de cada una de vosotras. Eli, Moisés, Jazel, y demás compañeros no-jacobinos de despacho y departamento, gracias por las

risas, las horas de comer, las cervezas de viernes tarde y tantas otras cosas.

Amigues i amics de fora del món de la ciència, però que m'heu acompanyat en aquest camí, fent-lo tan més fàcil que no sé com me n'hagués sortit sense vosaltres. Sou enterament responsables de que jo no hagi acabat parant boja durant aquests quatre anys i mig. Sou el contrapunt que necessito per relativitzar i veure les coses amb una perspectiva difícil de guanyar quan et capfiques tant en un projecte, i mai ho podré agrair prou. En aquest últim tram de bogeria, de fet, no hagués menjat, ni vestit, ni probablement sobreviscut sense l'Alba i l'Emma cuidant-me, cuinant per a mi, recollint-me la roba estesa, i en general, tenint una paciència enorme amb la meva invasió de la casa per treballar a totes hores.

The three months I spent in Saint Andrews have been one of the highlights of this already amazing journey. And they wouldn't have been half as good without everyone at CREEM being so incredibly nice to me, without Claudia and Valentin being so welcoming and friendly, and without Andy and his immense patience and kindness in explaining to me what I'm sure were the most basic statistical concepts. Of course, I am also incredibly grateful to Janine for accepting me there, for teaching me so, so much, and for always encouraging me to learn more. Lastly, a big, huge thank you to Lucía, for being as awesome as she always was and for introducing me to her group of friends from the very beginning of my time there.

Por último, a mi familia. No hay como agradecer una vida de apoyo a esta carrera loca que he escogido, de paciencia y comprensión con las

ausencias por trabajo de campo, estancias, congresos y demás. Ha sido un largo camino hasta aquí, y lo que queda por delante, pero vaya donde vaya, siempre me dais un motivo para volver.

Abstract

Movement is a widespread characteristic in the animal kingdom —occurring at many spatiotemporal scales— with consequences at an individual, population, species, and even ecosystem level. It is a very diverse character, with many different drivers that stem from the way in which individuals interact with their environment. Of these, one of the most important is the distribution of resources, particularly for migratory and foraging movements. In migration, the search for an optimal environment involves movement at large spatiotemporal scales, following seasonal changes in resource distribution. In foraging movements, the search for resources happens at small spatiotemporal scales, and involves different strategies to optimise the search and capture of food, including the ability to obtain foraging cues from conspecifics. In seabirds, movement —at large and local scales— has deep repercussions in their life-history traits, evolutionary history, morphology, physiology and behaviour, which makes them a very valuable study group to understand the role, the causes and consequences of migratory and foraging movements in the ecology of marine top-predators.

The study of migratory and foraging movements has been revolutionised by the development of smaller, cheaper and better tracking devices, promoting multi-colony, population and even species approaches to the study of animal movement, but which also come with a set of methodological challenges that have to be addressed in order to make unbiased inferences of space and habitat use at population or species level from individual movement data.

In this thesis, we develop methods to test the possible biases introduced by the use of individual tracking data to infer distribution at a population or species level. we then apply these tools to a multi-colony dataset of non-breeding locations of Cory's (*Calonectris borealis*), Scopoli's (*C. diomedea*) and Cape Verde (*C. edwardsii*) shearwaters, to study their migratory connectivity and non-breeding habitat segregation at the colony, population and species level. Lastly, we apply state-of-the-art spatial models to study foraging distributions of three neighbouring colonies of Cory's shearwaters, detect the segregation among them and unravel the environmental and behavioural drivers of this segregation.

I developed several functions in the R environment aimed at the detection of the effects of individual site fidelity and temporal variability in the inference of spatial use at a colony or population level, and to calculate the degree in which the movements of a single population can be representative of those of the entire species. These tools are applicable to individual movement data regardless of the species or tracking device. we also used these tools to demonstrate the spatial and ecological segregation between the non-breeding distributions of three taxa of *Calonectris* shearwaters studied, as well as detecting a stronger degree of migratory connectivity at a population than at a colony level, indicating that individuals of different colonies within a population mix in the non-breeding areas, but birds from different populations do not, which has important implications for their population dynamics and for their conservation and management. Lastly, we demonstrated segregation among the foraging distributions of three neighbouring colonies of Cory's shearwaters, both in the waters surrounding the colony

and in distant, foraging grounds, finding evidences of both environmental and behavioural drivers behind this segregation, and suggesting a mechanism through which transfer of information between individuals can be shaping the distributions of foraging seabirds.

The results of this thesis provide relevant tools for the field of movement ecology, as they can be used for analysing movements of mobile species, regardless of species, tracking device or spatiotemporal scale. In addition, they are relevant for the field of seabird ecology as they provide insights into the causes of space and habitat use in long-ranging pelagic seabirds.

Resum

El moviment és una característica omnipresent en el regne animal, a les més diverses escales espacio-temporals i amb conseqüències a diferents nivells (individual, poblacional, específic i fins i tot ecosistèmic). La força impulsora del moviment més predominant i generalitzada és distribució dels recursos en el medi, important tant a gran escala (moviments migratoris) com en els moviments diaris de recerca d'aliment a escala petita. En les aus marines en particular, el moviment té repercussions profundes en les seves característiques morfològiques i de comportament, en la seva història de vida, i la seva història evolutiva. El desenvolupament de dispositius de seguiment més petits, barats i precisos ha promocióat la proliferació d'estudis del moviment animal des d'un punt de vista multi-colònia, de població i fins i tot d'espècie. En aquesta tesi, desenvolupo diferents funcions per testar els biaixos introduïts en l'estudi del moviment, a través de dades de seguiment individual, a nivell de població o espècie. Posteriorment, utilitzo aquestes eines per a analitzar la connectivitat migratòria i la segregació dels hàbitats d'hivernada, des d'un punt de vista multi-colònia, de les baldrigues cendroses de l'Atlàntic (*Calonectris borealis*), del Mediterrani (*C. diomedea*), i de Cap Verd (*C. edwardsii*). Per últim, aplico un mètode innovador de modelatge espacial per a estudiar les distribucions d'alimentació de baldrigues cendroses de l'Atlàntic criant en tres colònies veïnes, per detectar-ne la segregació i descobrir-ne les causes, tant ambientals com comportamentals, incloent com els diferents mecanismes de transferència d'informació entre individus poden afectar a aquestes distribucions. Els resultats d'aquesta

tesi tenen rellevància per a la ecologia del moviment en general, ja que són aplicables a dades de moviment individual de qualsevol espècie, independentment de l'aparell de seguiment utilitzat, i en el camp de l'ecologia de les aus marines, ja que proporciona nous coneixements sobre els diferents factors afectant la distribució i l'ús de l'espai, tant durant la cria com durant la hivernada, en aus pelàgiques.

TABLE OF CONTENTS

Introduction	1
1.1 Animal movement: an overview	1
1.2 Migratory movements	5
1.2.1 <i>Overview of migratory movements</i>	5
1.2.2 <i>Migration in seabirds</i>	8
1.3 Foraging movements	10
1.3.1 <i>Overview of foraging movements</i>	10
1.2.2 <i>Foraging movements in seabirds</i>	14
1.4 Methods to study seabird movement	17
1.4.1 <i>Overview</i>	17
1.4.2 <i>The study of space use</i>	22
1.4.3 <i>The study of habitat use</i>	24
1.5 Uses of tracking data for conservation	26
1.6 Aims	27
1.7 References	29
Detecting recurrent sources of variability in animal tracking studies	47
2.1 Abstract	47
2.2 Introduction	49
2.3 Methods	52
2.3.1 <i>Testing for individual site fidelity</i>	52
2.3.2 <i>Testing for temporal variability</i>	53
2.3.3 <i>Testing for spatial differences</i>	54
2.3.4 <i>Empirical application</i>	55
2.4 Results	64
2.4.1 <i>Individual site fidelity effect</i>	64
2.4.2 <i>Inter-annual differences in distribution</i>	64
2.4.3 <i>Spatial differences in species representativeness</i>	65
2.4.4 <i>Non-breeding distributions of three species of <i>Calonectris</i> shearwaters</i>	67
2.5 Discussion	72

2.6	References	75
------------	-------------------	-----------

Migratory connectivity and non-breeding habitat segregation across geographical scales in three closely related seabird taxa		81
3.1	Abstract	81
3.2	Introduction	83
3.3	Methods	86
3.3.1	<i>Studied taxa and sampling design</i>	86
3.3.2	<i>Tracking data</i>	87
3.3.3	<i>Migratory connectivity</i>	89
3.3.4	<i>Environmental data and ecological niche modelling</i>	92
3.4	Results	96
3.4.1	<i>Migratory connectivity</i>	96
3.4.2	<i>Ecological niche modelling</i>	101
3.5	Discussion	107
3.5.1	<i>Ecological segregation</i>	107
3.5.2	<i>Spatial structure (within a given taxon)</i>	111
3.5.3	<i>Concluding remarks</i>	114
3.6	References	114

Environmental and behavioural drivers of foraging segregation in a long-range seabird: insights from Bayesian spatial models within the INLA framework		123
4.1	Abstract	123
4.2	Introduction	124
4.3	Methods	130
4.3.1	<i>Study species and area</i>	130
4.3.2	<i>Environmental data</i>	134
4.3.3	<i>Tracking data</i>	137
4.3.4	<i>Overlap analysis</i>	138
4.3.4	<i>Hierarchical Bayesian modelling</i>	139

4.4	Results	149
4.4.1	<i>Overlap analysis</i>	149
4.4.2	<i>Hierarchical Bayesian modelling</i>	151
4.4.3	<i>Model evaluation</i>	151
4.4.4	<i>Environmental covariates</i>	152
4.4.5	<i>Spatial effects</i>	158
4.5	Discussion	158
4.5.1	<i>Conclusions</i>	163
4.6	References	164

Discussion	173
-------------------	------------

5.1	Addressing sources of variability in the inference from individual tracking data	174
5.2	Space and habitat use in pelagic seabirds	180
5.3	Implications for conservation of the findings of this work	188
5.4	Literature	190

Conclusions	195
--------------------	------------

1

Introduction

1.1 Animal movement: an overview

Movement is a fundamental trait for most animal species, occurring at the most diverse spatial and temporal scales. It is driven by processes that result from the interaction of individuals with their environment (i.e. extrinsic drivers, including biotic and abiotic characteristics), and can be modulated by individual preferences (i.e. intrinsic drivers). These driving processes can be a response to short-term goals, such as search for resources, reproduction, and avoiding risks such as predation, but can also be a response to long-term fitness implications such as avoidance of inbreeding or long-term population survival (Table 1; Holyoak et al. 2008, Avgar et al. 2013).

Table 1

Main drivers of animal movement

Driver	Mechanism(s)	Reference
<u>Extrinsic</u>		
<i>Distribution of resources</i>	Optimisation of resource acquisition Seasonal migration	Matthiopoulos 2003, Trevail et al. 2019
<i>Landscape structure</i>	Adjustment of search behaviour to landscape heterogeneity	Bastille-Rousseau et al. 2017, Tucker et al. 2019
<i>Threat distribution</i>	Predator avoidance Pathogen avoidance	Guzman et al. 2019, Westerdahl et al. 2014
<i>Presence and distribution of conspecifics</i>	Territorialism Competition avoidance Attraction Search of mate for reproduction	Maher and Lott 1995, Stillman et al. 2000
<i>Avoidance of inbreeding</i>	Natal dispersal	Pusey 1987
<i>Population survival</i>	Metapopulation dynamics (emigration/immigration among populations)	Morales et al. 2010
<u>Intrinsic</u>		
<i>Memory</i>	Home range fidelity Non-breeding area fidelity	Fagan et al. 2013, Lafontaine et al. 2017
<i>Personality</i>	Exploration – exploitation trade-off	Spiegel et al. 2017, Patrick et al. 2017
<i>Sex</i>	Competition avoidance Different reproductive investment Different energetic requirements	Wearmouth and Sims 2008
<i>Diet</i>	Costs of food acquisition and digestion Spatial distribution of food	Tucker et al. 2014
<i>Size</i>	Energy requirements and interactions with neighbours	Jetz 2004
<i>Age/experience</i>	Learning	Riotte-Lambert and Weimerskirch 2013

Environmental characteristics are one of the main drivers of animal movement, and particularly resource distribution can affect movement across diverse spatiotemporal scales (from seasonal, long-scale migrations to daily, short-range foraging movements within an individual home range). Despite being usually studied in two main blocks within movement ecology: **migration** and **foraging movements** share the common driver of the optimisation of resource search, and are relevant across species, regardless of sex, age, experience or any other intrinsic individual factor. In addition, the two have common impacts and repercussions at many levels: (1) in the individual fitness and evolutionary processes, (2) in the structuring of communities and populations, (3) in the species responses to environmental change, and (4) in the ecosystem functioning (Hooten et al. 2017).

At an individual level, movement can have consequences on fitness since it is the mechanism for resource acquisition, and nutrient intake has an obvious effect on individual body condition, survival and reproduction (Parker et al. 2009). In addition to the proximal effects of movement on individual condition, detrimental non-lethal effects of resource acquisition in one season can carry over and affect survival or reproductive output in the subsequent season (Harrison et al. 2011). At a population level, the spatial structure of animal populations can range from closed populations (assumed free flow of individuals within a population) to structured subpopulations connected among them by immigration/emigration, and the degree of connectivity among populations (dependent on movement) will have important consequences for population dynamics (Hanski and Gaggiotti 2004). At a species level, the degree of spatial flexibility of a given species (determined by the

mobility of its individuals and the degree of philopatry and spatial fidelity they present) will be a key aspect to determine the species ability to adapt to an ever-changing environment, therefore, species with different degrees of site fidelity and breeding philopatry could respond differently to climate change (Ronce 2007; Robinson et al. 2009). Lastly, mobile individuals also have an effect on the ecosystems that they inhabit: at a landscape level, animals that move across habitats and ecosystems can provide important services such as pollination, seed dispersal, and movement of nutrients (Lundberg and Moberg 2003; Savage 2019).

Key taxa where both (migratory and foraging) movements are crucial to understand are seabirds. They inhabit at the interface of three media: they breed on land, fly through the air, but also feed and spend a significant part of their lives at sea. Thus, their movements have deep repercussions in life history traits, evolutionary history, morphology, physiology and behaviour, including here their foraging behaviour (Hamer et al. 2002). As most marine top predators, seabirds present extreme life-history traits resulting from an adaptation to forage in an environment, the ocean, which at fine scales shows a reduced availability and low spatiotemporal predictability of resources, but at large scales can vary predictably both in space (upwelling areas are almost constantly highly productive) and temporally (seasonality). Such extreme life-history traits include low reproductive outputs, reducing the energetic demands of breeding adults to one or a few chicks per breeding season (Lewison et al. 2004); long lifespans and low adult (natural) mortalities that increase life-long reproductive success (Weimerskirch 2002); and deferred maturities that allow immature birds to acquire the

ability to find prey at sea by spending the first years of their lives exploring the vast ocean and developing their own foraging skills (Grecian et al. 2018; Ramos et al. 2019). All these characteristics make seabirds a very valuable study group to understand the role, the causes and consequences of migratory and foraging movements in marine top-predators, which remains a key issue in animal ecology.

1.2 Migratory movements

1.2.1 Overview of migratory movements

Defining migration is not an easy task. The most obvious behaviours that have traditionally been considered migrations are the “to-and-fro movements over dramatic distances between breeding and wintering¹ grounds to secure optimal environmental conditions at all times” (Thomson 1926). However, since this definition of migration was proposed, many other types of movement have been described which, despite not having a return trip or not being seasonal, also fall under the function of securing optimal environmental conditions at all times (e.g. change of environments in different ontogenic stages of insects and some fish; Dingle and Drake 2007). This called for a more general definition of migration, which could include a wider range of migratory movements but not so wide as to be devoid of ecological meaning. Dingle (1996)

¹ The areas used by populations outside the breeding season have traditionally been called wintering grounds, because the first studied species breed mostly in summer. However, these should more accurately be called non-breeding grounds, since not all species breed in summer and, even for those species that do, if they cross the Equator to reach the non-breeding grounds it will be summer in there too.

described migration based on five characteristics, so any migratory population should present at least one of these: (1) persistent movement between two habitats where resources are available; (2) straightened out movement, the direction of which may or may not be decided and maintained through navigation behaviours; (3) migrating organisms are undistracted by stimuli that would otherwise attract them, like resources or favourable habitats, although some birds and insects might feed *en route* to replenish energy reserves and even hold temporary territories; (4) behavioural changes before and/or after migration, with pre-migratory hyperphagia being the most clear and known of these; and (5) reallocation of energy specifically to support migration, which implies trade-offs with other functions such as reproduction. According to these characteristics, migration is a widespread character, present in many taxonomic groups from insects, fish and reptiles to birds and mammals, and happens in all ways of locomotion (on ground, in water, on air, even drifting with currents or “parachuting” through silk threads like some spiders do; Hayashi et al. 2015). It also presents many variants that complicate its study: from an individual point of view an animal can be an obligate migrant (if they can’t skip a migration cycle) or facultative migrant (if they can decide whether to migrate based on environmental or intrinsic conditions; Chapman et al. 2011), and from a population point of view migration can be partial (when only some of the individuals of the population migrate, while others stay in either the breeding or non-breeding grounds; Chapman et al. 2011a), or differential (when only a section of the population, e.g. only males, or only immatures, migrate; Cristol et al. 1999). The presence of migration in many taxonomic groups

and the flexibility and variability of its expression provide some clues that explain its evolutionary origin (Box 1).

Box 1: evolutionary origin and ontogeny of migration

Migratory behaviour has appeared and disappeared many times in the evolutionary history, which hints to the possibility of it appearing from the activation and/or regulation of characters that exist, dormant, in resident populations. Adaptation to seasonality is one of the possible evolutionary drivers of migration: if resources or favourable habitats were spatiotemporally unstable, but predictable, populations could compensate this seasonal fluctuation by migrating to where the resources are (Alerstam et al. 2003).

Another, more recent explanation proposes that the primary driver of seasonal migration is the maintenance of fidelity to breeding sites. In this scenario, animals' movement is primed by the seasonal fluctuations of the environment, but what actually drives the evolution of migration is their determination to go back to a site where they have bred successfully (Winger et al. 2019).

Although in some species there is a clear genetic predisposition to migrate (Chapman et al. 2015; Yoda et al. 2017), many others, such as some marine mammals and some birds, acquire migratory behaviour through social learning (Palacín et al. 2011; Whitehead 2017; O'Corry-Crowe et al. 2018), and others follow cues and perform complex path integration (Müller and Wehner 2010).

Migratory species have often a more complicated population structure than resident species, which oscillates depending on the degree of connectivity between the breeding and non-breeding distributions of the species (Bauer et al. 2016). Species with strong migratory connectivity maintain the spatial structure of the breeding distribution during the non-breeding period, so animals —or populations— that breed in proximity spend all the annual cycle close by, and those that are distant from each other during breeding will be so too during the non-breeding

period, minimising the possibilities for mixing between populations (Webster et al. 2002; Marra et al. 2019). Therefore, the degree of migratory connectivity of a species has relevance not only for its population structure, but also for conservation. For example, species that experience a high degree of interpopulation mixing during the non-breeding season will be globally affected by any perturbations in the non-breeding grounds. Conversely, if the degree of mixing is low, a local perturbation might only affect one breeding population, having less effect in the entire species. However, in this case, the risk of local extinction is higher than in well-mixed populations (Ponchon et al. 2015). Similarly, populations that breed in a concentrated area but spread across the entire non-breeding range might be more affected by non-breeding habitat shrinkage, while they will be more equipped to deal with habitat degradation and loss than populations with low non-breeding spread (Finch et al. 2017).

1.2.2 Migration in seabirds

Many seabirds are migratory, performing cyclical, seasonal movements between breeding and non-breeding grounds (Dingle 1996). Some of the most impressive migrations in the world occur at sea, such as the migration of the Arctic tern (*Sterna paradisaea*) which perform the longest recorded animal migration, of up to 60,000 km (Egevang et al. 2010; Alerstam et al. 2019). The seasonal variation of environmental conditions is the main proposed driver of seabird migration, which could buffer the variation in resource availability by birds tracking optimal environmental conditions in an endless summer (Fig. 1; Alerstam et al. 2003, Shaffer et al. 2005, González-Solís et al. 2007).

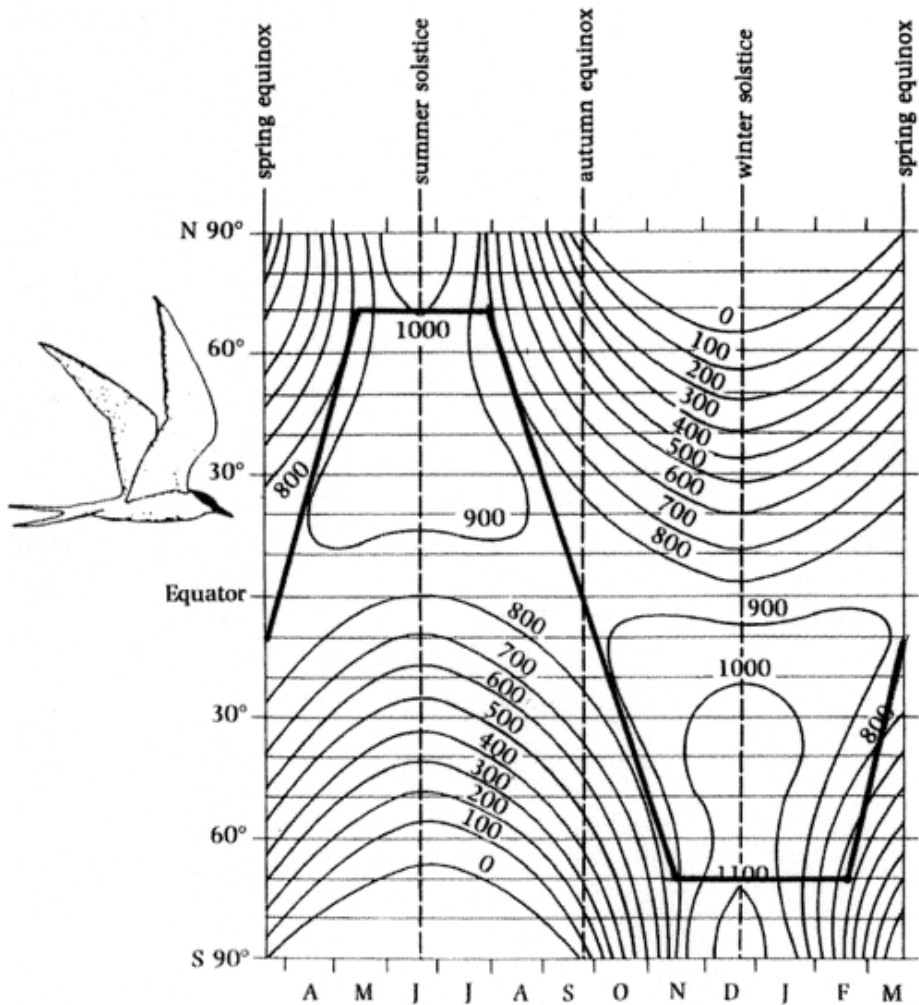


Figure 1

Daily solar energy ($\text{cal}\cdot\text{cm}^{-2}$) reaching the Earth at different latitudes and times of the year. The thick line represents the trajectory of an Arctic tern (*Sterna paradisaea*) in its migratory trip from the breeding colonies in the Arctic to the non-breeding grounds in the Antarctic. (reproduced from Alerstam et al. 2003, with permission of The Licensor through PLSclear license no 17391)

The study of migration in seabirds was at first hampered by the difficulty of obtaining information on the destination of individuals away from their breeding colonies. From the first studies using observations from

land, boat surveys, and ring recoveries (Camphuysen and Van der Meer 2001), to the first tracking devices used to follow birds in their migratory trips (Weimerskirch and Wilson 2000), the study of migration in seabirds has grown into a profusion of long-term (Desprez et al. 2018), multi-colony (Bogdanova et al. 2017), and even multi-species (Ramos et al. 2017) studies, facilitated by the development of smaller, more precise and cheaper tracking devices. This technological advancement has allowed us to gather detailed information at an individual level that, when scaled up to higher hierarchical levels (i.e. population, species, community), can provide insights into the effects of migrations on ecosystems (e.g. Mueller et al. 2014) or population processes (Morales et al. 2010), reveal effects of human alteration on animal behaviour (Tucker et al. 2018), or provide insights into conservation and management (Fraser et al. 2018). Particularly relevant the possibility of tracking several populations at once, opening a new venue for studying migratory populations from a metapopulation point of view (Fort et al. 2012; Ramos et al. 2015), which can reveal large-scale patterns not visible with single-colony studies, as well as issues of conservation concern that could affect seemingly unrelated populations (Bogdanova et al. 2017; Sherley et al. 2017).

1.3 Foraging movements

1.3.1 Overview of foraging movements

The idea that animals restrict their movements to a certain limited area is not new and, indeed, it was present as far back as Darwin’s *On the Origin of Species* (Darwin 1861). Later, Burt laid down the definition of home range as “that area traversed by an individual in its normal

activities of food gathering, mating, and caring for young” (Burt 1943). What Burt called “food gathering” is studied now as foraging behaviour. Foraging movements are those performed by animals in search of resources, which can range in extension from very localised movements to wide-ranging movements of several thousand kilometres (Pinsky and McCauley 2019). All animals need to allocate energy to survive and reproduce, but the search for resources is also energetically costly. Natural selection favours the development of behavioural and physiological characteristics that contribute to maximising foraging efficiency (Trevail, Green, Sharples, Jeffrey A. Polton, et al. 2019), so the scale at which this foraging behaviour occurs is determined by the energetic requirements of foragers and the spatial distribution of prey (Guzman et al. 2019).

Since resources are rarely uniformly distributed, adaptive characteristics will be those that maximise foraging efficiency improving an individual’s ability to identify patches suitable for foraging (e.g. following environmental cues indicative of high productivity; Hansen et al. 2016), to locate resources within those patches (e.g. by sight or smell; Potier et al. 2019), and to pursue and catch prey (e.g. plunge diving in certain species of seabirds, adaptations to deep-water sprints in pilot whales; Aguilar Soto et al. 2008, Thiebault et al. 2016). Among these characteristics maximising foraging efficiency, the ability to obtain information from their foraging conspecifics will also be a relevant factor, with direct effect on the movement patterns at an individual level, but also at population and species levels.

All the information that an animal obtains by interacting with the environment is personal information. This information can be private and inaccessible to others (i.e. memory) or non-private, and accessible to others through the production of social information (Fig. 2). In environments where resources are spatiotemporally variable, cues or signals generated from social information might be a key contributor to the optimisation of foraging strategies. Animals that forage in groups (or at least close to conspecifics) can make use of public (non-private) information to improve the fitness of their foraging decisions. This public information can be inadvertently generated or produced as a communication signal with evolutionary adaptive value (Fig. 2; Danchin 2004). Obviously, this information transfer can only occur when individuals of the same species forage close enough for the cues or signals to be transmitted. The fact that foraging in group is a widespread behaviour, even in species that are not strictly social, is a proof of its influence on fitness at a population level (Ward and Webster 2016).

The information can be transferred between individuals at the place where the resources occur or at a common location where individuals breed, rest, or shelter. When information is transmitted at the resource location, we talk about local enhancement, and when it is transmitted at a common location, we talk about information centres.

Local enhancement is based in social attraction among conspecifics regardless of mate selection (Kiestler 1979). During foraging, the presence of individuals in a foraging patch cues conspecifics on the location of the patch, particularly in instances when the foraging conspecifics are more conspicuous than resources (Buckley 1997).

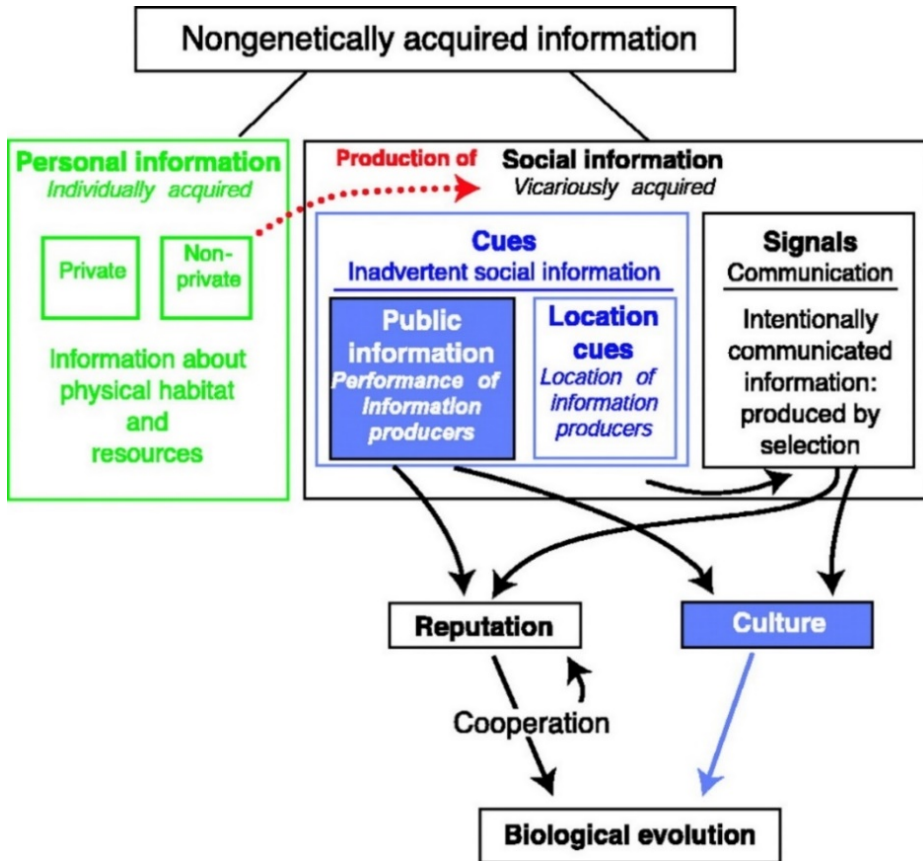


Figure 2

From all the information an individual acquires on the environment (green box) the non-private information can be transmitted to conspecifics (red arrow). If the information is transmitted willingly, we talk about communication through signals (black box) and if it is transmitted inadvertently, we talk about social information transfer through cues (blue box). From Danchin, E. (2004). Public Information: From Nosy Neighbors to Cultural Evolution. *Science*, 305(5683), 487–491. Reprinted with permission from AAAS”

The **information centre hypothesis** was developed by Ward and Zahavi (1972) to define and explain group foraging in birds that breed or roost communally. According to this hypothesis, knowledge on the location of food or good quality feeding sites is transmitted among individuals in common places where birds assemble in large numbers. It

has been suggested that the benefits of this information transfer have been the driving force behind the evolution of the roosting behaviour and coloniality in birds (Buckley 1997). Experimental and theoretical studies support the existence of information transfer between individuals in communal locations, not only for birds but also for mammals and fishes (Danchin 2004 and references therein, Agee and Monfils 2018).

1.2.2 Foraging movements in seabirds

Since seabirds breed on land but forage at sea, breeding individuals act as central place foragers during the breeding period, having to commute back and forth between the colony and the foraging areas to attend their rearing duties (Orians and Pearson 1979). The size of the colonies is determined, rather than by reproductive success or availability of breeding sites, by the availability of resources in the nearby waters, and thus, colony size often relates positively to the size of foraging area.

According to the Ashmole’s halo hypothesis, colony sizes are regulated by density-dependent effects of competition in reproductive success (rather than adult survival; Ashmole 1963). The visible consequence of this regulation is a relationship between colony size and foraging area size (Furness and Birkhead 1984). Since foraging individuals will deplete resources in the waters surrounding the colony, individuals from larger colonies will need to travel farther, thus creating a wider “halo” of resource depletion around the colony (Jovani et al. 2015; Oppel et al. 2015). The hinterland model (Cairns 1989) expands Ashmole’s model considering the interaction between neighbouring colonies. It proposes that, according to optimal foraging theory, birds of a given colony will forage only in the surrounding waters closer to their colony of origin than

to any other colonies, an area that Cairns referred to as *hinterland*. By this logic, geographic distribution of colonies will determine the size of the hinterlands and, thus, the colony sizes. Moreover, Cairns predicted that birds from neighbouring colonies would have segregated foraging areas, with the line of equidistance between colonies being the boundary between foraging areas. The inter-colony segregation predicted by the Hinterland model has been later proved for many different species of seabirds (reviewed in Bolton et al. 2018).

In addition to knowing in which direction to leave the colony, seabirds need to search and find their prey every foraging trip, since having to go back to the colony to attend their rearing duties makes them lose contact with the prey (Thiebault et al. 2014). In this regard, it has been suggested that the benefit of using public information might be one of the forces behind the evolution of coloniality in seabirds, since breeding in large aggregations of conspecifics facilitates the acquisition of information (Ward and Zahavi 1972; Buckley 1997). In fact, although colonial breeding is present in 13% of all bird species, it is concentrated in seabirds, with 96% breeding colonially (Sachs et al. 2007). This is another characteristic of foraging in the interface between two environments (air and water): birds have to constantly lose sight of their resources and consequently they are not able to monopolise them or defend them territorially. Thus, based on the principle of information transfer, for seabirds a great evolutionary advantage by breeding colonially.

Although this transfer of information is problematic to prove in seabirds, several studies, based on gannets and cormorants, have attempted to

detect it and even model it. In these species with short foraging ranges, the information conveyed by returning individuals, on the direction of profitable foraging areas, is used directly by those departing individuals, as suggested for Cape gannets (*Morus capensis*) by Grémillet et al. 2004. Using individual based models, first, Wakefield et al. (2013), and later, Boyd et al. (2016) proved the importance of information transfer events, both by local enhancement and by the transfer of information at the colony, to explain the foraging distributions of northern gannets (*M. bassanus*) and Peruvian boobies (*Sula variegata*), respectively. More recently, Jones et al. (2018) analysed departure times and overlap of first foraging patches in Australasian gannets (*M. serrator*) to demonstrate the existence of social foraging behaviours. Alternatively, for Guanay cormorants (*Phalacrocorax bouganvili*), another colony of Australasian gannets, and common guillemots (*Uria aalge*), it has been suggested that the transfer of information occurs in the near-colony rafts rather than on the colony (Burger 1997; Weimerskirch et al. 2010; Machovsky-Capuska et al. 2014).

In the case of seabirds with long foraging ranges, there are now several studies reporting segregation of foraging distributions among neighbouring colonies, such as shearwaters or albatrosses, which can perform foraging trips, lasting several days and spanning hundreds of kilometres (Bolton et al. 2018). However, the relevance of social behaviours and the transfer of information, either at sea or at the colonies or rafts has yet to be proved for any of them.

Much in the same way as the technological developments of the past decades revolutionised the study of seabird migration, the study of

seabird foraging distributions has been boosted by the development of smaller and less costly GPS tracking devices. This has allowed not only to follow the foraging movements of smaller species (Hedd et al. 2018), but also to approach the study of foraging distributions and spatial use at larger scales. That is, tracking individuals from different colonies simultaneously allows us now to address issues that escape the scope of single-colony studies, providing insights into inter-colony segregation (Wakefield et al. 2013), inter- and intra-specific competition (Wakefield et al. 2017), and conservation at a population or species scale (Ramos et al. 2013; Lascelles et al. 2016).

1.4 Methods to study seabird movement

1.4.1 Overview

The availability of increasingly detailed information on the spatiotemporal movements at an individual level has revolutionised the field of seabird movement ecology (Burger and Shaffer 2008). However, having larger and more precise datasets is not enough to answer ecological questions; we also need to know which questions can be answered with each type of data (i.e., the hypotheses we can pose), and how to answer them (i.e., the methods to be used). To ask the correct questions and obtain the correct answers, statistical methods have to evolve together with the data, otherwise we run the risk of—at best—underusing all the potential data has to offer and—at worst—misusing it and drawing wrong conclusions (Hebblewhite and Haydon 2010).

Movement can be studied from two perspectives: the **Lagrangian perspective** considers movement from the point of view of the observed

animal, focusing on individual movement, while the **Eulerian perspective** considers movement from the point of view of the observer, taking a “snapshot” of the position of all individuals in a given area at a given moment or period of time. The former perspective is, thus, more suitable for studying the consequences of movement at population levels (Table 2; Turchin 1998, Phillips et al. 2019), while Eulerian approaches to study seabird movement data allow exploring **space use** (e.g., distribution and overlap or segregation among populations) or **habitat use** (e.g. species distribution models).

For a long time, the only type of spatial data available to seabird researchers was that provided by direct observations, boat or plane surveys or re-sighting of ringed birds (Camphuysen and Van der Meer 2001). The irruption of tracking devices providing numerous individual positioning data in the seabird ecology arena has boosted the development of Lagrangian approaches to infer population and species processes from individual movement data (Box 2; Patterson et al. 2017). Lagrangian analytical approaches not only consider the location coordinates of several individuals, but also, the characteristics of their movement (i.e. distance between consecutive movements, heading, or turning angle) and they can also incorporate additional data of the environment as well as data of the physiological status of the individuals to the final model.

Tracking data, despite being individual based, can also be used to analyse population distributions from a Eulerian approach, by grouping the locations taken within a given time frame to turn movement data into static locations. In this process of inferring higher-level distributions

from individual tracking data, it is important to consider the different individual preferences can bias the inferred population distribution when only a few animals are tracked, and that presence-only data can introduce some biases in habitat use models (Table 2; Holdo and Roach 2013).

Box 2: the three most common types of Lagrangian models

Hidden Markov Models (HMM) and State Space Models (SSM) deal with discrete data at regular time intervals. They relate each location (observed state) to a hidden (unobserved) behavioural state. Space State Models, however, are able to incorporate uncertainty in the locations, and for that reason are sometimes used simply to correct location data. Diffusion models are less accessible to non-statistician users and thus less common, but deal with time as a constant (and therefore are able to deal with irregular time interval) to relate it to the behavioural states of the individual at each time point (shearwater illustration below courtesy of Manel Risa).

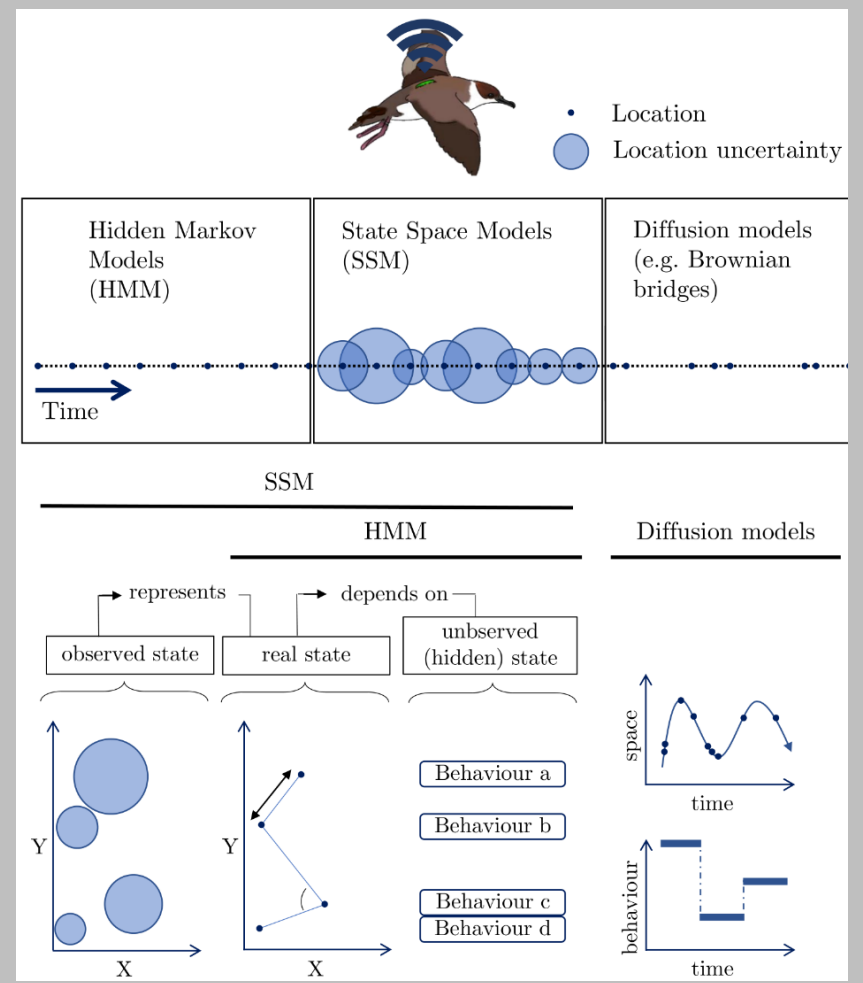


Table 2

Differences between the Eulerian and Lagrangian approaches to study animal movement

<i>Approach</i>	Eulerian	Lagrangian
<i>Focus</i>	Population redistribution	Individual movement
<i>Data collection</i>	Surveys / Mass mark recapture	Tracking / Individual mark recapture
<i>Survey area</i>	Fixed	Depends on species mobility
<i>Time framework</i>	Fixed	Depends on device time-life, potentially unlimited
<i>It measures</i>	Location of individuals Distribution patterns (aggregation, dispersion)	Movement characteristics (speed, acceleration). It can also measure physiology of individual
<i>Environmental variables</i>	Can be collected at the same time and place presence/abundance is being surveyed	Can be obtained from remote sensing and matched to track with precision depending on spatiotemporal resolution of environmental products Tracking devices can also sense environmental information ² .
<i>Limitations</i>	Depends on species or individual detectability Possible to confound temporal and spatial patterns ³ Does not provide data on breeding status, sex, age, or origin of individuals	Cost and logistical challenges can limit the number of individuals tagged or species (size range) that we can study. Population variability and individual site fidelity might complicate inference at higher levels. Presence only data might require some modifications to be used for habitat modelling

² The information gathered by these devices does not depend on resolution of external sources of data, so it is 100% accurate

³ if all individuals gather in the afternoon, we might think the abundance is higher in the regions we happen to sample in the afternoon

1.4.2 The study of space use

Methods to study space use, at both individual and population level, can go from the simple calculation of a minimum convex polygon of all positions to the most complex state-of-the-art methods. A minimum convex polygon is the smaller polygon that includes all the locations of an individual or population, and has been used in the past as a proxy for the home range (Ludynia et al. 2012). Its area, however, is very dependent on the number of locations and much affected by the presence of outliers. When comparing distributions of different individuals or populations obtained from tracking data, this can lead to the over-estimation of the foraging area of the more intensely sampled, producing erroneous conclusions with relation to their space use. In addition, a minimum convex polygon contains all the locations from an individual, negating the home range condition of “areas frequently used by an individual” as it contains all areas used, i.e. those used frequently and unfrequently. Lastly, a minimum convex polygon does not provide information of how animal positions are located inside the polygon, whether they are regularly distributed, clustered around a central location, or forming different small clusters in different locations. For these and other methodological issues, estimation of habitat use through minimum convex polygons was in time replaced by kernel density estimation (González-Solís et al. 2000; Wood et al. 2000). This probabilistic method provides a bivariate probability density function, the utilisation distribution (Winkle 1975), describing the probability of finding an individual at any location. From the utilisation distributions, the 95% contour was quickly adapted as indicator of the home range and the 50% contour as an indicator of the core use area. In spite of

representing an improvement over previous methods, kernel density estimation does not consider the spatial autocorrelation between consecutive points. There has been some debate regarding whether that constitutes an issue and skews space use estimation (e.g. Börger et al. 2006), but kernel density estimation continues to be a widely used method to estimate space use in movement ecology (Diop et al. 2018; Trevail, Green, Sharples, Jeff A. Polton, et al. 2019). However, more sophisticated methods for calculating kernel density surfaces have been developed and applied to seabird tracking data more recently. Movement-based kernel density estimation can incorporate time, distance, and habitat into estimates of home range (Walter et al. 2015), and therefore, they are particularly suited to study movement paths such as migratory corridors (Tracey et al. 2014). Brownian bridge kernels can incorporate error in locations, and are thus very well suited to model unprecise location data such as that obtained from some satellite platforms (e.g., ARGOS) or from geolocators (Kranstauber 2019).

All these ever-evolving techniques to study and represent space use employ individual data to scale up the behaviour of a few individuals to the level of population, or even species. However, this inference is not straightforward, as there are many biases and sources of variability that can greatly affect the results we extract at a higher level, and they should be considered from the first stages of study design. First, individuals from different colonies might respond differently to the environment or have different behaviours and spatial distributions (Young et al. 2009; Yamamoto et al. 2012), so we cannot assume that the conclusions gathered from a single colony can represent properly the entire population or species. Second, despite being based on individual

decisions, an animal's use of space is linked to the optimal use of resources, and therefore, affected by the environmental dynamism (Wolf and Trillmich 2007). Consequently, the time scale within which we pool data must also be considered to avoid introducing spatial variability in distribution data (Dias et al. 2011; Paiva et al. 2013). Lastly, if tracks from individuals from different populations are available, before pooling all tracks together to infer population-level behaviours we should consider both colony size and sample size of each population in the modelling, to avoid biases due to different relative abundances and sampling effort (e.g., Fig. 5 in Ramos et al. 2013).

1.4.3 The study of habitat use

The study of habitat use in seabirds has also evolved since the first qualitative studies simply plotting seabird locations over maps of environmental variables (Cherel and Weimerskirch 1995). Such pioneering studies divided environmental variables into categories and used hypothesis testing models (e.g. Mann-Whitney U-tests in Waugh et al. 1999) to demonstrate environmental preferences of individuals or populations. This approach rather oversimplifies the complex relationships of birds with their environment. The increasing quality and availability of both environmental and tracking data, as well as the improvement in computational power boosted the development of more complex models where space was discretised in regular cells and a value of one or zero corresponding to presence or absence of tracking data was assigned to each cell. Then, a Bernoulli distribution could be assigned to this data and used to model it in relation to values of environmental variables at each cell. If abundance information for each cell was available, instead of presence/absence, a Poisson distribution could be

used. In addition, the use of Generalised Linear Models allowed to make predictions in unsampled areas if the values of the environmental variables in those areas were known. Soon, generalised additive models were used to account for non-linear relationships between seabird abundance and environmental variables (Wakefield et al. 2012). From there, increasingly complex methods have been incorporated to model the relationship between seabird abundance/distribution and the environment, including approximations borrowed from the information theory framework (maximum entropy; Phillips et al. 2006, Elith et al. 2011) or machine learning algorithms (Humphries et al. 2018).

When choosing from the wide array of species distribution modelling techniques, it is important to consider what the most appropriate method is for the data we are modelling (Oppel et al. 2012; Quillfeldt et al. 2017), or even take advantage of the increasing current computational power to produce ensemble models that average predictions of different algorithms minimising thus the biases introduced by each of them when used as stand-alone models (e.g. Pereira et al. 2018).

Despite the rapid evolution of the species distribution modelling methodology, some issues remain to be addressed. Most of these methodologies involve the aggregation of data (animal presence and environmental information) into pre-defined cells. This makes these models scale-dependent, since results can vary depending on the choice of cell size (Renner and Warton 2013). In addition, we often lose information, since we can only model processes occurring at the scale of our cell sizes or larger. Another important methodological issue is that, in commonly used species distribution models, animal presence or

abundance at each cell is often related to the corresponding values of the environmental variables without considering their potential spatial autocorrelation (i.e. “everything is related to everything else, but near things are more related than distant things”; Tobler 1970). Thus, although used to model spatial distribution of species, such models are not truly spatial.

Point pattern models are well known in statistical circles, but have been seldom used in the field of ecology because they are complex, computationally costly, and most of them have to be fitted in a Bayesian framework (Illian and Burslem 2017). However, they can model the distribution of points continuously, preserving the spatial correlation structure present in the data and avoiding the gridding of the data that inevitably causes loss of small-scale information. New statistical techniques developed for this type of models have made them easier to apply and less computationally costly, which make them now more approachable for ecologists (Illian and Burslem 2017). In fact, this type of truly spatial models have successfully been applied to Eulerian data (Paradinas et al. 2015; Soriano-Redondo et al. 2019), and could eventually be adapted to model Lagrangian data.

1.5 Uses of tracking data for conservation

Tracking data can have effect on conservation at different levels, from pure scientific pursuits to short-time applied studies; first, through basic research which can add to the knowledge on species movements and behaviour, and in turn may have long term applications in conservation; second, by highlighting the transnationality of animal movement and thus the need for international agreements for management; third,

through the discovery of migratory pathways or abundance hotspots that might be of conservation concern and have an effect on population dynamics; and lastly, by providing information relevant to dynamic management, such as temporal closures or protection against bycatch in marine species (McGowan et al. 2017).

Particularly in the marine environment, the main driver for developing tracking studies is the collection of distribution data otherwise difficult to obtain due the impracticality of survey techniques. Year-round tracking data has facilitated the study of migratory connectivity, providing insights into the drivers of population declines caused by non-lethal carry over effects that impact the breeding success of the subsequent season (e.g. Bogdanova et al. 2017), or lethal anthropogenic impacts in the non-breeding distributions and migratory flyways (e.g. Sherley et al. 2017)

1.6 Aims

The main aims of this dissertation are to provide ecological insights into the at-sea distribution of marine top-predators year-round at a multi-colony scale, and to develop methodological tools for the inference of animal distributions of populations and species from individual tracking data. Within this general aim, we pursued two specific objectives:

- (1) To address the different sources of variability introduced by the use of individual tracking data in the inference of space and habitat use at a colony, multi-colony, population or species level
 - Providing tools to detect the existence of biases and variability introduced by different sampling efforts between individuals

and populations, and by temporal variability in the distributions of mobile species inferred from individual tracking data.

- Demonstrating the applicability of spatially explicit, continuously indexed, point pattern models to individual tracking data
- (2) To study the space and habitat use of pelagic seabirds at a multi-colony level, its drivers, and implications
- during the non-breeding season, for three closely related taxa of pelagic seabirds, tracked from several colonies across the entire breeding distribution, to ultimately understand their spatial and habitat segregation, as well as their migratory connectivity at colony, population and species levels,
 - during the breeding season, for three neighbouring colonies, to detect among-colony segregation and to ultimately understand the environmental and behavioural drivers of this segregation.
- (3) To provide tools that confer robustness to studies using individual tracking data to define conservation and management strategies of mobile species.

In the **first chapter**, we discuss the three main sources of variability in multi-colony tracking studies and provide tools to test the effect of these variabilities in the spatial distributions at a population level. we then test these tools in a multi-colony, multi-species dataset of year-round locations of three pelagic species of seabirds.

In the **second chapter**, we use this multi-species dataset to study their non-breeding distribution from a multi-colony point of view, providing

evidences of migratory connectivity and among-species ecological segregation.

In the **third chapter**, we apply, for the first time, a point pattern model within a Bayesian framework to a set of animal tracking data that provides a continuous-space approach, avoiding the need to aggregate the data in cells and, thus, the loss of information that comes with it. With this method, we detect spatial structure within the foraging distributions of birds breeding in three neighbouring colonies, providing robust insights into the ecological and behavioural drivers of these distributions.

1.7 References

- Agee LA, Monfils M-H. 2018. Effect of demonstrator reliability and recency of last demonstration on acquisition of a socially transmitted food preference. *R Soc Open Sci.* 5(6):172391. doi:10.1098/rsos.172391. <https://royalsocietypublishing.org/doi/10.1098/rsos.172391>.
- Aguilar Soto N, Johnson MP, Madsen PT, Díaz F, Domínguez I, Brito A, Tyack P. 2008. Cheetahs of the deep sea: deep foraging sprints in short-finned pilot whales off Tenerife (Canary Islands). *J Anim Ecol.* 77(5):936–947. doi:10.1111/j.1365-2656.2008.01393.x. <http://doi.wiley.com/10.1111/j.1365-2656.2008.01393.x>.
- Alerstam T, Bäckman J, Grönroos J, Olofsson P, Strandberg R. 2019. Hypotheses and tracking results about the longest migration: The case of the arctic tern. *Ecol Evol.*(June):ece3.5459. doi:10.1002/ece3.5459. <https://onlinelibrary.wiley.com/doi/abs/10.1002/ece3.5459>.
- Alerstam T, Hedenström A, Åkesson S. 2003. Long-distance migration: evolution and determinants. *Oikos.* 103:247–260.
- Ashmole NP. 1963. The regulation of numbers of tropical oceanic birds. *Ibis* (Lond 1859). 103b(3):458–473. doi:10.1111/j.1474-919X.1963.tb06766.x. <http://doi.wiley.com/10.1111/j.1474-919X.1963.tb06766.x>.

- Avgar T, Mosser A, Brown GS, Fryxell JM. 2013. Environmental and individual drivers of animal movement patterns across a wide geographical gradient. Mysterud A, editor. *J Anim Ecol.* 82(1):96–106. doi:10.1111/j.1365-2656.2012.02035.x. <http://doi.wiley.com/10.1111/j.1365-2656.2012.02035.x>.
- Bastille-Rousseau G, Gibbs JP, Yackulic CB, Frair JL, Cabrera F, Rousseau L-P, Wikelski M, Kümmereth F, Blake S. 2017. Animal movement in the absence of predation: environmental drivers of movement strategies in a partial migration system. *Oikos.* 126(7):1004–1019. doi:10.1111/oik.03928. <http://doi.wiley.com/10.1111/oik.03928>.
- Bauer S, Lisovski S, Hahn S. 2016. Timing is crucial for consequences of migratory connectivity. *Oikos.* 125(5):605–612. doi:10.1111/oik.02706. <http://doi.wiley.com/10.1111/oik.02706>.
- Bogdanova MI, Butler A, Wanless S, Moe B, Anker-Nilssen T, Frederiksen M, Boulinier T, Chivers LLS, Christensen-dalsgaard S, Descamps S, et al. 2017. Multi-colony tracking reveals spatio-temporal variation in carry-over effects between breeding success and winter movements in a pelagic seabird. *Mar Ecol Prog Ser.* 578:167–181. doi:doi.org/10.3354/meps12096. [accessed 2018 Apr 10]. <http://www.int-res.com/abstracts/meps/v578/p167-181/>.
- Bolton M, Conolly G, Carroll M, Wakefield ED, Caldow R. 2018. A review of the occurrence of inter-colony segregation of seabird foraging areas and the implications for marine environmental impact assessment. *Ibis (Lond 1859).* doi:10.1111/ibi.12677. <http://doi.wiley.com/10.1111/ibi.12677>.
- 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. *J Anim Ecol.* 75(6):1393–1405. doi:10.1111/j.1365-2656.2006.01164.x.
- Boyd C, Grünbaum D, Hunt GL, Punt ADSE, Weimerskirch H, Bertrand S. 2016. Effectiveness of social information used by seabirds searching for unpredictable and ephemeral prey. *Behav Ecol.* 27(4):1223–1234. doi:10.1093/beheco/arw039.
- Buckley NJ. 1997. Spatial-Concentration Effects and the Importance of Local Enhancement in the Evolution of Colonial Breeding in Seabirds. *Am*

- Nat. 149(6):1091–1112. doi:10.1086/286040.
<https://www.journals.uchicago.edu/doi/10.1086/286040>.
- Burger AE. 1997. Arrival and departure behavior of common murrelets at colonies: Evidence for an information halo? *Waterbirds*. 20(1):55–65.
- Burger AE, Shaffer SA. 2008. Application of tracking and data-logging technology in research and conservation of seabirds. *Auk*. 125(2):253–264.
- Burt WH. 1943. Territoriality and Home Range Concepts as Applied to Mammals. *J Mammal*. 24(3):346. doi:10.2307/1374834.
<https://academic.oup.com/jmammal/article-lookup/doi/10.2307/1374834>.
- Cairns DK. 1989. The regulation of seabird colony size: a hinterland model. *Am Nat*. 134(1):141–146.
- Camphuysen KCJ, Van der Meer J. 2001. Pelagic distribution, moult and (sub-) specific status of Cory's shearwaters *Calonectris [d.] diomedea/borealis* wintering off southern Africa. *Mar Ornithol*. 29:89–96.
- Chapman BB, Brönmark C, Nilsson J-Å, Hansson LA. 2011. The ecology and evolution of partial migration. *Oikos*. 120(12):1764–1775. doi:10.1111/j.1600-0706.2011.20131.x.
<http://doi.wiley.com/10.1111/j.1600-0706.2011.20131.x>.
- Chapman BB, Hulthén K, Blomqvist DR, Hansson LA, Nilsson J-Å, Brodersen J, Anders Nilsson P, Skov C, Brönmark C. 2011. To boldly go: individual differences in boldness influence migratory tendency. *Ecol Lett*. 14(9):871–876. doi:10.1111/j.1461-0248.2011.01648.x.
<http://doi.wiley.com/10.1111/j.1461-0248.2011.01648.x>.
- Chapman JW, Reynolds DR, Wilson K. 2015. Long-range seasonal migration in insects: Mechanisms, evolutionary drivers and ecological consequences. *Ecol Lett*. 18(3):287–302. doi:10.1111/ele.12407.
- Cherel Y, Weimerskirch H. 1995. Seabirds as indicators of marine resources:black-browed albatrosses feeding on ommastrephid squids in Kerguelen waters. *Mar Ecol Prog Ser*. 129:295–300. doi:10.3354/meps129295. <http://www.int-res.com/abstracts/meps/v129/p295-300/>.

- Cristol DA, Baker MB, Carbone C. 1999. Differential migration revisited: latitudinal segregation by age and sex class. In: Nolan V, Ketterson ED, Thompson CF, editors. *Current Ornithology*. New York: Kluwer Academic/Plenum Publishers. p. 33–88.
- Danchin E. 2004. Public Information: From Nosy Neighbors to Cultural Evolution. *Science* (80-). 305(5683):487–491. doi:10.1126/science.1098254. <http://www.sciencemag.org/cgi/doi/10.1126/science.1098254>.
- Darwin C. 1861. *On the Origin of Species by Means of Natural Selection*. 3rd ed. London: John Murray.
- Desprez M, Jenouvrier S, Barbraud C, Delord K, Weimerskirch H. 2018. Linking oceanographic conditions, migratory schedules and foraging behaviour during the non-breeding season to reproductive performance in a long-lived seabird. *Funct Ecol*.(March 2017):1–14. doi:10.1111/1365-2435.13117. <http://doi.wiley.com/10.1111/1365-2435.13117>.
- Dias MP, Granadeiro JP, Phillips RA, Alonso H, Catry P. 2011. Breaking the routine: individual Cory’s shearwaters shift winter destinations between hemispheres and across ocean basins. *Proc R Soc B*. 278(1713):1786–1793. doi:10.1098/rspb.2010.2114. <http://rspb.royalsocietypublishing.org/cgi/doi/10.1098/rspb.2010.2114>.
- Dingle H. 1996. *Migration: the biology of life on the move*. New York: Oxford University Press.
- Dingle H, Drake VA. 2007. What Is Migration? *Bioscience*. 57(2):113–121. doi:10.1641/B570206. <http://academic.oup.com/bioscience/article/57/2/113/228325/What-Is-Migration>.
- Diop N, Zango L, Beard A, Ba CT, Ndiaye PI, Henry L, Clingham E, Opper S, González-Solís J. 2018. Foraging ecology of tropicbirds breeding in two contrasting marine environments in the tropical Atlantic. *Mar Ecol Prog Ser*. 607:221–236. doi:10.3354/meps12774.
- Egevang C, Stenhouse IJ, Phillips RA, Petersen A, Fox JW, Silk JRD. 2010. Tracking of Arctic terns *Sterna paradisaea* reveals longest animal migration. *Proc Natl Acad Sci*. 107(5):2078–2081.

- Elith J, Phillips SJ, Hastie T, Dudík M, Chee YE, Yates CJ. 2011. A statistical explanation of MaxEnt for ecologists. *Divers Distrib.* 17(1):43–57. doi:10.1111/j.1472-4642.2010.00725.x.
- Fagan WF, Lewis MA, Auger-Méthé M, Avgar T, Benhamou S, Breed G, LaDage L, Schlägel UE, Tang W, Papastamatiou YP, et al. 2013. Spatial memory and animal movement. Clobert J, editor. *Ecol Lett.* 16(10):1316–1329. doi:10.1111/ele.12165. [accessed 2019 Aug 18]. <http://doi.wiley.com/10.1111/ele.12165>.
- Finch T, Butler SJ, Franco AMA, Cresswell W. 2017. Low migratory connectivity is common in long-distance migrant birds. Chapman J, editor. *J Anim Ecol.* 86(3):662–673. doi:10.1111/1365-2656.12635. <http://doi.wiley.com/10.1111/1365-2656.12635>.
- Fort J, Pettex E, Tremblay Y, Lorentsen SH, Garthe S, Votier SC, Pons J-B, Siorat F, Furness RW, Grecian WJ, et al. 2012. Meta-population evidence of oriented chain migration in northern gannets (*Morus bassanus*). *Front Ecol Environ.* 10(5):237–242. doi:10.1890/110194. <http://www.esajournals.org/doi/abs/10.1890/110194>.
- Fraser KC, Davies KTA, Davy CM, Ford AT, Flockhart DTT, Martins EG. 2018. Tracking the Conservation Promise of Movement Ecology. *Front Ecol Evol.* 6:150. doi:10.3389/fevo.2018.00150. [accessed 2018 Oct 1]. <https://www.frontiersin.org/article/10.3389/fevo.2018.00150/full>.
- Furness RW, Birkhead TR. 1984. Seabird colony distributions suggest competition for food supplies during the breeding season. *Nature.* 311(5987):655–656. doi:10.1038/311655a0. <http://www.nature.com/articles/311655a0>.
- González-Solís J, Croxall JP, Oro D, Ruiz X. 2007. Trans-equatorial migration and mixing in the wintering areas of a pelagic seabird. *Front Ecol Environ.* 5(6):297–301. doi:10.1890/1540-9295(2007)5[297:TMAMIT]2.0.CO;2. [accessed 2017 Jul 8]. [http://doi.wiley.com/10.1890/1540-9295\(2007\)5\[297:TMAMIT\]2.0.CO;2](http://doi.wiley.com/10.1890/1540-9295(2007)5[297:TMAMIT]2.0.CO;2).
- González-Solís J, Croxall JP, Wood AG. 2000. Sexual dimorphism and sexual segregation in foraging strategies of northern giant petrels *Macronectes halli* during the incubation period. *Oikos.* 90(2):390–398.

- Grecian WJ, Lane J V, Michelot T, Wade HM, Hamer KC. 2018. Understanding the ontogeny of foraging behaviour: insights from combining marine predator bio-logging with satellite-derived oceanography in hidden Markov models. *J R Soc Interface*. 15(143):20180084. doi:10.1098/rsif.2018.0084. [accessed 2018 Jun 25]. <http://www.ncbi.nlm.nih.gov/pubmed/29875281>.
- Grémillet D, Dell’Omo G, Ryan PG, Peters G, Ropert-Coudert Y, Weeks SJ. 2004. Offshore diplomacy, or how seabirds mitigate intra-specific competition: a case study based on GPS tracking of Cape gannets from neighbouring colonies. *Mar Ecol Prog Ser*. 268:265–279. doi:10.3354/meps268265.
- Guzman LM, Germain RM, Forbes C, Straus S, O’Connor MI, Gravel D, Srivastava DS, Thompson PL. 2019. Towards a multi-trophic extension of metacommunity ecology. Brose U, editor. *Ecol Lett*. 22(1):19–33. doi:10.1111/ele.13162. <https://onlinelibrary.wiley.com/doi/abs/10.1111/ele.13162>.
- Hamer KC, Schreiber EA, Burger J. 2002. Breeding Biology, Life Histories and Life History-Environment Interactions in Seabirds. In: Schreiber EA, Burger J, editors. *Biology of Marine Birds*. Boca Raton, Florida: CRC Press. p. 217–261.
- Hansen MJ, Ward AJW, Fürtbauer I, King AJ. 2016. Environmental quality determines finder-joiner dynamics in socially foraging three-spined sticklebacks (*Gasterosteus aculeatus*). *Behav Ecol Sociobiol*. 70(6):889–899. doi:10.1007/s00265-016-2111-5. <http://dx.doi.org/10.1007/s00265-016-2111-5>.
- Hanski I, Gaggiotti OE, editors. 2004. *Ecology, Genetics, and Evolution of Metapopulations*. Burlington, MA: Elsevier Academic Press.
- Harrison XA, Blount JD, Inger R, Norris DR, Bearhop S. 2011. Carry-over effects as drivers of fitness differences in animals. *J Anim Ecol*. 80(1):4–18. doi:10.1111/j.1365-2656.2010.01740.x. <http://doi.wiley.com/10.1111/j.1365-2656.2010.01740.x>.
- Hayashi M, Bakkali M, Hyde A, Goodacre SL. 2015. Sail or sink: novel behavioural adaptations on water in aerially dispersing species. *BMC Evol Biol*. 15(1):118. doi:10.1186/s12862-015-0402-5. <http://dx.doi.org/10.1186/s12862-015-0402-5>.

- Hebblewhite M, Haydon DT. 2010. Distinguishing technology from biology: a critical review of the use of GPS telemetry data in ecology. *Philos Trans R Soc B Biol Sci.* 365(1550):2303–2312. doi:10.1098/rstb.2010.0087. [accessed 2018 Mar 7]. <http://www.ncbi.nlm.nih.gov/pubmed/20566506>.
- Hedd A, Pollet IL, Mauck RA, Burke CM, Mallory ML, McFarlane Tranquilla LA, Montevecchi WA, Robertson GJ, Ronconi RA, Shutler D, et al. 2018. Foraging areas, offshore habitat use, and colony overlap by incubating Leach’s storm-petrels *Oceanodroma leucorhoa* in the Northwest Atlantic. Mettke-Hofmann C, editor. *PLoS One.* 13(5):e0194389. doi:10.1371/journal.pone.0194389. [accessed 2018 May 21]. <http://dx.plos.org/10.1371/journal.pone.0194389>.
- Holdo RM, Roach RR. 2013. Inferring animal population distributions from individual tracking data: Theoretical insights and potential pitfalls. *J Anim Ecol.* 82(1):175–181. doi:10.1111/j.1365-2656.2012.02031.x.
- Holyoak M, Casagrandi R, Nathan R, Revilla E, Spiegel O. 2008. Trends and missing parts in the study of movement ecology. *Proc Natl Acad Sci.* 105(49):19060–19065. doi:10.1073/pnas.0800483105. [accessed 2018 Mar 7]. <http://www.ncbi.nlm.nih.gov/pubmed/19060194>.
- Hooten MB, Johnson DS, McClintock BT, Morales JM. 2017. *Animal Movement*. CRC Press Taylor & Francis Group 6000 Broken Sound Parkway NW, Suite 300 Boca Raton, FL 33487-2742: CRC Press. <https://www.taylorfrancis.com/books/9781466582156>.
- Humphries G, Magness DR, Huettmann F, editors. 2018. *Machine Learning for Ecology and Sustainable Natural Resource Management*. Cham: Springer International Publishing. <http://link.springer.com/10.1007/978-3-319-96978-7>.
- Illian JB, Burslem DFRP. 2017. Improving the usability of spatial point process methodology: an interdisciplinary dialogue between statistics and ecology. *AStA Adv Stat Anal.* 101(4):1–26. doi:10.1007/s10182-017-0301-8.
- Jetz W. 2004. The Scaling of Animal Space Use. *Science (80-)*. 306(5694):266–268. doi:10.1126/science.1102138. <http://www.sciencemag.org/cgi/doi/10.1126/science.1102138>.

- Jones TB, Patrick SC, Arnould JPY, Rodríguez-Malagón MA, Wells MR, Green JA. 2018. Evidence of sociality in the timing and location of foraging in a colonial seabird. *Biol Lett.* 14(7):20180214. doi:10.1098/rsbl.2018.0214. <http://rsbl.royalsocietypublishing.org/lookup/doi/10.1098/rsbl.2018.0214>.
- Jovani R, Lascelles BG, Garamszegi LZ, Mavor R, Thaxter CB, Oro D. 2015. Colony size and foraging range in seabirds. *Oikos*.(September 2015):n/a-n/a. doi:10.1111/oik.02781. [accessed 2015 Oct 26]. <http://doi.wiley.com/10.1111/oik.02781>.
- Kiester AR. 1979. Conspecifics as cues: A mechanism for habitat selection in the Panamanian grass anole (*Anolis aeneus*). *Behav Ecol Sociobiol.* 5(4):323–330. doi:10.1007/BF00292522. <http://link.springer.com/10.1007/BF00292522>.
- Kranstauber B. 2019. Modelling animal movement as Brownian bridges with covariates. *Mov Ecol.* 7(1):22. doi:10.1186/s40462-019-0167-3. <https://movementecologyjournal.biomedcentral.com/articles/10.1186/s40462-019-0167-3>.
- Lafontaine A, Drapeau P, Fortin D, St-Laurent MH. 2017. Many places called home: the adaptive value of seasonal adjustments in range fidelity. *J Anim Ecol.* 86(3):624–633. doi:10.1111/1365-2656.12645.
- Lascelles BG, Taylor PR, Miller MGR, Dias MP, Opper S, Torres L, Hedd A, Le Corre M, Phillips RA, Shaffer SA, et al. 2016. Applying global criteria to tracking data to define important areas for marine conservation. Visconti P, editor. *Divers Distrib.* 22(4):422–431. doi:10.1111/ddi.12411. [accessed 2017 Apr 5]. <http://onlinelibrary.wiley.com/doi/10.1111/ddi.12411/full>.
- Lewison RL, Crowder LB, Read AJ, Freeman SA. 2004. Understanding impacts of fisheries bycatch on marine megafauna. *Trends Ecol Evol.* 19(11):598–604. doi:10.1016/j.tree.2004.09.004.
- Ludynia K, Kemper J, Roux JP. 2012. The Namibian Islands' Marine Protected Area: Using seabird tracking data to define boundaries and assess their adequacy. *Biol Conserv.* 156:136–145. doi:10.1016/j.biocon.2011.11.014. <http://dx.doi.org/10.1016/j.biocon.2011.11.014>.

- Lundberg J, Moberg F. 2003. Mobile Link Organisms and Ecosystem Functioning: Implications for Ecosystem Resilience and Management. *Ecosystems*. 6(1):0087–0098. doi:10.1007/s10021-002-0150-4. <http://link.springer.com/10.1007/s10021-002-0150-4>.
- Machovsky-Capuska GE, Hauber ME, Libby E, Amiot C, Raubenheimer D. 2014. The contribution of private and public information in foraging by Australasian gannets. *Anim Cogn*. 17(4):849–858. doi:10.1007/s10071-013-0716-x. [accessed 2019 Jul 27]. <http://link.springer.com/10.1007/s10071-013-0716-x>.
- Marra PP, Cohen E, Harrison A-L, Studds CE, Webster M. 2019. Migratory Connectivity. In: Crooks KR, Sanjayan MA, editors. *Encyclopedia of Animal Behavior*. New York: Elsevier. p. 643–654. <https://linkinghub.elsevier.com/retrieve/pii/B9780128096338207619>.
- Matthiopoulos J. 2003. The use of space by animals as a function of accessibility and preference. *Ecol Modell*. 159(2–3):239–268. doi:10.1016/S0304-3800(02)00293-4. <https://linkinghub.elsevier.com/retrieve/pii/S0304380002002934>.
- McGowan J, Beger M, Lewison RL, Harcourt RG, Campbell HA, Priest M, Dwyer RG, Lin H-Y, Lentini P, Dudgeon C, et al. 2017. Integrating research using animal-borne telemetry with the needs of conservation management. du Toit J, editor. *J Appl Ecol*. 54(2):423–429. doi:10.1111/1365-2664.12755. [accessed 2017 May 23]. <http://doi.wiley.com/10.1111/1365-2664.12755>.
- Morales JM, Moorcroft PR, Matthiopoulos J, Frair JL, Kie JG, Powell RA, Merrill EH, Haydon DT. 2010. Building the bridge between animal movement and population dynamics. *Philos Trans R Soc B Biol Sci*. 365(1550):2289–2301. doi:10.1098/rstb.2010.0082. <https://royalsocietypublishing.org/doi/10.1098/rstb.2010.0082>.
- Mueller T, Lenz J, Caprano T, Fiedler W, Böhning-Gaese K. 2014. Large frugivorous birds facilitate functional connectivity of fragmented landscapes. Elphick C, editor. *J Appl Ecol*. 51(3):684–692. doi:10.1111/1365-2664.12247. <http://doi.wiley.com/10.1111/1365-2664.12247>.
- Müller M, Wehner R. 2010. Path Integration Provides a Scaffold for Landmark Learning in Desert Ants. *Curr Biol*. 20(15):1368–1371.

doi:10.1016/j.cub.2010.06.035.

<https://linkinghub.elsevier.com/retrieve/pii/S0960982210007827>.

O’Corry-Crowe G, Suydam R, Quakenbush L, Potgieter B, Harwood L, Litovka D, Ferrer T, Citta J, Burkanov V, Frost K, et al. 2018. Migratory culture, population structure and stock identity in North Pacific beluga whales (*Delphinapterus leucas*). Li S, editor. PLoS One. 13(3):e0194201. doi:10.1371/journal.pone.0194201.

<https://dx.plos.org/10.1371/journal.pone.0194201>.

Oppel S, Beard A, Fox D, Mackley E, Leat E, Henry L, Clingham E, Fowler N, Sim J, Sommerfeld J, et al. 2015. Foraging distribution of a tropical seabird supports Ashmole’s hypothesis of population regulation. Behav Ecol Sociobiol. doi:10.1007/s00265-015-1903-3.

<http://link.springer.com/10.1007/s00265-015-1903-3>.

Oppel S, Meirinho A, Ramírez I, Gardner B, O’Connell AF, Miller PI, Louzao M. 2012. Comparison of five modelling techniques to predict the spatial distribution and abundance of seabirds. Biol Conserv. 156:94–104. doi:10.1016/j.biocon.2011.11.013.

<http://dx.doi.org/10.1016/j.biocon.2011.11.013>.

Orians GH, Pearson NE. 1979. On the theory of central place foraging. In: Horn DJ, Mitchell RD, Stairs GR, editors. Analysis of ecological systems. Columbus: Ohio State University Press. p. 155–177.

Paiva VH, Geraldes P, Marques V, Rodríguez R, Garthe S, Ramos JA. 2013. Effects of environmental variability on different trophic levels of the North Atlantic food web. Mar Ecol Prog Ser. 477(August):15–28. doi:10.3354/meps10180.

Palacín C, Alonso JC, Alonso JA, Magaña M, Martín CA. 2011. Cultural transmission and flexibility of partial migration patterns in a long-lived bird, the great bustard *Otis tarda*. J Avian Biol. 42(4):301–308. doi:10.1111/j.1600-048X.2011.05395.x.

<http://doi.wiley.com/10.1111/j.1600-048X.2011.05395.x>.

Paradinas I, Conesa D, Pennino MG, Muñoz F, Fernández AM, López-Quílez A, Bellido JM. 2015. Bayesian spatio-temporal approach to identifying fish nurseries by validating persistence areas. Mar Ecol Prog Ser. 528:245–255. doi:10.3354/meps11281. <http://www.int-res.com/abstracts/meps/v528/p245-255/>.

- Parker KL, Barboza PS, Gillingham MP. 2009. Nutrition integrates environmental responses of ungulates. *Funct Ecol.* 23(1):57–69. doi:10.1111/j.1365-2435.2009.01528.x. <http://doi.wiley.com/10.1111/j.1365-2435.2009.01528.x>.
- Patrick SC, Pinaud D, Weimerskirch H. 2017. Boldness predicts an individual's position along an exploration-exploitation foraging trade-off. Morand-Ferron J, editor. *J Anim Ecol.* 86(5):1257–1268. doi:10.1111/1365-2656.12724. [accessed 2017 Aug 6]. <http://doi.wiley.com/10.1111/1365-2656.12724>.
- Patterson TA, Parton A, Langrock R, Blackwell PG, Thomas L, King R. 2017. Statistical modelling of individual animal movement: an overview of key methods and a discussion of practical challenges. *ASTA Adv Stat Anal.*:1–40. doi:10.1007/s10182-017-0302-7.
- Pereira JM, Krüger L, Oliveira N, Meirinho A, Silva A, Ramos JA, Paiva VH. 2018. Using a multi-model ensemble forecasting approach to identify key marine protected areas for seabirds in the Portuguese coast. *Ocean Coast Manag.* 153(July 2017):98–107. doi:10.1016/j.ocecoaman.2017.12.014. <https://doi.org/10.1016/j.ocecoaman.2017.12.014>.
- Phillips EM, Horne JK, Zamon JE, Felis JJ, Adams J. 2019. Does perspective matter? A case study comparing Eulerian and Lagrangian estimates of common murre (*Uria aalge*) distributions. *Ecol Evol.*(November 2018):ece3.5083. doi:10.1002/ece3.5083. <https://onlinelibrary.wiley.com/doi/abs/10.1002/ece3.5083>.
- Phillips SJ, Anderson RP, Schapire RE. 2006. Maximum entropy modeling of species geographic distributions. *Ecol Modell.* 190:231–259. doi:10.1016/j.ecolmodel.2005.03.026.
- Pinsky M, McCauley D. 2019. Home range and body size data compiled from the literature for marine and terrestrial vertebrates. doi:10.1575/1912/bco-dmo.752795.1. <https://hdl.handle.net/1912/23932>.
- Ponchon A, Garnier R, Grémillet D, Boulinier T. 2015. Predicting population responses to environmental change: the importance of considering informed dispersal strategies in spatially structured population models. Heikkinen R, editor. *Divers Distrib.* 21(1):88–100.

doi:10.1111/ddi.12273. <http://doi.wiley.com/10.1111/ddi.12273>.

- Potier S, Duriez O, Célérier A, Liegeois JL, Bonadonna F. 2019. Sight or smell: which senses do scavenging raptors use to find food? *Anim Cogn.* 22(1):49–59. doi:10.1007/s10071-018-1220-0. <http://dx.doi.org/10.1007/s10071-018-1220-0>.
- Pusey AE. 1987. Sex-biased dispersal and inbreeding avoidance in birds and mammals. *Trends Ecol Evol.* 2:295–299.
- Quillfeldt P, Engler JO, Silk JRDD, Phillips RA. 2017. Influence of device accuracy and choice of algorithm for species distribution modelling of seabirds: a case study using black-browed albatrosses. *J Avian Biol.* 1990(December 2016):1–7. doi:10.1111/jav.01238. <http://doi.wiley.com/10.1111/jav.01238>.
- Ramos R, Carlile N, Madeiros J, Ramírez I, Paiva VH, Dinis HA, Zino F, Biscoito M, Leal GR, Bugoni L, et al. 2017. It is the time for oceanic seabirds: Tracking year-round distribution of gadfly petrels across the Atlantic Ocean. *Divers Distrib.* 23:794–805.
- Ramos R, Granadeiro JP, Rodríguez B, Navarro J, Paiva VH, Bécares J, Reyes-González JM, Fagundes I, Ruiz A, Arcos JM, et al. 2013. Meta-population feeding grounds of Cory's shearwater in the subtropical Atlantic Ocean: implications for the definition of Marine Protected Areas based on tracking studies. *Divers Distrib.* 19(10):1284–1298.
- Ramos R, Morera-Pujol V, Cruz-Flores M, López-Souto S, Brothers M, González-Solís J. 2019. A geolocator-tagged fledgling provides first evidence on juvenile movements of Cory's Shearwater *Calonectris borealis*. *Bird Study.* 66(2):283–288. doi:10.1080/00063657.2019.1638341. <https://doi.org/10.1080/00063657.2019.1638341>.
- Ramos R, Sanz V, Militão T, Bried J, Neves VC, Biscoito M, Phillips RA, Zino F, González-Solís J. 2015. Leapfrog migration and habitat preferences of a small oceanic seabird, Bulwer's petrel (*Bulweria bulwerii*). *J Biogeogr.* 42(9):1651–1664. doi:10.1111/jbi.12541. <http://doi.wiley.com/10.1111/jbi.12541>.
- Renner IW, Warton DI. 2013. Equivalence of MAXENT and Poisson Point Process Models for Species Distribution Modeling in Ecology.

- Biometrics. 69(1):274–281. doi:10.1111/j.1541-0420.2012.01824.x.
- Riotte-Lambert L, Weimerskirch H. 2013. Do naive juvenile seabirds forage differently from adults? *Proc R Soc B*. 280(1768):20131434. doi:10.1098/rspb.2013.1434. [accessed 2017 Oct 23]. <http://www.ncbi.nlm.nih.gov/pubmed/23926153>.
- Robinson R, Crick H, Learmonth J, Maclean I, Thomas C, Bairlein F, Forchhammer M, Francis C, Gill J, Godley B, et al. 2009. Travelling through a warming world: climate change and migratory species. *Endanger Species Res*. 7(2):87–99. doi:10.3354/esr00095. <http://www.int-res.com/abstracts/esr/v7/n2/p87-99/>.
- Ronce O. 2007. How Does It Feel to Be Like a Rolling Stone? Ten Questions About Dispersal Evolution. *Annu Rev Ecol Evol Syst*. 38(1):231–253. doi:10.1146/annurev.ecolsys.38.091206.095611.
- Sachs JL, Hughes CR, Nuechterlein GL, Buitron D. 2007. Evolution of coloniality in birds: a test of hypotheses with the red-necked grebe (*Podiceps grisegena*). *Auk*. 124(2):628. doi:10.1642/0004-8038(2007)124[628:eociba]2.0.co;2.
- Savage C. 2019. Seabird nutrients are assimilated by corals and enhance coral growth rates. *Sci Rep*. 9(1):1–10. doi:10.1038/s41598-019-41030-6. <http://dx.doi.org/10.1038/s41598-019-41030-6>.
- Shaffer SA, Tremblay Y, Weimerskirch H, Scott D, Thompson DR, Sagar PM, Moller H, Taylor GA, Foley DG, Block BA, et al. 2005. Migratory shearwaters integrate oceanic resources across the Pacific Ocean in an endless summer. *Proc Natl Acad Sci*. 8(37):240. doi:10.4324/9781315687315.
- Sherley RB, Ludynia K, Dyer BM, Lamont T, Makhado AB, Roux JP, Scales KL, Underhill LG, Votier SC. 2017. Metapopulation Tracking Juvenile Penguins Reveals an Ecosystem-wide Ecological Trap. *Curr Biol*. 27(4):563–568. doi:10.1016/j.cub.2016.12.054. <http://dx.doi.org/10.1016/j.cub.2016.12.054>.
- Soriano-Redondo A, Jones-Todd CM, Bearhop S, Hilton GM, Lock L, Stanbury A, Votier SC, Illian JB. 2019. Understanding species distribution in dynamic populations: a new approach using spatio-temporal point process models. *Ecography (Cop)*. 42(6):1092–1102.

doi:10.1111/ecog.03771.

<https://onlinelibrary.wiley.com/doi/abs/10.1111/ecog.03771>.

Spiegel O, Leu ST, Bull CM, Sih A. 2017. What's your move? Movement as a link between personality and spatial dynamics in animal populations. *Ecol Lett.* 20(1):3–18. doi:10.1111/ele.12708. <http://doi.wiley.com/10.1111/ele.12708>.

Thiebault A, Mullers R, Pistorius P, Meza-Torres MA, Dubroca L, Green DB, Tremblay Y. 2014. From colony to first patch: processes of prey searching and social information in Cape Gannets. *Auk.* 131(4):595–609. doi:10.1642/AUK-13-209.1. <http://www.bioone.org/doi/abs/10.1642/AUK-13-209.1>.

Thiebault A, Semeria M, Lett C, Tremblay Y. 2016. How to capture fish in a school? Effect of successive predator attacks on seabird feeding success. *J Anim Ecol.* 85(1):157–167. doi:10.1111/1365-2656.12455.

Thomson AL. 1926. Problems of bird-migration. London: H.F. & G. Witherby.

Tobler WR. 1970. A Computer Movie Simulating Urban Growth in the Detroit Region. *Econ Geogr.* 46:234. doi:10.2307/143141. <https://www.jstor.org/stable/143141?origin=crossref>.

Tracey J a., Sheppard J, Zhu J, Wei F, Swaisgood RR, Fisher RN. 2014. Movement-Based Estimation and Visualization of Space Use in 3D for Wildlife Ecology and Conservation. Sueur C, editor. *PLoS One.* 9(7):e101205. doi:10.1371/journal.pone.0101205. <http://dx.plos.org/10.1371/journal.pone.0101205>.

Trevaill AM, Green JA, Sharples J, Polton Jeffrey A., Arnould JPYJPY, Patrick SC. 2019. Environmental heterogeneity amplifies behavioural response to a temporal cycle. *Oikos.* 128(4):1–12. doi:10.1111/oik.05579. <https://doi.org/10.1016/j.marpol.2018.08.024>.

Trevaill AM, Green JA, Sharples J, Polton Jeff A., Miller PI, Daunt F, Owen E, Bolton M, Colhoun K, Newton S, et al. 2019. Environmental heterogeneity decreases reproductive success via effects on foraging behaviour. *Proc R Soc B Biol Sci.* 286(1904). doi:10.1098/rspb.2019.0795.

Tucker MA, Alexandrou O, Bierregaard RO, Bildstein KL, Böhning-Gaese K, Bracis C, Brzorad JN, Buechley ER, Cabot D, Calabrese JM, et al.

2019. Large birds travel farther in homogeneous environments. Boucher-Lalonde V, editor. *Glob Ecol Biogeogr.* 28(November 2018):576–587. doi:10.1111/geb.12875. <http://doi.wiley.com/10.1111/geb.12875>.
- Tucker MA, Böhning-Gaese K, Fagan WF, Fryxell JM, Van Moorter B, Alberts SC, Ali AH, Allen AM, Attias N, Avgar T, et al. 2018. Moving in the Anthropocene: Global reductions in terrestrial mammalian movements. *Science* (80-). 359(6374):466–469. doi:10.1126/science.aam9712.
- Tucker MA, Ord TJ, Rogers TL. 2014. Evolutionary predictors of mammalian home range size: body mass, diet and the environment. *Glob Ecol Biogeogr.* 23(10):1105–1114. doi:10.1111/geb.12194. <http://doi.wiley.com/10.1111/geb.12194>.
- Turchin P. 1998. Quantitative analysis of movement: measuring and modeling population redistribution in animals and plants. Sunderland: Sinauer Associates.
- Wakefield ED, Bodey TW, Bearhop S, Blackburn J, Colhoun K, Davies R, Dwyer RG, Green JA, Grémillet D, Jackson AL, et al. 2013. Space partitioning without territoriality in gannets. *Science* (80-). 341:68–70. <http://www.sciencemag.org/cgi/doi/10.1126/science.1236077>.
- Wakefield ED, Owen E, Baer J, Carroll MJ, Daunt F, Dodd SG, Green JA, Guilford T, Mavor RA, Miller PI, et al. 2017. Breeding density, fine-scale tracking, and large-scale modeling reveal the regional distribution of four seabird species. *Ecol Appl.* 27(7):2074–2091. doi:10.1002/eap.1591. <http://doi.wiley.com/10.1002/eap.1591>.
- Wakefield ED, Phillips RA, Belchier M. 2012. Foraging black-browed albatrosses target waters overlaying moraine banks -a consequence of upward benthic-pelagic coupling? *Antarct Sci.* 24(3):269–280. doi:10.1017/S0954102012000132. https://www.cambridge.org/core/product/identifier/S0954102012000132/type/journal_article.
- Walter WD, Onorato DP, Fischer JW. 2015. Is there a single best estimator? Selection of home range estimators using area-under-the-curve. *Mov Ecol.* 3(1):10. doi:10.1186/s40462-015-0039-4. ???
- Ward A, Webster M. 2016. Sociality: The Behaviour of Group-Living Animals.

Cham: Springer International Publishing.
<http://link.springer.com/10.1007/978-3-319-28585-6>.

- Ward P, Zahavi A. 1972. The importance of certain assemblages of birds as “information-centres” for food-finding. *Ibis* (Lond 1859). 115(4):517–534. doi:10.1111/j.1474-919X.1973.tb01990.x. <http://doi.wiley.com/10.1111/j.1474-919X.1973.tb01990.x>.
- Waugh SM, Weimerskirch H, Cherel Y, Shankar U, Prince PA, Sagar PM. 1999. Exploitation of the marine environment by two sympatric albatrosses in the Pacific Southern Ocean. *Mar Ecol Prog Ser*. 177:243–254.
- Wearmouth VJ, Sims DW. 2008. Sexual segregation in marine fish, reptiles, birds and mammals: behaviour patterns, mechanisms, and conservation implications. In: Sims D W, editor. *Advances in marine biology*. Academic Press. Elsevier. p. 107–170.
- Webster MS, Marra PP, Haig SM, Bensch S, Holmes RT. 2002. Links between worlds: unraveling migratory connectivity. *Trends Ecol Evol*. 17:76–83.
- Weimerskirch H. 2002. Seabird demography and its relationship with the marine environment. In: Schreiber EA, Burger J, editors. *Biology of marine birds*. CRC Press. p. 115–136.
- Weimerskirch H, Bertrand S, Silva J, Marques JC, Goya E. 2010. Use of social information in seabirds: Compass rafts indicate the heading of food patches. Gratwicke B, editor. *PLoS One*. 5(3):e9928. doi:10.1371/journal.pone.0009928. [accessed 2019 Jul 27]. <https://dx.plos.org/10.1371/journal.pone.0009928>.
- Weimerskirch H, Wilson RP. 2000. Oceanic respite for wandering albatrosses. *Nature*. 406:955–956.
- Westerdahl H, Bensch S, Nilsson J-Å, O’Connor E, Sehgal R, Tesson S, Hasselquist D. 2014. Pathogens and hosts on the move. In: Hansson L-A, Åkesson S, editors. *Animal Movement Across Scales*. Oxford University Press. p. 126–148. <http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199677184.001.0001/acprof-9780199677184-chapter-8>.
- Whitehead H. 2017. Gene–culture coevolution in whales and dolphins. *Proc Natl Acad Sci*. 114(30):7814–7821. doi:10.1073/pnas.1620736114. <http://www.pnas.org/lookup/doi/10.1073/pnas.1620736114>.

- Winger BM, Auteri GG, Pegan TM, Weeks BC. 2019. A long winter for the Red Queen: rethinking the evolution of seasonal migration. *Biol Rev.* 94(3):737–752. doi:10.1111/brv.12476.
- Winkle W Van. 1975. Comparison of Several Probabilistic Home-Range Models. *J Wildl Manage.* 39(1):118. doi:10.2307/3800474. <https://www.jstor.org/stable/3800474?origin=crossref>.
- Wolf JBW, Trillmich F. 2007. Beyond habitat requirements: individual fine-scale site fidelity in a colony of the Galapagos sea lion (*Zalophus wolfebaeki*) creates conditions for social structuring. *Oecologia.* 152(3):553–567. doi:10.1007/s00442-007-0665-7. <http://link.springer.com/10.1007/s00442-007-0665-7>.
- Wood AG, Naef-Daenzer B, Prince PA, Croxall JP. 2000. Quantifying habitat use in satellite-tracked pelagic seabirds: application of kernel estimation to albatross locations. *J Avian Biol.* 31:278–286.
- Yamamoto T, Takahashi A, Oka N, Shirai M, Yamamoto M, Katsumata N, Sato K, Watanabe S, Trathan PN. 2012. Inter-colony differences in the incubation pattern of streaked shearwaters in relation to the local marine environment. *Waterbi.* 35(2):248–259. doi:10.1675/063.035.0207. <http://www.waterbirds.org/journal>.
- Yoda K, Yamamoto T, Suzuki H, Matsumoto S, Müller M, Yamamoto M. 2017. Compass orientation drives naïve pelagic seabirds to cross mountain ranges. *Curr Biol.* 27(21):R1152–R1153. doi:10.1016/j.cub.2017.09.009. [accessed 2017 Nov 10]. <http://linkinghub.elsevier.com/retrieve/pii/S0960982217311739>.
- Young LC, Vanderlip C, Duffy DC, Afanasyev V, Shaffer S a. 2009. Bringing home the trash: Do colony-based differences in foraging distribution lead to increased plastic ingestion in Laysan Albatrosses. *PLoS One.* 4.

2

Detecting recurrent sources of variability in animal tracking studies

2.1 Abstract

Over the last decades, the study of animal movement through tracking data has grown exponentially, exceeding the expectations of researchers in the field of movement ecology. This has posed new challenges in the analysis process, specifically when inferring higher-level distributions (i.e. colony, population, species) from individual data. Sources of variability such as individual site fidelity, environmental stochasticity over time, and spatial variability in movement patterns must be considered, and their effects identified and corrected for, to produce accurate estimates of spatial distribution.

We developed a set of procedures to detect the effect of these three sources of variability in the distribution of groups of animals when inferred from individual tracking data. These procedures are applicable to any set of tracking data regardless of the species or tracking technique. We validated the applicability of the method on a data set containing 1,346 year-round migratory trips from 805 individuals of three closely related seabird species breeding in 34 different colonies in the Mediterranean Sea and in the Atlantic Ocean, sampled during a 10-year period.

Using our newly developed procedure we were able to demonstrate that there was no effect of individual site fidelity or environmental stochasticity on the at-sea distribution of birds for any of the three species we considered. We were also able to identify variability in the non-breeding distributions of birds from different colonies, with significant effects of the distance to the population's centre, and of the latitude or longitude on the colonies' representativeness at the species level.

This work provides a useful and much-needed tool for researchers using animal tracking data to model species distributions or establish conservation measures. Accounting for these sources of variability has become essential in the context of the globalisation of science, where collaborations and tracking data repositories are making the analysis of very large data sets increasingly common.

2.2 Introduction

Although movement is a widespread characteristic in the animal kingdom, the study of movement ecology had not flourished until recently, due to the scarcity of available data and difficulties in obtaining them (Dingle 1996). In the past decades, the advent of animal-borne tracking technology 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 (Ferrerias 2001), and even ecosystem-wide (Courbin et al. 2014) tracking studies. Such large data sets have outgrown the available analysis techniques and revealed sources of variability in the estimation of higher-level spatial distributions (i.e. colony, population, species) from individual movement data that are yet to be properly addressed (Gutowsky et al. 2015).

There are numerous sources of variability when scaling-up from individual tracks to the space use of higher-level groups, and we must understand their origins and effects. Among the most important are: the presence of individual site fidelity (Spiegel et al. 2017), the temporal variability in environmental conditions (Paiva et al. 2013), and the spatial variability in the use of space (Frederiksen et al. 2012).

Fidelity to a geographic area is a well-documented phenomenon, present in animal species from three phyla (Switzer 1993), and can be related to breeding or foraging behaviour, and to social interactions (Giuggioli and Bartumeus 2012). 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 (Schmidt 2004).

Although most studies of individual site fidelity refer to the individuals' return to the same breeding area (Baylis et al. 2015), site fidelity also occurs outside the breeding period (Robillard et al. 2018). Regardless, its effect must be considered when individual tracking data is used for inferring a population's distribution, as it can bias results towards areas preferred by the more well-represented animals (Giuggioli and Bartumeus 2012; Lascelles et al. 2016).

Temporal changes in environmental conditions 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 and Trillmich 2007). Herds of nomadic herbivores, for instance, move following peaks in productivity of grasslands (Aikens et al. 2017), and specific route characteristics and non-breeding areas of migratory birds can be linked to changing environmental conditions (Dias et al. 2011).

Failing to consider the entire breeding range of a species when it spans a heterogeneous environment can lead to underestimating the space use of the species. Individuals from different areas will be exposed to different environmental conditions or ecological pressures, and will thus exhibit a differential use of space. In fact, it has been shown that the size of the home ranges calculated using tracking data can be correlated to the size of the study area (Nekolny et al. 2017). Thus, maximising the extent of the sampled area as well as the number of animals tracked 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 non-overlapping distributions outside of the

breeding season, tracking individuals from only a few breeding populations could lead to an erroneous estimation of their non-breeding areas (Webster et al. 2002).

Marine top predators are a particularly useful group to study these sources of variability in tracking data since they show geographically widespread distributions and diverse movement and migratory patterns (Yurkowski et al. 2018). Among them, *Calonectris* shearwaters are medium-sized Procellariiformes that perform 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 Atlantic and Mediterranean coasts, and their non-breeding distributions are composed of discrete pelagic areas in the Atlantic and Indian Oceans (González-Solís et al., 2007; Dias et al., 2011). All of these characteristics make them a suitable study group to investigate the effects of the aforementioned sources of variability.

In this work, we aim to provide a set of tools to understand the effects that (a) individual site fidelity, (b) environmental variability and (c) the extent of sampling effort, have on the distributions of mobile species tracked using animal-borne devices, regardless of tracking method, habitat or characteristics of the species. To do so, we collated, for the first time, a data set of 1,346 year-round tracks from 805 individuals of three *Calonectris* shearwater species breeding in up to 34 colonies. This constitutes a robust and diverse data set and provides a relevant example to demonstrate the applicability of our method.

2.3 Methods

We describe the procedures to test the effects of individual, temporal and spatial variation in a tracking data set. Our method is applicable to any set 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. We also provide an example using the non-breeding distributions of *Calonectris* shearwaters tracked with Global Location Sensors (GLS; Wilson, Ducamp, Rees, Culik, & Niekamp, 1992).

2.3.1 *Testing for individual site fidelity*

To detect the bias caused by individuals preferring certain areas, we propose a method that consists of obtaining an estimate of space use for each trip using the Kernel Density Estimate method (KDE, Worton, 1989). Firstly, for every possible pair of trips regardless of the individual that performed them, it calculates their spatial overlap. We selected the Bhattacharyya affinity (Bhattacharyya 1943) since most ecologists are familiar with it, but the function allows the selection of any other method available in the `kerneloverlap()` function provided in the `adehabitathR` package in R (Calenge 2006). This produces a square matrix containing values of overlap of all pairwise combinations of trips. Secondly, from the resulting matrix, the method selects all values corresponding to the overlap of two trips from different individuals, and groups them into a vector containing “between individual” values. All the values corresponding to the overlap of trips from the same individual are grouped into a vector of “within individual” values. Thirdly, the two vectors are compared through a bootstrap version of the Kolmogorov-Smirnov (K-S) test (Abadie 2002), from the package `Matching` (Sekhon

2011), which is suitable for non-randomised samples and does not assume continuous distributions. The two vectors are indeed not independent, as the same trip that is part of a “within individual” overlap can be part of a “between individual” overlap. Moreover, the values are not continuous as they are restricted from 0 (absence of overlap) to 1 (full overlap) with many 0's in both vectors (many instances of no overlap), which would cause ties between them. As with the original K-S test, the bootstrap version does not assume a Gaussian distribution. Rejecting the null hypothesis means that “between individual” and “within individual” vectors are samples from different distributions, i.e. have different means, standard deviations, and overall shapes.

2.3.2 Testing for temporal variability

The aim of this test is not to understand why or how changes in distributions occur over time, but only to test for their existence. In the present example, we test variability among years, but the user can select the temporal unit. The existence of temporal variability can be detected with an approach that mimics that used to test the effect of individual site fidelity (see above): the aim is to check whether variability in individual space use within a year is the same as between years. The function is the same as above, but using *year* as a grouping variable; the “between group” values are overlaps of all pairs of trips from the different years, while the “within group” values are overlaps of all pairs of trips from the same year. In this way, the K-S tests whether “between year” overlaps have the same distribution as “within year” overlaps. Rejecting the null hypothesis means that the variability is different between compared to within years, proving the existence of a temporal effect in space use.

2.3.3 *Testing for spatial differences*

We propose a method to test the representativeness of each spatial unit with respect to the entire species in spatially structured populations with large spatial ranges. In our example, we use breeding colonies (from now on *colonies*) as spatial units, but this would also apply to populations, sub-ranges, or arbitrarily delimited areas of the entire distribution. Our method works in two steps.

In the first step, we generate a simulated distribution for the entire species. Since empirical tracking data will have biases related to differential sampling efforts and population sizes in all the colonies sampled, we first use the data to generate an unbiased simulation of the non-breeding distribution of the entire species, using the `simulateDistribution()` function (Table A1). It first calculates, for each colony, a colony level KDE by pooling all individual trips. It then generates random locations with a spatial distribution proportional to the KDE, and in a number that is proportional to the population size of the colony (Fig. A1). This is run for every colony and pools the random locations from all colonies to generate a simulated data set that mimics the non-breeding distribution of the entire species.

In the second step, we calculate each colony's representativeness of the entire species based on inclusion with the `bootstrapColony()` function. For each colony, it calculates, sequentially, the core areas (50% UDs) of an increasing number of trips and then calculates the percentage of locations from the simulated species distribution (generated with the `simulateDistribution()` function) that are included in those core areas. This is bootstrapped at each sample size. As the number of

selected trips increases, the percentage of inclusion increases as well, until it stabilises at a point when adding more trips to the data set no longer influences the spatial distribution. A non-linear regression model calculates the value of this asymptote, the inclusion value at which continuing to increase the sample size will not increase inclusion further (Fig. A2). We regard this value as the maximum inclusion value (MIV) possible for that colony. Because this MIV is calculated based on 50% UD, it ranges between 0 and 0.5 (i.e. even if a sample represents the entire species correctly, 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. The function then multiplies the MIV obtained by 100, producing the value of “species representativeness”, defined as the percentage of points of the entire species that the 50% UD of the sampled colony includes.

All the developed functions and their arguments are defined and explained in the supplementary material (Table A1), and the code can be found in Annex 2.

2.3.4 Empirical application

We demonstrate the use of these functions on a data set 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 3). Locations were obtained from GLS (R. P. Wilson, Ducamp, Rees, Culik, & Niekamp, 1992), which registers ambient light and provides one or

two positions per day deriving longitude from the time of twilight and latitude from the length of the light period. The data set contains individuals tracked for up to 8 consecutive years, which allows us to also test the effect of individual site fidelity. Details of GLS deployment for each of the colonies can be found in Table A2.

2.3.4.1 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® (Biotrack Ltd.) software, or the **GeoLight** package in R (Lisovski and 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, Chambers, & Fedak (1992) to remove other highly inaccurate locations. Phenological states (migrating, breeding, wintering) were assigned using custom-made R routines and confirmed by visual inspection.

2.3.4.2 Individual site fidelity

To test the effect of individual site fidelity on non-breeding distributions, we selected, for each species, the colony where more individuals had been tracked repeatedly (Table A3). We ran the **IndEffectTest()** function for each colony, using individual as the grouping variable and following Lascelles et al. (2016), a Scale of 186 km, which corresponds to the average error of the GLS locations (Phillips et al. 2004). We obtained a *P* value for each of them, corresponding to the testing of the H_0 : the

values of “between individual” overlap have the same distribution as the values of “within individuals” overlap. We applied a threshold α value of 0.05 to reject the H_0 .

2.3.4.3 Temporal differences

Since in our data set each of the trips corresponded to a year, and to avoid confounding inter-annual differences with individual site fidelity, 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. With the resulting data set (Table 3), we ran the `IndEffectTest()` test for each colony, using year as the grouping variable, and obtained a P value for each of them, corresponding to the testing of the H_0 : the values of “between years” overlap have the same distribution as the values of “within year” overlap. Again, we rejected this hypothesis at an α value of 0.05.

Table 3

number of tracks obtained each year at the 13 colonies of Cory's shearwaters, 19 colonies of Scopoli's shearwaters and 2 colonies of Cape Verde shearwaters. The bottom row contains the total number of tracks obtained each year, and the end column the total number of tracks obtained from each colony. For Veneguera, Pantaleu and Curral Velho, in brackets, the number of tracks used for the inter-annual differences test

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	
	Central Azores	Graciosa			8	7	10							25
		Pico					15							15
	West Azores	Corvo			6	2	2						10	
	Iberian coast	Berlenga		10					14				13	37
		Sisargas											2	2
	Chafarinas	Chafarinas			5				4	2		5		16
	Terreros	Terreros										1	1	2
		Selvagem		30	28	40	50	43	28					219
	Canary Islands + Selvagens	Montaña Clara							18	12	15	11		56
		Timanfaya											15	15
		Veneguera			27 (12)	28 (12)	27 (12)	44 (15)	61 (14)	31 (13)	18 (10)	44 (14)	29 (15)	309
Scopoli's shearwater	Palomas	Palomas									18	8	26	
	Chafarinas	Chafarinas			3			4	3		4	2	16	
	Balearic Islands	Cala Morell									10	13	23	

		NaForadada									6	6	
		NaPobra									9	9	
		Pantaleu			28 (10)	26 (10)	24 (10)	6 (6)	9 (8)		10 (7)	103	
		Porquerolles					7					7	
	France	Riou					27					27	
		Frioul					8					8	
	Corsica and Sardinia	Giraglia					24					24	
		Lavezzi					18					18	
		Filfla						5	7			12	
		Malta				1			3		9	13	
	Sicilian Channel	Gozo				1		1	2			4	
		Linosa			22	27	11	7	17	6		90	
		Zembra								13	13	26	
	Middle Adriatic	Tremiti					7	1				8	
	Strofades	Strofades			3	2		2	2			9	
	Aegean Sea	Paximada								8	8	16	
Cape Verde shearwater	Cape Verde	CurralVelho	13 (6)	8 (4)	11 (5)	9 (4)	10 (5)	5	3	2	10 (6)	71	
		Raso	13	9	5	2						29	
		Total	66	72	129	164	221	268	76	75	144	131	1346

2.3.4.4. Spatial differences

Before proceeding with the analysis, we tested the level of local representativeness of each colony sampled (Table 4) and selected only the colonies with a local representativeness value $>80\%$ (Lascelles et al., 2016). With the remaining colonies, we simulated the distribution of each species using the `simulatedDistribution()` function. We obtained a set of simulated locations for each species (Table 4), which was used in the following step. Finally, we ran the `bootstrap_Colony()` function, which calculated the species representativeness value for each colony, i.e. a measure of how well individual space use for that colony represents space use by the entire species.

Table 4

Sample size, local representativeness (in %), number of simulated locations and species representativeness for each colony. Colonies with a sample size < 4 or local representativeness < 80.0% (in italics) were considered non-representative and species representativeness was not calculated

Species	Population	Colony	Sample size	Local representativeness	Simulated locations	Species representativeness				
Cory's shearwater	East Azores	Vila	71	92.4	79,990	50.9				
		Central Azores	Faial	22			89.8	57,130	40.3	
	<i>Graciosa</i>		<i>24</i>	<i>73.4</i>						
	<i>Pico</i>		<i>14</i>	<i>67.0</i>						
	West Azores		<i>Corvo</i>	<i>9</i>	<i>68.4</i>					
	Iberian Coast	Berlenga	36	95.7	9,800	39.7				
		<i>Sisargas</i>	<i>2</i>							
	Alboran Sea	Chafarinas	10	77.8						
		<i>Terreros</i>	<i>2</i>							
	Canary Islands + Selvagens		Selvagem	211	98.5	295,400	38.8			
			Montaña Clara	55	94.8			113,750	30.5	
			Timanfaya	14	86.2					19,500
Veneguera			301	99.0	10,000					
Scopoli's shearwater	Palomas	Palomas	25	91.4		670	36.9			
	Chafarinas	<i>Chafarinas</i>	<i>5</i>	<i>59.7</i>						
Balearic Islands		Cala Morell	22	90.9		10,000	49.5			
		<i>Na Foradada</i>	<i>5</i>	<i>56.7</i>						
		<i>Na Pobra</i>	<i>8</i>	<i>62.4</i>						
		Pantaleu	101	97.5	2,100			49.8		
France	<i>Porquerolles</i>	<i>6</i>	<i>61.4</i>							

		Riou	26	94.5	2,800	38.5
		Frioul	7	83.2	700	47.5
	Corsica and Sardinia	Giraglia	23	96.3	380	36.0
		Lavezzi	17	93.9	3,000	35.2
	Sicilian Channel	Filfla	11	85.0	2,000	47.0
		Malta	12	82.9	15,500	46.5
		<i>Gozo</i>	<i>3</i>	<i>89.8</i>		
		Linosa	89	97.4	50,000	51.9
		Zembra	25	93.4	1,137,200	44.3
	Middle Adriatic	<i>Tremiti</i>	<i>7</i>	<i>71.3</i>		
	Strofades	Strofades	8	88.3	69,250	31.3
	Aegean Sea	Paximada	15	97.3	12,450	31.4
Cape Verde shearwater	East Cape Verde	Curral Velho	70	99.4	330	49.5
	West Cape Verde	Raso	28	97.9	19,599	51.0

To further understand possible causes and implications of any differences found in representativeness for Cory's and Scopoli's shearwater data, we used linear models (data from Cape Verde were excluded since we had data from only two colonies). To compensate for small sample size, 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 OLS method provided by the `lm()` function. (Cribari-Neto et al. 2007). Before running the models, we calculated a KDE of all known breeding locations for each species (Table A4), weighted by their colony size, and then calculated the centroid of the 95% UD. We took this centroid as being the species centre of mass (*species centre*, hereafter). We calculated the straight-line distance of each colony to this species centre as a measure of how centric each colony is within the species' breeding range, to include it as a predictor in the model. For Scopoli's shearwaters, we modelled the representativeness value against colony latitude and longitude (and their interaction), distance to the species 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, so we ran the full model without the interaction between latitude and longitude. For both species, we used variance inflated factor (*VIF*) values of the predictor variables and diagnostic plots to sequentially remove variables in order to improve model fit.

2.3.4.5 Non-breeding distributions of three species of *Calonectris* shearwaters

To avoid biases due to different sample sizes from each colony, and due to different colony sizes, we used the following procedure for plotting distributions: for each colony, we followed the procedure from Lascelles et al. (2016) to obtain a rasterised distribution. We obtained 22 raster layers, 7 for Cory’s shearwater colonies, 13 for Scopoli’s shearwater colonies and 2 for Cape Verde shearwater colonies. We normalised them so the sum of all cell values equals one, and multiplied them by the corresponding colony size. Finally, for each species, we added them up to plot an accurate, unbiased representation of the distribution of the species. All analyses were performed in R (R Core Team 2019) version 3.4.4.

2.4 Results

2.4.1 Individual site fidelity effect

In general, the overlap among non-breeding distributions was much higher in Cape Verde shearwaters (equipped on Curral Velho) than in the other two species, but the comparison of “between individual” and “within individual” overlap values did not show statistically significant differences for any of them (Table 5; Fig. A3).

2.4.2 Inter-annual differences in distribution

Again, the highest values of overlap were observed in Cape Verde shearwaters, but the comparison of “between year” and “within year” overlap values showed no significant differences for any of the species (Table 5; Fig. A4).

Table 5

Output from the individual site fidelity and year effect tests. For each species, median and interquartile ranges are provided, as well as the Kolmogorov-Smirnov (K-S) statistic and the P value.

		Between group overlap		Within group overlap		K-S test	
		Mean	SD	Mean	SD	D	P value
Individual effect	Cory's shearwater	0.069	0.116	0.084	0.128	0.064	0.050
	Scopoli's shearwater	0.087	0.144	0.094	0.161	0.074	0.980
	Cape Verde shearwater	0.252	0.130	0.269	0.130	0.114	0.230
Year effect	Cory's shearwater	0.065	0.115	0.065	0.117	0.014	1.000
	Scopoli's shearwater	0.071	0.132	0.058	0.120	0.056	0.933
	Cape Verde shearwater	0.251	0.132	0.232	0.132	0.107	0.579

2.4.3 *Spatial differences in species representativeness*

First, we ran the `bootstrap()` function from Lascelles et al. (2016) to test the representativeness of each sample at a local level. We excluded three of the 34 sampled colonies, which had less than four trips. From the 31 remaining colonies, we obtained values of local representativeness $>80\%$ in 22 of them (Table 4). With these, we proceeded to test their representativeness at a species level.

We simulated non-breeding distributions from each breeding colony and obtained a data set that contained a total of more than 10^5 locations for Cory's shearwaters, more than 10^6 for Scopoli's shearwaters and more than 10^4 for Cape Verde shearwaters (Table 4).

With the `colonyBootstrap()` function we calculated how well 50% UDs of each colony represented the simulated distribution of related species. Values of representativeness for Cory's shearwater colonies ranged between 21.2% and 40.3%; for Scopoli's shearwater colonies from 31.3% to 51.9%; and for Cape Verde shearwaters from 49.5% to 51.0% (Table 4). After running the full linear model for the Cory's shearwater data we first removed longitude due to high *VIF* values indicating collinearity problems. The resulting model showed no collinearity issues, but none of the predictors had a significant effect. We dredged the model (function `dredge()` from the package `MuMIn`, Barton, 2018) and averaged all resulting models with $\Delta\text{AIC} < 2.0$. The averaged model selected only latitude as a predictor, so we reran that model. The resulting model had an adjusted R^2 of 0.478, and there was a small but significant effect of latitude on the representativeness value (Table 6; Fig. A5). For Scopoli's shearwater, after running the full model we removed the interaction between latitude and longitude and the distance to the geographic centre, as they showed high *VIF* values. After confirming a poor fit of the model by visual inspection of residual plots, we also removed the latitude from the model predictors since colonies of this species are located along a West-East gradient, and the population size, since it is correlated to the distance to the species centre. The resulting model had an adjusted R^2 of 0.387 (Table 6). Distance to the species centre had a small but significant effect on the representativeness, with a P value < 0.05 and 95% confidence intervals (CI) not overlapping 0 (Table 6; Fig. A5).

Table 6

Main parameters from the models obtained from the selected generalised linear model. Continuous predictors are mean-centred and scaled by 1 SD. Effects with a P value < 0.05 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

Species	Effect	Estimate	P value	95% CI
Cory's shearwater	Intercept	0.390	0.000	0.290, 0.418
	Latitude	0.075	0.028	0.012, 0.138
	Adjusted R^2	0.478		
Scopoli's shearwater	Intercept	0.409	0.000	0.337, 0.442
	Longitude	-0.031	0.102	-0.069, 0.007
	Distance to species centre	-0.041	0.015	-0.072, -0.010
	Sample size	-0.027	0.152	-0.012, 0.065
	Adjusted R^2	0.387		

*2.4.4 Non-breeding distributions of three species of *Calonectris* shearwaters*

Tests for individual site fidelity and temporal variability showed they had no effect in our data, so we used all tracks from all the individuals, and data from all years together, to plot the non-breeding distribution of each species. We also plotted, for every species, the non-breeding 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 non-breeding area was off the coast of South Africa, at the confluence between the Agulhas and Benguela currents and the Agulhas current retroflexion (Fig. 3a and A6). In single-area distributions, however, we observed that the main non-breeding area for animals breeding in the Azores was the Agulhas retroflexion, with secondary areas in the North and South Central Atlantic (Fig. 3b); for Iberian

coast animals the main non-breeding area was the confluence between the Agulhas and Benguela currents (Fig. 3c); and for animals from the Canary Islands and Selvagens, the main non-breeding area was the southern Benguela current, with the Canary current, off the coast of Western Sahara and Senegal, as a secondary area (Fig. 3d). For Scopoli's shearwaters, the main non-breeding area was the Canary Current (Fig. 4a), but the distribution of non-breeding birds changed if we considered only animals from the eastern, central, or western colonies (Fig. 4b, 4c and 4d). The only non-breeding area of Cape Verde shearwaters was off the southern coast of Brazil (Fig. 5).

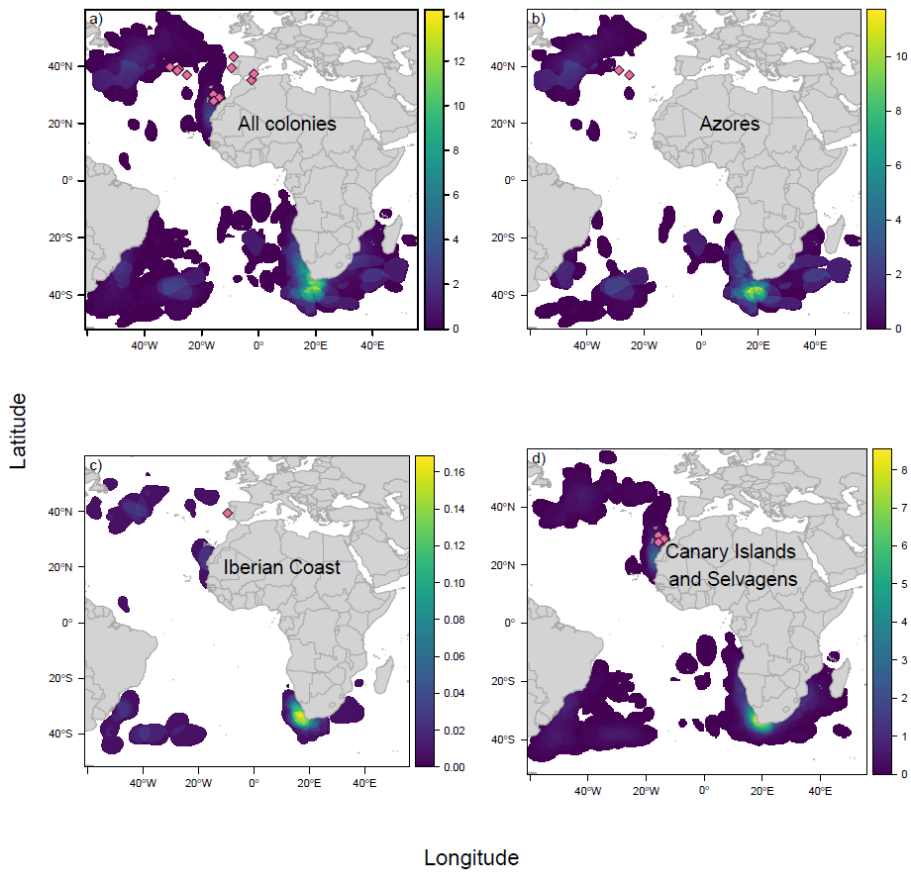


Figure 3

Non-breeding 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' non-breeding distribution (a), and using data only from the representative colonies in (b) Azores, (c) the Iberian coast, and (d) Canary Islands and Selvagens. Pink diamonds in (a) show the location of all sampled colonies. In (b), (c) and (d), pink diamonds show representative colonies for each population. Scales show the number of individuals per $0.1^{\circ} \times 0.1^{\circ}$ cell.

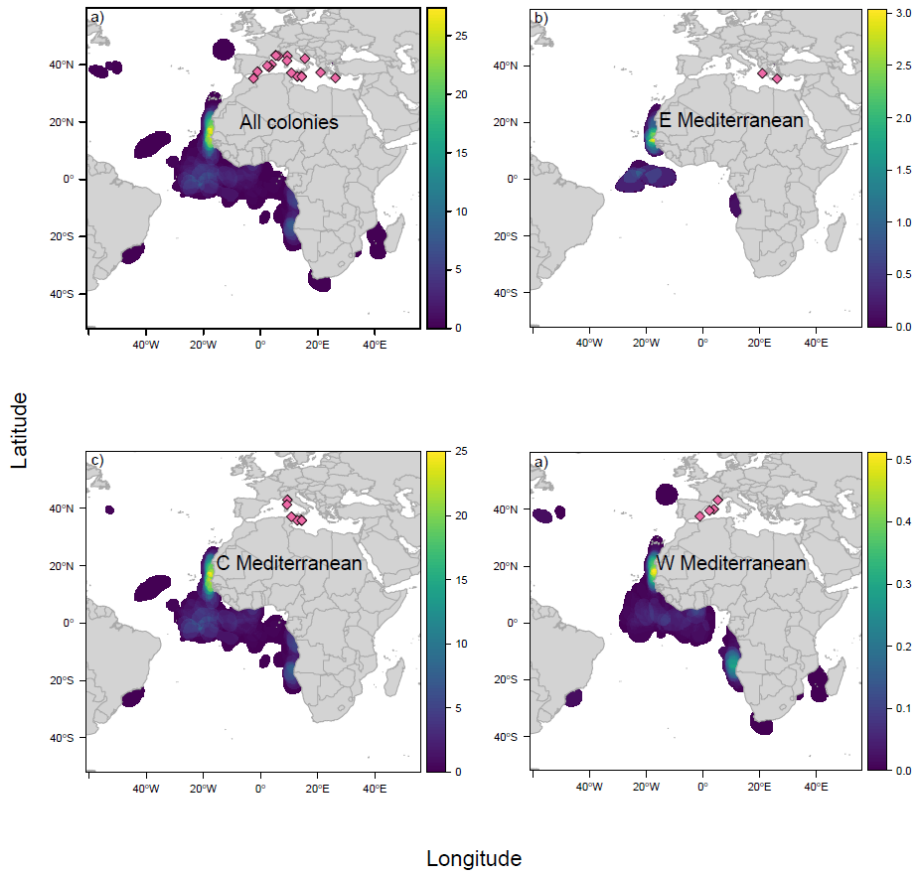


Figure 4

Non-breeding 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' non-breeding 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), (c) and (d), pink diamonds show representative colonies for each population. Scales show the number of individuals per $0.1^{\circ} \times 0.1^{\circ}$ cell.

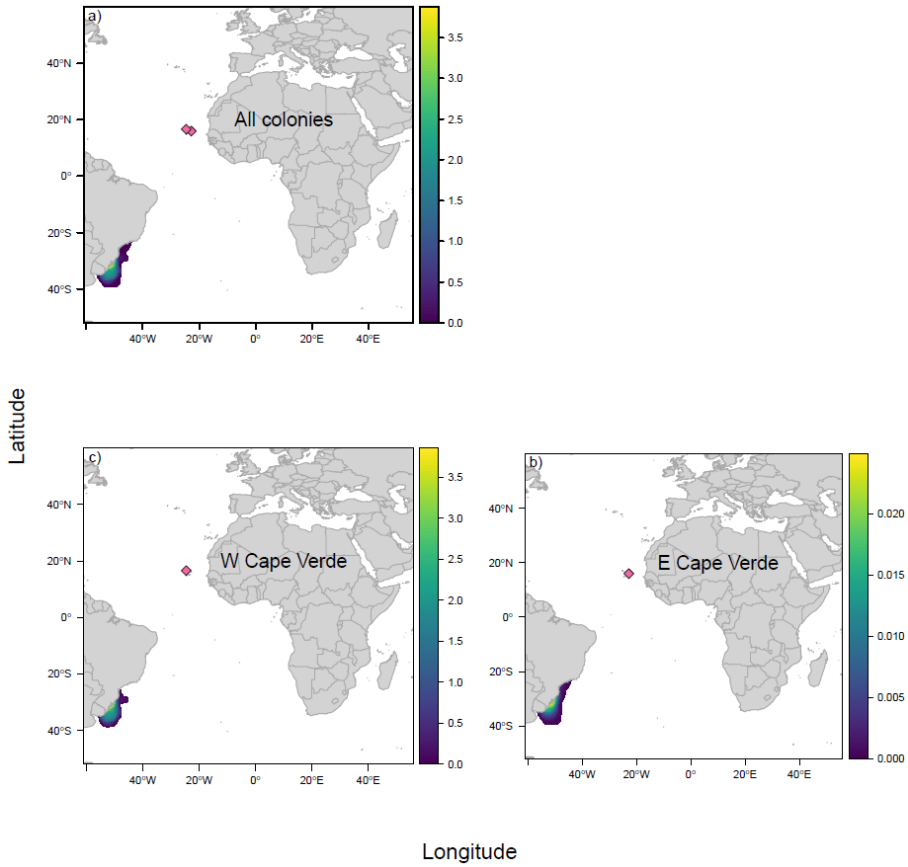


Figure 5

Non-breeding 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' non-breeding distribution (a), and using data only from the representative colonies in (b) eastern Cape Verde and (c) western Cape 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 $0.1^{\circ} \times 0.1^{\circ}$ cell.

2.5 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 (individual site fidelity, temporal variability, and spatial variability), potentially applicable to a wide variety of tracking data sets. The individual site fidelity test is useful for data sets with uneven amounts of data for each individual. It gives an idea of how much individual site fidelity affects the distribution at a population level. It should be noted that the lack of an effect does not mean that individuals do not show site fidelity. For example, in populations where all animals use the same areas, it would be possible to not detect a significant effect of individual site fidelity since overlaps are equally high in between- and within-individual comparisons.

The test for inter-annual differences gives a measure of distribution variation at the population level. We would expect this test to find a significant effect of temporal variability in adaptable species that live in predictable environments. Resilient species, or those that live in relatively constant environments, would have the same distributions year after year, resulting in non-significant effects of temporal variability on the distributions.

The representativeness test calculates how well a population represents the entire species distribution. The drawback of this function is that it requires a good knowledge of the population distribution to generate a simulated location data set against which we can compare our real location data. In spatially structured populations, we would expect the representativeness 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 non-breeding areas), we would expect the representativeness to be relatively high and similar for most colonies.

The application of the proposed method to our data set allowed us to demonstrate that there is no effect of individual site fidelity or temporal variability in the distributions of any of the three species studied, while the level of population representativeness was relatively high, but variable, for all 22 breeding colonies that we tested.

Overlap values were higher in Cape Verde shearwaters than in the other two species, as we expected since this species has only one wintering area. Neither between nor within group overlaps differed significantly in any of the species. Concordantly, evidence suggested a lack of individual consistency in the use of foraging areas during breeding and in several migratory and non-breeding parameters for both Cory's and Scopoli's shearwaters (Dias et al., 2011; Müller, Massa, Phillips, & Dell'Omo, 2014; Courbin et al., 2018, but see Navarro & González-Solís, 2009). At a population level, there are studies reporting inter-annual consistency in non-breeding distribution and migration phenology of seabirds (Müller et al. 2014; Legrand et al. 2016), and our results agreed with these findings, showing that between- and within-year overlaps in space use were not significantly different for any of the three species.

In our example with these three phylogenetically close species that share non-breeding areas, we indeed found that the species with a most restricted breeding and non-breeding distributions (Cape Verde shearwater) had the highest representative values. Conversely, the

species with more structured breeding distributions, that also spend the non-breeding period in several discrete areas, showed the lowest representativeness values. For Cory's shearwaters, the representativeness increased with latitude, but the effect of the relationship was small. This small effect was probably a by-product of the fact that the northernmost colonies in the Azores archipelago hold *ca.* 65% of the estimated world population of the species (Fontaine et al. 2011). For Scopoli's shearwaters, representativeness was negatively affected by distance to the species centre, showing that colonies closer to the centre of mass of the breeding distribution were more representative of the entire species. Although the relationship with longitude was not significant, we found that birds breeding in the easternmost colonies used the northernmost non-breeding areas preferentially. This was likely due to the fact that they had to travel longer distances to reach any of the non-breeding areas, which made their use of the southernmost non-breeding locations less likely (Ramos 2019).

These findings have implications for conservation studies. When the aim is to define areas where the studied species is most abundant, it would be more convenient to concentrate sampling efforts in colonies or breeding areas near the centre of mass of the population. However, when the aim is to define the entire distribution of a species, i.e. any region in which an individual from the studied species can be found, a sampling strategy including colonies or areas throughout the range of the species would be more appropriate.

Collaborative data sets, meta-population studies and their applications

to predicting distributions of animals from unsampled areas (e.g. Wakefield et al., 2017; Péron, Authier, & Grémillet, 2018) and to informing conservation and management policies (Hays et al. 2019) are becoming common-place in the current scientific context. Thus, accounting for these sources of variability has become essential. Our method aims to provide an objective protocol for the detection of three of the main sources of variability that can be used by movement ecologists working with a wide diversity of tracking data types.

2.6 References

- Abadie A. 2002. Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models. *J Am Stat Assoc.* 97(457):284–292.
- Aikens EO, Kauffman MJ, Merkle JA, Dwinnell SPH, Fralick GL, Monteith KL. 2017. The greenscape shapes surfing of resource waves in a large migratory herbivore. *Ecol Lett.* 20(6):741–750. doi:10.1111/ele.12772.
- Barton K. 2018. MuMIn: multi-model inference. :1–63.
- Baylis AMM, Orben RA, Arnould JPY, Peters K, Knox T, Costa DP, Staniland IJ. 2015. Diving deeper into individual foraging specializations of a large marine predator, the southern sea lion. *Oecologia.* doi:10.1007/s00442-015-3421-4. <http://link.springer.com/10.1007/s00442-015-3421-4>.
- Bhattacharyya A. 1943. On a measure of divergence between two statistical populations defined by their probability distribution. *Bull Calcutta Math Soc.* 35:99–109.
- BiotrackLtd. Biotrack Suite.
- 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. *J Anim Ecol.* 75(6):1393–1405. doi: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. *Ecol Modell.* 197(3–4):516–519. doi:10.1016/j.ecolmodel.2006.03.017.

<http://linkinghub.elsevier.com/retrieve/pii/S0304380006001414>.

- Courbin N, Besnard A, Péron C, Saraux 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. Nathan R, editor. *Ecol Lett.* 21(7):1043–1054. doi:10.1111/ele.12970. [accessed 2018 Jul 6]. <http://doi.wiley.com/10.1111/ele.12970>.
- Courbin N, Fortin D, Dussault C, Courtois R. 2014. Logging-induced changes in habitat network connectivity shape behavioral interactions in the wolf-caribou-moose system. *Ecol Monogr.* 84(2):265–285. doi:10.1890/12-2118.1.
- Cribari-Neto F, Souza TC, Vasconcellos KLP. 2007. Inference Under Heteroskedasticity and Leveraged Data. *Commun Stat - Theory Methods.* 36(10):1877–1888. doi:10.1080/03610920601126589. <http://www.tandfonline.com/doi/abs/10.1080/03610920601126589>.
- Dias MP, Granadeiro JP, Phillips RA, Alonso H, Catry P. 2011. Breaking the routine: individual Cory's shearwaters shift winter destinations between hemispheres and across ocean basins. *Proc R Soc B.* 278(1713):1786–1793. doi:10.1098/rspb.2010.2114. <http://rspb.royalsocietypublishing.org/cgi/doi/10.1098/rspb.2010.2114>.
- Dingle H. 1996. Migration: the biology of life on the move. New York: Oxford University Press.
- Ekstrom PA. 2004. An advance in geolocation by light. *Mem Natl Inst Polar Res.* 58:210–226.
- Ferreras P. 2001. Landscape structure and asymmetrical inter-patch connectivity in a metapopulation of the endangered Iberian lynx. *Biol Conserv.* 100(1):125–136. doi:10.1016/S0006-3207(00)00213-5.
- Fontaine R, Gimenez O, Bried J. 2011. The impact of introduced predators, light-induced mortality of fledglings and poaching on the dynamics of the Cory's shearwater (*Calonectris diomedea*) population from the Azores, northeastern subtropical Atlantic. *Biol Conserv.* 144(7):1998–2011. doi:10.1016/j.biocon.2011.04.022. <http://dx.doi.org/10.1016/j.biocon.2011.04.022>.

- Frederiksen M, Moe B, Daunt F, Phillips RA, Barrett RT, Bogdanova MI, Boulinier T, Chardine JW, Chastel O, Chivers LS, et al. 2012. Multicolony tracking reveals the winter distribution of a pelagic seabird on an ocean basin scale. *Divers Distrib.* 18(6):530–542. doi:10.1111/j.1472-4642.2011.00864.x.
- Giuggioli L, Bartumeus F. 2012. Linking animal movement to site fidelity. *J Math Biol.* 64(4):647–656. doi:10.1007/s00285-011-0431-7. <http://link.springer.com/10.1007/s00285-011-0431-7>.
- González-Solís J, Croxall JP, Oro D, Ruiz X. 2007. Trans-equatorial migration and mixing in the wintering areas of a pelagic seabird. *Front Ecol Environ.* 5(6):297–301. doi:10.1890/1540-9295(2007)5[297:TMAMIT]2.0.CO;2. [accessed 2017 Jul 8]. [http://doi.wiley.com/10.1890/1540-9295\(2007\)5\[297:TMAMIT\]2.0.CO;2](http://doi.wiley.com/10.1890/1540-9295(2007)5[297:TMAMIT]2.0.CO;2).
- Gutowsky SE, Leonard ML, Conners MG, Shaffer SA, Jonsen ID. 2015. Individual-level Variation and Higher-level Interpretations of Space Use in Wide-ranging Species: An Albatross Case Study of Sampling Effects. *Front Mar Sci.* 2(November):93. doi:10.3389/fmars.2015.00093. <http://journal.frontiersin.org/Article/10.3389/fmars.2015.00093/abstract>.
- Hays GC, Bailey H, Bograd SJ, Bowen WD, Campagna C, Carmichael RH, Casale P, Chiaradia A, Costa DP, Cuevas E, et al. 2019. Translating Marine Animal Tracking Data into Conservation Policy and Management. *Trends Ecol Evol.* xx:1–15. doi:10.1016/j.tree.2019.01.009. <https://linkinghub.elsevier.com/retrieve/pii/S0169534719300242>.
- Klaassen RHG, Hake M, Strandberg R, Koks BJ, Trierweiler C, Exo K-MM, Bairlein F, Alerstam T. 2014. When and where does mortality occur in migratory birds? Direct evidence from long-term satellite tracking of raptors. *J Anim Ecol.* 83(1):176–184. doi:10.1111/1365-2656.12135.
- Lascelles BG, Taylor PR, Miller MGR, Dias MP, Oppel S, Torres L, Hedd A, Le Corre M, Phillips RA, Shaffer SA, et al. 2016. Applying global criteria to tracking data to define important areas for marine conservation. Visconti P, editor. *Divers Distrib.* 22(4):422–431. doi:10.1111/ddi.12411. [accessed 2017 Apr 5].

<http://onlinelibrary.wiley.com/doi/10.1111/ddi.12411/full>.

- 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 barau*, and predicted changes induced by global warming. *Mar Ecol Prog Ser.* 550:235–248. doi:10.3354/meps11710. <http://www.int-res.com/abstracts/meps/v550/p235-248/>.
- Lisovski S, Hahn S. 2012. GeoLight - processing and analysing light-based geolocator data in R. *Methods Ecol Evol.* 3(6):1055–1059. doi:10.1111/j.2041-210X.2012.00248.x.
- Long JA. 2019. jtools: Analysis and Presentation of Social Scientific Data. <https://cran.r-project.org/package=jtools>.
- McConnell BJ, Chambers C, Fedak MA. 1992. Foraging ecology of southern elephant seals in relation to the bathymetry and productivity of the Southern Ocean. *Antarct Sci.* 4:393–398.
- MigrateTechnology. 2012. IntiProc.
- Müller M, Massa B, Phillips RA, Dell'Omo G. 2014. Individual consistency and sex differences in migration strategies of *Scopelis* shearwaters *Calonectris diomedea* despite systematic year differences. *Curr Zool.* 60(5):631–641.
- Navarro J, González-Solís J. 2009. Environmental determinants of foraging strategies in Cory's shearwaters *Calonectris diomedea* breeding on the Canary Island, NE Atlantic. *Mar Ecol Prog Ser.* 378:259–267.
- Nekolny SR, Denny M, Biedenbach G, Howells EM, Mazzoil M, Durden WN, Moreland L, Lambert JD, Gibson QA. 2017. Effects of study area size on home range estimates of common bottlenose dolphins *Tursiops truncatus*. *Curr Zool.* 63(6):693–701. doi:10.1093/cz/zox049.
- Paiva VH, Geraldés P, Marques V, Rodríguez R, Garthe S, Ramos JA. 2013. Effects of environmental variability on different trophic levels of the North Atlantic food web. *Mar Ecol Prog Ser.* 477(August):15–28. doi:10.3354/meps10180.
- Péron C, Authier M, Grémillet D. 2018. Testing the transferability of track-based habitat models for sound marine spatial planning. *Divers*

- Distrib.(June):1–16. doi:10.1111/ddi.12832.
<http://doi.wiley.com/10.1111/ddi.12832>.
- Phillips RA, Silk JRD, Croxall JP, Afanasyev V, Briggs DR. 2004. Accuracy of geolocation estimates for flying seabirds. *Mar Ecol Prog Ser.* 266:265–272. doi:10.3354/meps266265.
- R Core Team. 2019. R: A language and environment for statistical computing. <https://www.r-project.org/>.
- Ramos R. 2019. Crossing the Pillars of Hercules: Understanding transoceanic migrations of seabirds throughout their breeding range. *Ecol Evol.* 9(8):4760–4771. doi:10.1002/ece3.5079.
<https://onlinelibrary.wiley.com/doi/abs/10.1002/ece3.5079>.
- Robillard A, Gauthier G, Therrien JF, Bêty J. 2018. Wintering space use and site fidelity in a nomadic species, the snowy owl. *J Avian Biol.* 49(5):1–13. doi:10.1111/jav.01707.
- Schmidt KA. 2004. Site fidelity in temporally correlated environments enhances population persistence. *Ecol Lett.* 7(3):176–184. doi:10.1111/j.1461-0248.2003.00565.x.
- Sekhon JS. 2011. Multivariate and Propensity Score Matching Software with Automated Balance Optimization: The Matching Package for R. *J Stat Softw.* 42(7):1–52. doi:10.18637/jss.v042.i07.
<http://www.jstatsoft.org/v42/i07/>.
- Spiegel O, Leu ST, Bull CM, Sih A. 2017. What’s your move? Movement as a link between personality and spatial dynamics in animal populations. *Ecol Lett.* 20(1):3–18. doi:10.1111/ele.12708.
<http://doi.wiley.com/10.1111/ele.12708>.
- Switzer P V. 1993. Site fidelity in predictable and unpredictable habitats. *Evol Ecol.* 7:533–555.
- 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. doi:10.1080/00063659409477193. [accessed 2014 May 15].
<http://www.tandfonline.com/doi/abs/10.1080/00063659409477193>.
- Wakefield ED, Owen E, Baer J, Carroll MJ, Daunt F, Dodd SG, Green JA, Guilford T, Mavor RA, Miller PI, et al. 2017. Breeding density, fine-

- scale tracking, and large-scale modeling reveal the regional distribution of four seabird species. *Ecol Appl.* 27(7):2074–2091. doi:10.1002/eap.1591. <http://doi.wiley.com/10.1002/eap.1591>.
- Webster MS, Marra PP, Haig SM, Bensch S, Holmes RT. 2002. Links between worlds: unraveling migratory connectivity. *Trends Ecol Evol.* 17:76–83.
- Wilmers CC, Nickel BA, Bryce CM, Smith JA, Wheat RE, Yovovich V. 2015. The golden age of bio-logging: how animal-borne sensors are advancing the frontiers of ecology. *Ecology.* 96(7):1741–1753. doi:10.1890/14-1401.1. <http://www.esajournals.org/doi/abs/10.1890/14-1401.1>.
- Wilson RP, Ducamp JJ, Rees WG, Culik BM, Niekamp K. 1992. Estimation of location: global coverage using light intensity. In: *Wildlife telemetry: remote monitoring and tracking of animals.* p. 131–134.
- Wolf JBW, 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. doi:10.1007/s00442-007-0665-7. <http://link.springer.com/10.1007/s00442-007-0665-7>.
- Worton BJ. 1989. Kernel Methods for Estimating the Utilization Distribution in Home-Range Studies. *Ecology.* 70(1):164–168.
- Yurkowski DJ, Auger-Méthé M, Mallory ML, Wong SNP, Gilchrist G, Derocher AE, Richardson E, Lunn NJ, Hussey NE, Marcoux M, et al. 2018. Abundance and species diversity hotspots of tracked marine predators across the North American Arctic. *Divers Distrib.*(January 2018):328–345. doi:10.1111/ddi.12860. <http://doi.wiley.com/10.1111/ddi.12860>.

3

Migratory connectivity and non-breeding habitat segregation across geographical scales in three closely related seabird taxa

3.1 Abstract

In migratory species, migratory connectivity has relevant consequences in population dynamics, conservation and management, and even genetic mixing. In colonially breeding species, where the spatial structure is an important trait of their populations, a strong migratory connectivity (i.e. the spatial structure of the breeding population is maintained during the non-breeding period) can promote isolation and ecological divergence between populations which, ultimately, can affect the process of lineage segregation. Studying the differences in habitat use and environmental preferences among colonies, populations, or taxa, can improve our

understanding of the ecological segregation among them, and provide insights into the geographical structure of a species. We studied the non-breeding ecological niche and migratory connectivity of 805 birds breeding in 34 different colonies of the three Atlantic and Mediterranean *Calonectris* seabird taxa to understand at what geographical scales migratory connectivity, wintering habitat segregation, and differences in environmental preferences emerge. We calculated the migratory connectivity, and modelled their non-breeding ecological niches, at a taxa, population, and colony level. At a taxa level, we found a clear spatial segregation of non-breeding distributions between Cory's (*C. borealis*) and Scopoli's (*C. diomedea*) shearwaters, and a clear ecological segregation between Cory's and Cape Verde (*C. edwardsii*) shearwaters. At a population level, we found some structure in the migratory connectivity and non-breeding environmental preferences that was not maintained at a colony level. Our results provide evidences of non-breeding spatial segregation of Scopoli's shearwaters non-breeding area and that of the other two taxa, and of different environmental preferences among Cory's and Cape Verde shearwaters, despite sharing non-breeding grounds. In addition, the presence of migratory connectivity at a population level, but not at a colony level, suggests that birds from nearby colonies mix in the non-breeding areas among them, but not with birds from different populations, which has important implications not only for population dynamics but also for conservation and management.

3.2 Introduction

Migratory connectivity, or the link between breeding and non-breeding distributions in migratory populations, has relevant consequences, among others, for individual fitness, population dynamics and genetic mixing, as well as for conservation and management (Bauer et al. 2016; Finch et al. 2017). Understood as the preservation of the spatial structure between the breeding and non-breeding distribution of a species, it is usually classified along a continuum from weak to strong (Webster et al. 2002). Weak migratory connectivity occurs when the geographic structure of populations in the breeding grounds is not maintained in the non-breeding grounds (i.e. animals from different breeding populations mix during non-breeding periods), while strong migratory connectivity occurs when the geographic structure in the non-breeding grounds mirrors that of the breeding grounds (i.e. animals that breed closer to each other also spend the non-breeding period closer to each other, and farther from others, Finch et al. 2017). In species with weak migratory connectivity there are more opportunities for encounters between animals from different breeding populations in the non-breeding grounds, which can facilitate a potential change in breeding location, potentially leading to gene flow among populations or even hybridization and introgression (Tigano et al. 2015; Quillfeldt et al. 2017). Conversely, in species with strong migratory connectivity, the opportunities for encounters between animals of different populations are reduced, and thus the population genetic structure may be stronger (Burg and Croxall 2001).

In colonially breeding species, such as most seabirds, the biogeographical structure of their populations can promote the process of ecological divergence (Matthiopoulos et al. 2005). In these species of high flying capacity, the geographic structure (spatially isolated patches of suitable habitat in a continuum of unsuitable habitat, Hanski and Gilpin 1997) is mostly maintained by behavioural traits, such as philopatry or breeding site fidelity, rather than by distance or physical barriers (Matthiopoulos et al. 2005; Friesen et al. 2007). However, in migratory species, these behavioural barriers disappear during the non-breeding period, when individuals are no longer tethered to the colony by breeding duties. This allows them to either disperse randomly without specified non-breeding distributions, or migrate to specific non-breeding areas that might or might not be shared among colonies, which adds further complexity to their population structure (Esler 2000; Friesen 2015).

In this context, modelling the non-breeding ecological niche of seabirds and analysing at what biogeographic scales (i.e. colony, population and taxon) connectivity patterns occurs is essential to understand the scale of ecological divergence and, ultimately, to provide evidence of lineage separation and management units (Raxworthy et al. 2007). This can be particularly insightful when dealing with closely related taxa with incomplete lineage separation, i.e. when morphological divergence is subtle, reproductive isolation is not absolute or genetic differentiation is still small (Rissler and Apodaca 2007; Fišer et al. 2018). This is the case of the *Calonectris* complex, a group of four taxa of pelagic seabirds breeding colonially in islands and islets in subtropical waters of the Northern Hemisphere (Warham 1996). Three of them have a parapatric breeding distribution spanning the North-eastern coast of the Atlantic

Ocean, and the Mediterranean Sea, and spend the non-breeding season in several discrete areas of the Atlantic and Indian Oceans, with different degrees of inter-specific mixing (Reyes-González et al. 2017). Of these three, the Cape Verde shearwater (*C. edwardsii*; CALEDW) diverged between 7,000 and 9,000 years ago (Gómez-Díaz et al. 2006), and is consensually regarded as a full species (Hazevoet 1995). Cory's (*C. borealis*; CALBOR) and Scopoli's (*C. diomedea*; CALDIO) shearwaters have only recently been recognised as separate species (Sangster et al. 2012), and there is still debate as to whether they should be considered as such (Genovart et al. 2013). Despite their mainly parapatric breeding distribution, there is a small region around the strait of Gibraltar where CALBOR and CALDIO breed in sympatry, with colonies of both taxa at a very short distance and even a mixed colony where there have been reports of mixed pairs (Gómez-Díaz et al. 2006; Zidat et al. 2017). These characteristics make them a unique group to study ecological divergence through the analyses of non-breeding distributions and migratory connectivity at different hierarchical levels (colony, population, and taxon).

For this work, we have collated a dataset of 1,346 year-round tracks of 805 birds from these three *Calonectris* taxa, breeding in 34 different colonies spread along the Mediterranean and Atlantic basins that will allow us to explore the degree of non-breeding ecological niche segregation at different biogeographic scales (i.e. colony, population, species/taxon). Differentiation of migratory strategies and segregation in habitat and ecological niche should help us detect ecological divergence among populations or taxa. First, we expect to find low level of non-breeding segregation and low migratory connectivity among breeding

populations of the same taxon, but high level of segregation and high migratory connectivity among different taxa. Second, studying the environmental preferences at an intra-taxon and inter-taxon levels, we expect higher ecological segregation of the non-breeding ground among different taxa than among populations of a given taxon.

3.3 Methods

3.3.1 Studied taxa and sampling design

Calonectris shearwaters are medium-sized Procellariiforms, with long life span, high reproductive investment and low reproductive output (Navarro and González-Solís 2007; Ramos et al. 2012), strongly philopatric, and faithful to the breeding colonies (Mougin et al. 1999). Females lay a single egg per season, and incubation and chick-rearing duties are shared by both parents (Thibault et al. 1997; Granadeiro et al. 2006). CALBOR breeds mainly on Macaronesian islands (except Cape Verde) and off the West coast of the Iberian Peninsula and spends the non-breeding season in different areas of the Atlantic and Indian Oceans, off the coasts of Africa and South America as well as two oceanic areas of the Atlantic, one below the Labrador Peninsula and another in the South Atlantic over the mid-oceanic ridge (González-Solís et al. 2007; Dias et al. 2011). CALDIO breeds inside the Mediterranean and spends the non-breeding season in the Atlantic Ocean off the West coast of Africa and East coast of South America (see Figures X1, X2 and X3 in Chapter 1), and CALEDW breeds in the Cape Verde islands and spends the non-breeding season off the East coast of South America (Birdlife International 2019). The three taxa have similar phenology, arriving at the colonies for breeding in late February – early March. Laying occurs

around the second half of May, and the hatching period is around mid to late July. In late October or early November, chicks fledge and abandon the colony, not to return to breed until they reach about 6 years of age (Thibault et al. 1997; Warham 1997; Granadeiro et al. 1999; Mougin et al. 2000; Paiva et al. 2015).

Our dataset comprises year-round at-sea positions of adult birds breeding in 13 CALBOR colonies, 19 CALDIO colonies, and two CALEDW colonies, representing the entire breeding distribution of the three taxa, and obtained between the summers of 2006 and 2016 (Fig. 6).

3.3.2 Tracking data

Positions were obtained with Global Location Sensors (GLS; Wilson et al. 1992), which provide one or two positions per day, using daylight to calculate longitude from the time of twilight and latitude from the length of the day (light period). Detailed deployment information for each colony can be found in Table A2. Twilight events calculated from the light measurements were visually inspected and corrected when interferences near the twilights were detected. Locations were obtained from the light data using either Intiproc® (Migrate Technology Ltd.) or Biotrack® (Biotrack Ltd.) software, or the GeoLight package in R (Lisovski and Hahn 2012). To eliminate biases due to incorrect latitude estimations during or near equinoxes (Ekstrom 2004) we removed position data from 20 days before and after the equinoxes. In addition, we removed unrealistic positions by applying a quadratic speed filter following McConnell et al. (1992).

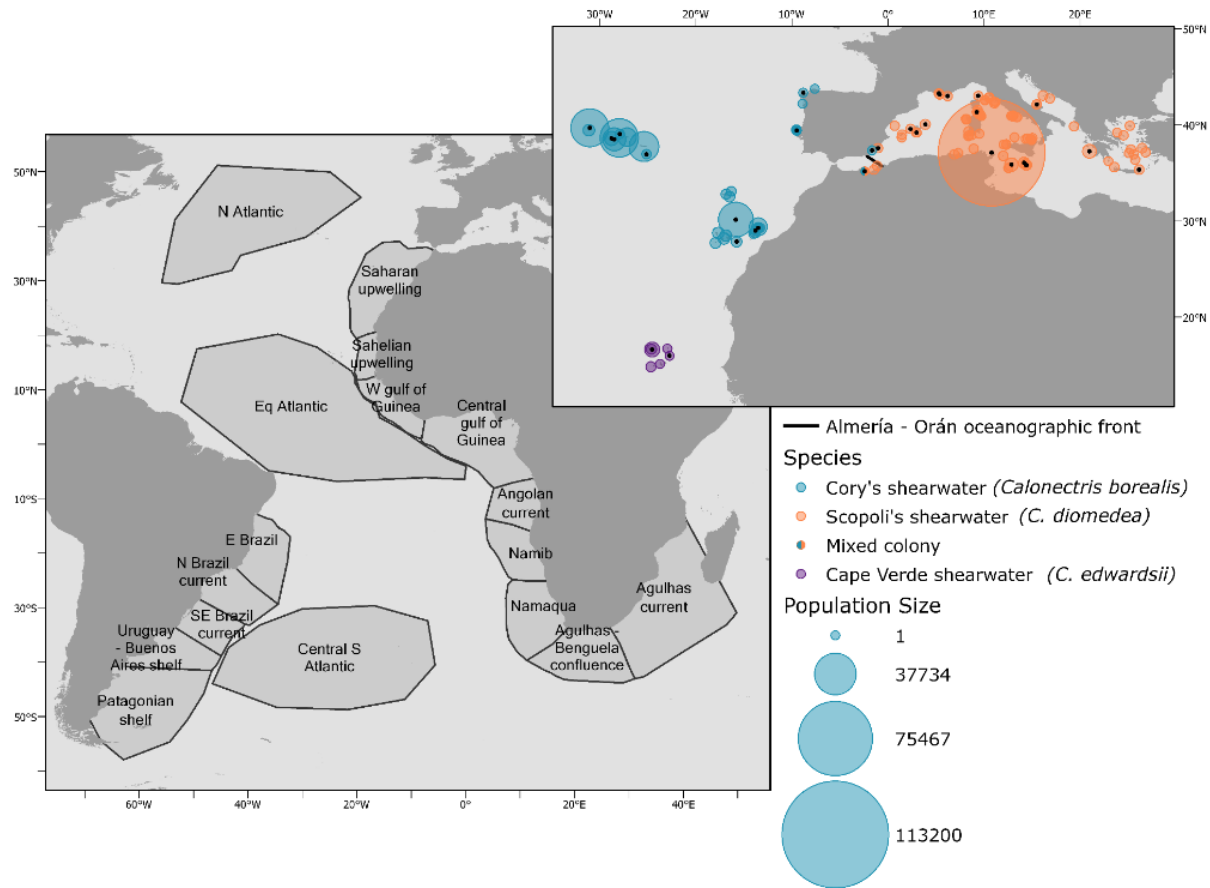


Figure 6 (previous page)

Breeding distribution and potential non-breeding areas of the three *Calonectris* shearwaters studied: a) known breeding locations of Cory's shearwaters (in blue), Scopoli's shearwaters (in orange) and Cape Verde shearwaters (in purple). The size of the coloured circle is proportional to the breeding population estimated for each colony (not linearly) and the colonies with black full circles inside are the ones where we have tracked birds. Notice that a mixed colony of CALBOR and CALDIO occurs in Chafarinas Islands, in the area of the Strait of Gibraltar. The black, continuous line represents the Almería – Orán oceanographic front (AEOF). b) Non-breeding areas of the three Palearctic *Calonectris* species as used for the analyses of this work. Adapted from Spalding et al. 2007;

We visually inspected tracks to assign a phenological state to each position. Based on the distance between consecutive positions and directionality of movement we classified periods as “residency” when for at least four days the distance between consecutive positions was short and the movement non-directional. The residency periods during the breeding season and around the breeding colonies were classified as breeding. The residency periods outside the breeding season were classified as main non-breeding (the longest) and staging (the rest of them). Movement periods (long distances between consecutive locations and directionality) were classified as migration. For our analysis we used only residency periods during non-breeding. Since our dataset contains individuals tracked for up to eight consecutive years, and along a 10-year period, we followed the methodology developed in Chapter 1 to check for the effect of individual site fidelity and temporal variability in the spatial distribution of each of the taxa.

3.3.3 *Migratory connectivity*

We divided the non-breeding distribution of the individuals using the Marine Ecoregions Of the World system (Spalding et al. 2007). These

regions extend only to 200 nautical miles offshore or to the 200-m isobath (whichever occurs first). Since *Calonectris* shearwaters can spend the non-breeding season farther from the coast, we extended these areas offshore to include most of the non-breeding positions but maintaining the limits between regions (Fig. 6). In addition, since this division does not include offshore areas, but some adult shearwaters do spend their non-breeding periods in the middle of the Atlantic Ocean, we generated three areas *de novo*: North Atlantic (55° W – 19° W, 30° N – 21° N), Equatorial Atlantic (52° W – 0° W, 6° S – 20° N), and central South Atlantic (46° W – 6° W, 48° S – 30° S; Fig. 6). To assign a non-breeding region to each individual trip we calculated, with the package `adehabitatHR` (Calenge 2011), the 5% kernel density utilisation distribution (UD, Worton 1989) of the non-breeding positions and assigned as non-breeding region the one occupied by the centroid of this kernel.

To group the breeding colonies of each taxon in populations, we created a buffer of 100 km around each colony, sampled and un-sampled (Table A4) using the function `gBuffer()` (package `rgeos`, Bivand and Rundel 2019). and considered those colonies whose buffers overlapped as one population.

We estimated the migratory connectivity following Cohen et al. (2017), which improves on the classical migratory connectivity method (Ambrosini et al. 2009). The method uses transition probabilities between specific breeding and non-breeding sites and can consider uneven sampling and different abundances. The resulting Migratory Connectivity value (MC) can take values between -1 and 1. MC close to 1 indicates that populations that breed close-by also spend the non-

breeding period close-by (strong migratory connectivity); MC close to -1 indicates that populations that breed close together spend the non-breeding period farther apart and MC close to 0 indicates that the relationship between breeding and non-breeding distances is no different from what expected at random (weak migratory connectivity). We calculated MC for each taxon taking first *colony* and then *population* (defined as all colonies with overlapping 100 km buffers, see above) as breeding sites, and then taking *taxon* as the breeding site. MC estimates are reported as median values (95% confidence interval, CI) along the manuscript. The functions developed by Cohen et al. (2017) are implemented in the package MigConnectivity (Hostetler and Hallworth 2018). We used the function `estMC()`, designed to estimate the value of MC taking into account differences in population size among origins, different sampling efforts, and the position error associated to the GLS data, with estimates of bias and variance obtained from modelling ground-truthing GLS data. For each combination of origin (breeding site) and destination (non-breeding site) we ran the `estMC()` function with 1,000 simulations (value recommended for GLS data) and 1,000 samples, with a number of maximum tries of 300 (more details can be obtained in Cohen et al. 2017).

To visually represent the migratory connectivity we constructed circular plots with the package `circlize` (Gu et al. 2014). To accurately represent how breeding and non-breeding distributions are connected, the widths of the sectors were calculated to be proportional to the population size, and the width of the links to be proportional to the number of individuals going from each origin to each destination, corrected by origin population size and sampling effort.

3.3.4 Environmental data and ecological niche modelling

To model the habitat preferences of shearwaters we considered environmental variables known to affect seabird distributions (Wakefield et al. 2009; Louzao et al. 2011). Sources of the environmental variables, units, and processing details can be found in Table 7. After checking pairwise correlations between all covariates, bathymetry (m, BAT) and sea floor slope (% , SLO) were included as static variables, and chlorophyll-A concentration ($\text{mg}\cdot\text{m}^{-3}$, CHLA), CHLA temporal variability ($\text{mg}\cdot\text{m}^{-3}$, CHLA_var), sea surface temperature ($^{\circ}\text{C}$, SST), SST spatial gradient (% , SST_grad), and salinity ($\text{g}\cdot\text{kg}^{-1}$, SAL) were included as dynamic environmental variables. Latitude and longitude were also included as variables to capture the spatial structure of the data and avoid the effects of spatial autocorrelation (Segurado et al. 2006). All environmental variables were scaled to have mean zero and standard deviation of one before entering the models, and the CHLA was back transformed after the analysis to retrieve original values.

Table 7

Environmental variables used as predictors in the ensemble models, their resolution, processing to final product and source where they were obtained. The variables retained after checking Spearman correlations are highlighted in bold.

Variable	Units	Original resolution	Final resolution	Processing	Source
Bathymetry	m	1 arcmin	0.5 ^o		doi:10.7289/V5C8276M
Slope	rad	1 arcmin	0.5 ^o	Calculated with raster::terrain	
ChlA concentration	mg·m ⁻³	4 km	0.5 ^o	log transformed to normalize values	https://bit.ly/2mqYiN9
ChlA gradient	mad	4 km	0.5 ^o	Calculated with raster::terrain	
ChlA temporal variability	%	4 km	0.5 ^o	$\frac{\max(CHLA) - \min(CHLA)}{\max(CHLA)} \times 100$	Oppel et al. 2012
Sea surface temperature	°C	1/12 ^o	0.5 ^o		https://bit.ly/2mpaldQ
SST Gradient	°C	1/12 ^o	0.5 ^o	Calculated with raster::terrain	
Salinity	PSU	1/12 ^o	0.5 ^o		https://bit.ly/2mpaldQ
SS Height	m	1/12 ^o	0.5 ^o		https://bit.ly/2mpaldQ
Ekman transport intensity	m·s ⁻¹	0.25 ^o	0.5 ^o		doi:10.1002/2014GL061773
Mixed Layer Thickness	m	1/12 ^o	0.5 ^o	Calculated from ChlA gradient	https://bit.ly/2mpaldQ
Net Primary Productivity	mg·m ⁻³ ·day ⁻¹	0.25 ^o	0.5 ^o		https://bit.ly/2lZmiqt
Phytoplankton concentration	Mmol·m ⁻³	0.25 ^o	0.5 ^o		https://bit.ly/2lZmiqt

Habitat preferences were evaluated through ensemble species distribution models (Marmion et al. 2009; Thuiller et al. 2009) using the function `ensemble_modelling()` from the package SSDM (Schmitt et al. 2017). Due to different sampling efforts, the numbers of at-sea locations for each colony in our dataset were not proportional to the respective colony sizes. For all models except at colony level, to ensure our at-sea locations data accurately represented the real proportions of birds from each colony found at sea, we simulated a number of non-breeding locations proportional to colony size using the locations obtained from the GLS data of each colony (following the methodology developed in Chapter 1). From this process we obtained a set of non-breeding simulated locations for each of the colonies of the three taxa. We modelled the non-breeding distributions at four different levels: we performed 25 models at a colony level (nine CALBOR colonies, 14 CALDIO colonies, and two CALEDW colonies); 17 models at a population level (seven CALBOR populations, eight CALDIO populations, and two CALEDW populations) pooling the non-breeding simulated data of the corresponding colonies; three models at a taxon level (CALBOR, CALDIO and CALEDW) pooling the non-breeding simulated data of all the colonies within each taxon; and one model at a supra-taxon level (the three taxa together) pooling the simulated non-breeding distributions of all the colonies. For each of these models we selected 2,000 presence data points from that pooled dataset, to speed up computation and to avoid differences in model fit caused by different number of presences (i.e. the models at a taxa or supra-taxa level would have more presences than any of the models at a population level). To perform this selection randomly we used the function `rthin()` from the

package `spatstat` (Baddeley et al. 2015). The function performs independent random thinning, producing a point pattern with the same relative density as the original one but a smaller number of points. These 2,000 randomly selected presences were the ones used to perform each of the environmental niche models.

To generate the ensemble models, from all the available modelling algorithms, we excluded generalised linear models because it has been reported to perform poorly with spatially structured data (Segurado et al. 2006). Thus, we used generalised additive models, multivariate regression splines, generalised boosted regressions models, classification tree analyses, random forests, maximum entropy, artificial neural networks, and support vector machines. Each model was validated by holdout cross-validation, using 70% of the dataset as training data, and we ran every algorithm 5 times and selected the best model from each of them. The participation of each algorithm to the final ensemble model was weighted by its AUC (Area Under the Curve, Fielding and Bell 1997) value, and only models with $AUC > 0.80$ were selected for the ensemble (except in the supra-taxon level model, where the AUC threshold to include models in the ensemble had to be lowered to 0.7 since no model had an $AUC > 0.8$). Variable importance was calculated as the score $(1-r) \cdot 100$ where r is the Pearson correlation of the full model prediction and the prediction of the model without the evaluated variable. Thus, the higher the value the more important the variable is for the model.

To measure similarity between non-breeding habitats of colonies within each taxa, (CALBOR and CALDIO, we excluded CALEDW because there are only two sampled colonies), we calculated pairwise Pearson

correlations between all possible pairs of model projections, as well as distances among colonies. We then used linear models (`lm()` function in R) to calculate the relationship among pairwise distances between colonies and pairwise Pearson correlations between corresponding model projections. All analyses were performed in R (version 3.6.0, R Core Team 2019)

3.4 Results

3.4.1 Migratory connectivity

Table 8

MC values (median and 95% CI) estimated at a species level (for the three taxa together), and at a population and colony levels for each species. For Cape Verde shearwater there is only analysis at a colony level because the two studied colonies belong to the same population.

Aggregation level	Species	Median	(95% CI)
Taxon	CALBOR	0.455	(0.418 – 0.494)
	CALDIO		
	CALEDW		
Population	CALBOR	0.186	(0.097 – 0.375)
Population	CALDIO	0.318	(0.246 – 0.496)
Colony	CALBOR	0.045	(0.006 – 0.101)
Colony	CALDIO	0.062	(0.022 – 0.161)
Colony	CALEDW	0.006	(-0.006 – 0.040)

Firstly, we estimated the migratory connectivity (MC) at a taxon level, i.e., grouping all individuals by taxon and estimating the migratory connectivity taking the three taxa as the three only possible origins, and all the non-breeding regions as possible destinations. From the 17 non-breeding regions available, CALBOR used a total of 15, of which it shares nine with either of the other two taxa (60.0%). CALDIO used a total of nine, of which seven were shared with at least one of the other two taxa (77.8%) and CALEDW shearwaters used a total of three non-

breeding regions, all of them shared (Fig. 7). Overall MC was 0.45 (0.42 – 0.49; Table 8 & Fig. A7).

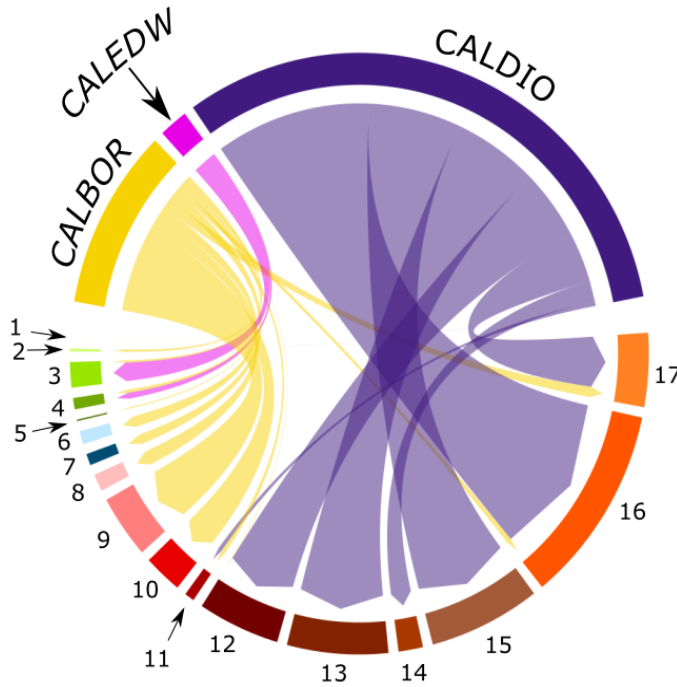


Figure 7.

Circular plot showing the links between each taxon (top half) and non-breeding area (bottom half). The widths of the sectors and links are proportional to the number of individuals from each taxon travelling to each of the non-breeding areas, after correcting for the different sampling effort in each colony and accounting for different population sizes.

Non-breeding areas: (1) Eastern Brazil, (2) Northern Brazil current – South-eastern Brazil, (3) South Brazil current, (4) Uruguay – Buenos Aires shelf, (5) Patagonian shelf, (6) North Atlantic, (7) Central-southern Atlantic (8) Agulhas current, (9) Agulhas – Benguela confluence, (10) Namaqua, (11) Namib, (12) Angolan current, (13) Central gulf of Guinea, (14) West gulf of Guinea, (15) Equatorial Atlantic, (16) Sahelian upwelling and (17) Saharan upwelling.

Within CALBOR, MC was 0.19 (0.10 – 0.37) at a population level and 0.04 (0.01 – 0.10) at a colony level (Table 8 & Fig. A7). Birds from the colony of Veneguera (West Canary Islands), Montaña Clara (East Canary Islands) and Selvagem (Salvages Islands) used most of the non-breeding regions (13, 12 and 12, respectively, out of 17). The Namaqua coast and the Agulhas – Benguela confluence non-breeding regions were used by all colonies from all populations, while the central coast of the Gulf of Guinea was only used by birds from Chafarinas (Fig. 8).

Within CALDIO, MC was 0.32 (0.25 – 0.50) at a population level and 0.06 (0.02 – 0.16) at a colony level (Table 8 & Fig. A7). The colonies of Pantaleu (Balearic Islands) and Palomas (South-east Spain) used most of the non-breeding areas (seven and eight, respectively), while Paximada (Aegean Sea) and Strofades (Eastern Ionian Sea) use only three of them. Of the non-breeding regions, the Agulhas current is exclusive to birds from Palomas, and used only by one individual, and the Northern Brazil current is used only by birds from two colonies, Frioul (France) and Lavezzi (Corsica – Sardinia). Conversely, the Sahelian coast was used by birds breeding in all colonies from all populations, while the Equatorial Atlantic was used by birds of all populations, and all colonies but Frioul (Fig. 9).

Within CALEDW, as each of the two colonies represented one population, there is only one measure of MC. The median MC value was 0.01 (-0.01 – 0.04; Table 8 & Fig. A7). The three non-breeding regions were used in a similar proportion by birds from the two colonies (Fig. 10)

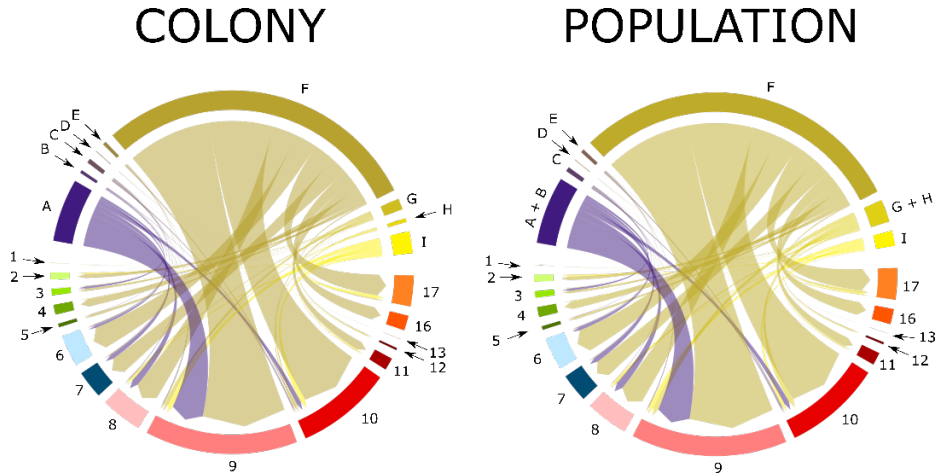


Figure 8

Circular plot showing the links between each colony (top half) and non-breeding area (bottom half) for CALBOR colonies (a) and populations (b). The width of the sectors and links are proportional to the number of individuals from each colony travelling to each of the non-breeding areas, after correcting for the different sampling effort in each colony and accounting for different population sizes.

Colonies: (A) Faial, (B) Graciosa, (C) Vila, (D) Chafarinas, (E) Berlenga, (F) Selvagem, (G) Montaña Clara, (H) Timanfaya and (I) Veneguera.

Populations: (A + B) Central Azores, (C) East Azores, (D) Chafarinas, (E) Berlengas, (F) Selvagem, (G + H) East Canary Islands, (I) West Canary Islands

Non-breeding areas: (1) Eastern Brasil, (2) Northern Brasil Current – South-eastern Brasil, (3) South Brasil Current, (4) Uruguay – Buenos Aires shelf, (5) Patagonian shelf, (6) North Atlantic, (7) Central-southern Atlantic (8) Agulhas current, (9) Agulhas – Benguela confluence, (10) Namaqua, (11) Namib, (12) Angolan current, (13) Gulf of Guinea Central, (16) Sahelian upwelling and (17) Saharan upwelling

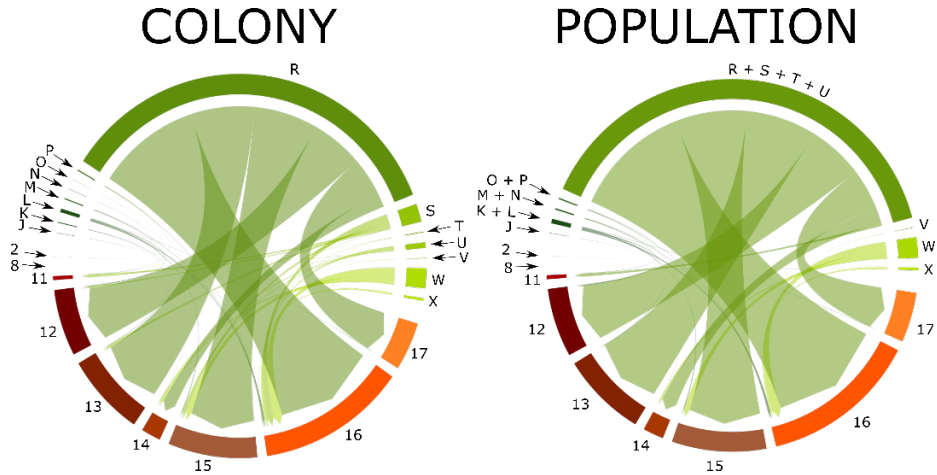


Figure 9

Circular plot showing the links between each colony (top half) and non-breeding area (bottom half) for CALDIO colonies (a) and populations (b). The width of the sectors and links are proportional to the number of individuals from each colony travelling to each of the non-breeding areas, after correcting for the different sampling effort in each colony and accounting for different population sizes.

Colonies: (J) Palomas, (K) Pantaleu, (L) Cala Morell, (M) Riou, (N) Frioul, (O) Giraglia, (P) Lavezzi, (R) Zembra, (S) Linosa, (T) Filfla, (U) Malta, (V) Tremiti, (W) Strofades, (X) Paximada.

Populations: (J) Palomas, (K + L) Balearic Islands, (M + N) France, (O + P) Corsica and Sardinia, (R + S + T + U) Sicilian Channel, (V) Middle Adriatic, (W) Strofades, (X) Aegean Sea

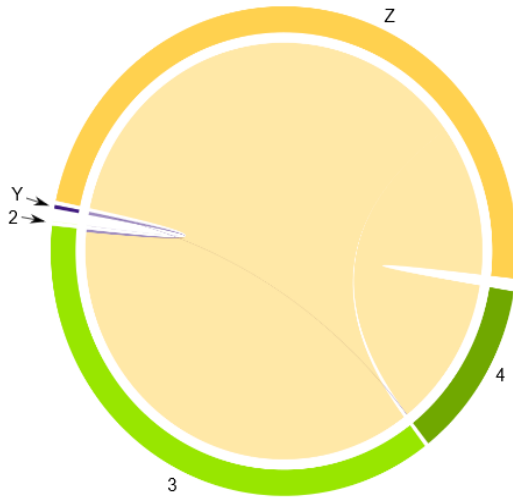


Figure 10

Circular plot showing the links between each colony (top half) and non-breeding area (bottom half) for CALEDW. The width of the sectors and links are proportional to the number of individuals from each colony travelling to each of the non-breeding areas, after correcting for the different sampling effort in each colony and accounting for different population sizes.

Colonies: (Y) Curral Velho, (Z) Raso.

Non-breeding areas: (2) Northern Brasil Current – South-eastern Brasil, (3) South Brasil Current, (4) Uruguay – Buenos Aires shelf.

3.4.2 *Ecological niche modelling*

From the ensembled models we obtained predicted non-breeding distributions at three levels: population, taxon and supra-taxon. At a population level all models for CALBOR had lower AUC (range 0.82 – 0.87) than any of the CALDIO (range 0.88 – 0.95) and CALEDW models (range 0.98 – 0.99). At a taxon level, once more, the worst fit was for CALBOR (AUC = 0.81) and the best for CALEDW (AUC = 0.98). For CALBOR, the taxon-level had a worse fit than any of the population-

level or colony-level models. For CALDIO the taxon-level model had a worse fit ($AUC = 0.90$) than all but three (out of eight) population-level models. For CALEDW the model fit at a population and taxon level were very similar. The model with the lowest AUC was the supra-taxon level model ($AUC = 0.78$; Table 9).

From the ensembled models we obtained importance scores that allowed us to know which environmental variables were more influential in the distribution of birds (Table 9). For the supra-taxa model the most important variable was CHLA (importance score 18.5). SST was the most important variable for CALBOR at a taxon level, but with an importance score (11.0) very similar to that of CHLA (10.3) and SST_grad (10.1). At population level, SST was the most important variable for three of the populations, while CHLA was the most important for the other four. For CALDIO the two most important variables at a taxon level were also CHLA and SST with similar importance scores (14.6 and 11.1 respectively). At a population level, CHLA was the most important variable for four populations, and SST for three. Only for one population (Aegean Sea) the most important variable was CHLA_var (11.8).

For CALEDW the most important variable was SST for all models. For the CALEDW models, although CHLA was the most important variable for both populations, at a taxon level the most important variable was SST. Nevertheless, the importance scores of both variables were very similar for all three models.

Table 9

Model evaluation parameters (AUC and variable importance ratio) for each of the models at a population, taxa and supra-taxa level. AUC ranges between 0 (worst fit) and 1 (best fit), and variable importance ratios range between 0 (least important) and 100 (most important). The most important environmental variable for each model is shown in bold, and the least important in italics

Group	Population	AUC	BAT	SLO	SST	CHLA	SAL	CHLA var	SST grad
CALBOR	Central Azores Is.	0.85	4.6	<i>4.0</i>	5.1	10.8	9.8	5.3	4.6
	East Azores Is.	0.82	4.7	<i>4.5</i>	7.8	13.9	9.1	7.8	6.1
	Chafarinas Is.	0.87	6.0	<i>4.6</i>	7.1	9.2	6.7	4.8	5.4
	Berlengas Is.	0.84	6.0	<i>5.0</i>	14.4	13.3	5.8	7.0	5.8
	Salvages Is.	0.84	5.3	<i>5.0</i>	16.0	13.1	7.7	6.7	8.9
	East Canary Is.	0.87	7.7	<i>5.0</i>	21.0	12.3	7.0	8.4	7.0
	West Canary Is.	0.86	12.4	<i>4.6</i>	11.1	19.3	5.4	7.9	5.6
	Taxon level model	0.81	<i>6.4</i>	6.5	11.0	10.3	9.9	7.0	10.1
CALDIO	Palomas	0.90	5.5	<i>5.1</i>	9.0	12.9	5.9	5.3	7.6
	Balearic Is.	0.90	4.2	<i>3.9</i>	6.8	10.7	6.1	4.3	8.7
	France	0.88	<i>3.8</i>	4.7	15.1	14.6	5.4	11.0	5.7
	Corsica – Sardinia	0.90	4.8	<i>4.3</i>	15.2	13.2	6.3	9.0	8.9
	Sicilian Channel	0.89	4.1	<i>3.9</i>	13.8	12.1	5.7	4.3	5.1
	Middle Adriatic	0.88	<i>3.8</i>	4.0	8.7	11.4	6.5	4.4	8.3
	Eastern Ionian Sea	0.95	4.3	<i>3.3</i>	9.6	13.3	3.7	5.2	4.0
	Aegean Sea	0.90	6.3	5.4	<i>5.1</i>	7.2	7.8	11.8	11.6
	Taxon level model	0.90	4.1	<i>3.8</i>	11.1	14.6	5.3	4.1	4.1
CALEDW	East Cape Verde	0.98	3.6	<i>3.3</i>	5.5	7.2	4.2	5.7	3.6
	West Cape Verde	0.99	<i>3.1</i>	3.5	5.2	6.5	5.5	4.9	4.5
		Taxon level model	0.98	<i>3.2</i>	3.7	6.8	5.6	4.3	3.9
Supra-taxon		0.78	4.8	<i>4.2</i>	9.1	18.5	6.5	4.9	7.2

To compare environmental preferences at a supra-taxon, taxon and population level we extracted the original value of the most important environmental variables at the presence points. At the supra-taxon model the most important variable was CHLA (median = 0.30 mg·m⁻³, CI = 0.08 – 2.78). At a taxon level, CALBOR preferred the lowest SST (19.4 °C [11.2 – 24.0]), while CALEDW chose more temperate values (22.0 °C [17.8 – 24.9]). For CALDIO the most important variable was CHLA, choosing lower values than the other two taxa (0.23 mg·m⁻³ [0.10 – 2.83]; Table 10; Fig. A8). At a population level, CHLA and SST were identified as the most important variables in the model in all but one model (for the Aegean Sea population, for which CHLA_var was most important). Median values of SST were similar for all CALBOR and CALEDW populations, all within a 2.1 °C and 0.1 °C range respectively, while there was more variability among the CALDIO populations (range of 4.1 °C). Within these, the Eastern Ionian Sea population selected the highest values (27.2 °C [2.4 – 28.8]) and Palomas the lowest (23.1 °C [18.1 – 28.0]). (Table 10 & Fig. A9). The selected CHLA values were in a range of 0.1 mg·m⁻³ for CALBOR, a range of 0.3 mg·m⁻³ for CALDIO and 0.2 mg·m⁻³ for CALEDW. Again, the maximum variability is among the CALDIO populations, with the minimum values selected by the Eastern Ionian Sea population (0.22 mg·m⁻³ [0.12 – 1.14]) and the maximum by the Palomas population (0.49 mg·m⁻³ [0.09 – 5.00]). The Aegean Sea population selected the highest value of CHLA_var (88.9% [20.1 – 97.6]; Table 10 & Fig. A9).

Pairwise correlations (representing similarity) among colony-level projections ranged between -0.08 and 0.90 for CALBOR and between 0.37 and 0.94 for CALDIO. The linear models showed that the similarity

between model projections was negatively related to distance between colonies for the two taxa, with significant correlations for both, and R-squared values of 0.57 and 0.40 respectively (Fig. 11).

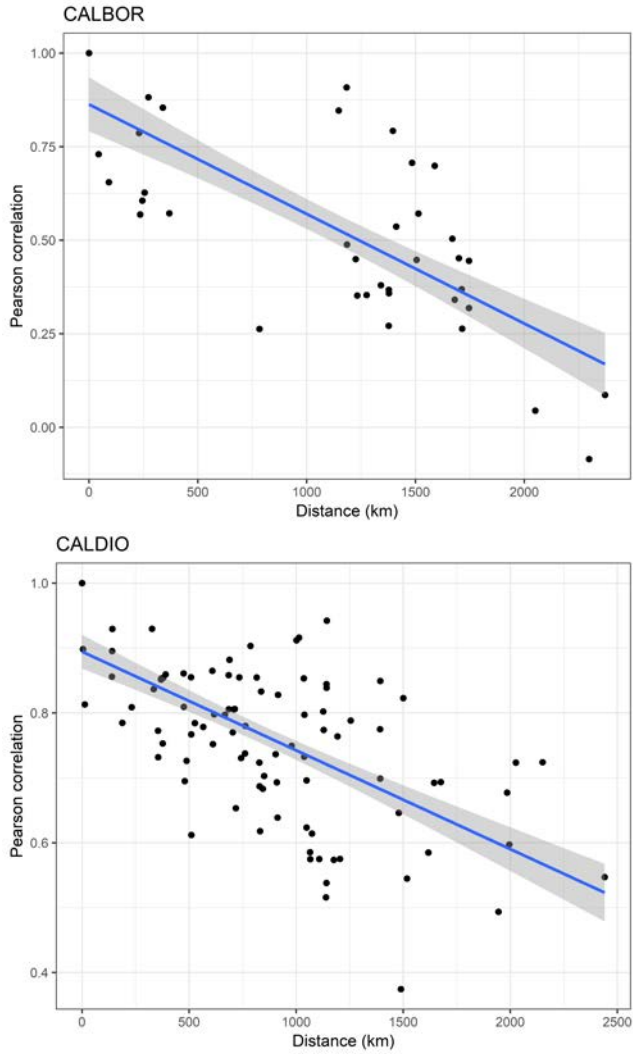


Figure 11

correlation of pairwise distances between colonies and similarity between respective non-breeding habitat model projections for CALBOR (a) and CALDIO (b). adjusted R-squared values of 0.57 and 0.40 respectively

Table 10

Median and 95% confidence interval (CI) of the values selected for each variable for each of the models at a colony, population and species level. Only the variables that appear as most important in at least one of the models are shown, and the most important environmental variable for each model is shown in bold

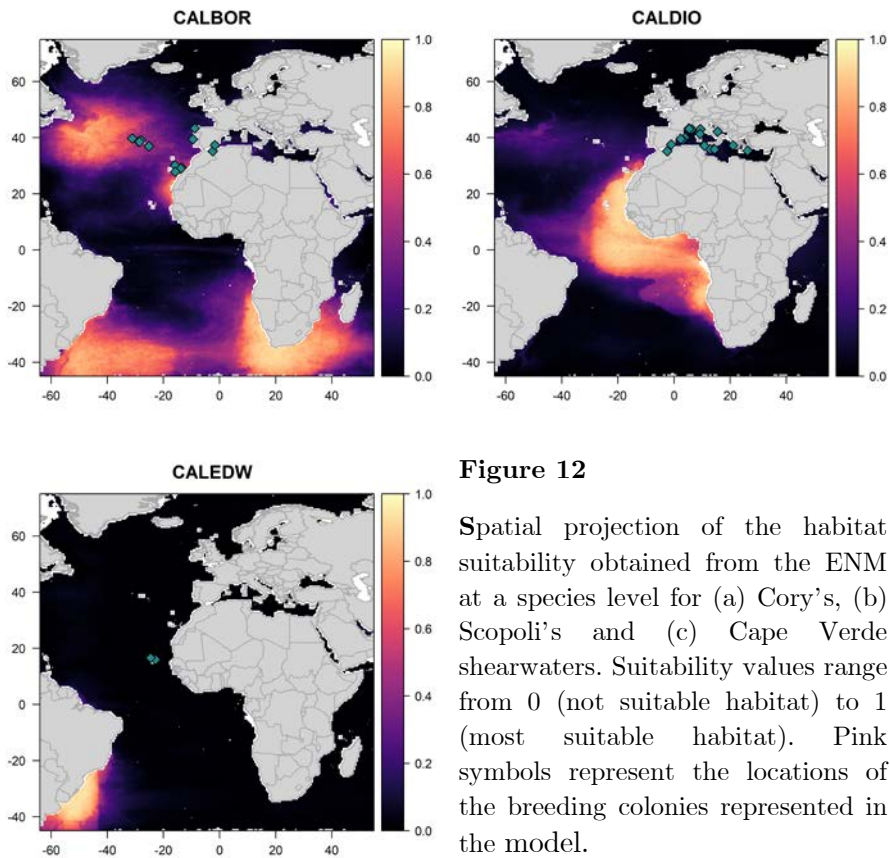
Group	Population	SST		CHLA		CHLA var	
		Median	95% CI	Median	95% CI	Median	95% CI
CALBOR	Central Azores Is.	19.4	11.2 - 24.0	0.35	0.08 - 0.87	40.8	20.5 - 76.1
	East Azores Is.	19.1	6.7 - 25.8	0.31	0.05 - 1.06	55.1	20.4 - 85.2
	Chafarinas Is.	21.2	12.4 - 28.7	0.38	0.08 - 4.76	59.7	22.1 - 97.3
	Berlengas Is.	19.3	11.3 - 25.2	0.39	0.09 - 2.05	43.5	19.9 - 83.9
	Salvages Is.	19.8	11.2 - 25.7	0.40	0.09 - 2.18	42.8	19.9 - 85.8
	East Canary Is.	21.0	15.2 - 27.8	0.38	0.08 - 2.90	47.5	20.2 - 90.4
	West Canary Is.	20.5	16.0 - 27.4	0.36	0.07 - 3.39	45.4	20.3 - 87.9
	Taxon level model	19.4	9.6 - 25.8	0.35	0.07 - 1.45	43.7	20.4 - 82.0
CALDIO	Palomas	23.1	18.1 - 28.0	0.49	0.09 - 5.00	76.9	24.4 - 97.5
	Balearic Is.	23.3	17.4 - 28.7	0.32	0.08 - 4.36	71.9	25.8 - 97.2
	France	25.1	18.4 - 28.8	0.26	0.06 - 2.62	77.7	30.0 - 97.5
	Corsica - Sardinia	24.9	18.3 - 28.8	0.28	0.10 - 2.94	79.1	31.2 - 97.2
	Sicilian Channel	25.7	17.3 - 28.8	0.23	0.08 - 4.64	67.0	20.7 - 97.2
	Middle Adriatic	23.1	15.8 - 28.6	0.34	0.09 - 4.31	71.9	22.0 - 97.4
	Eastern Ionian Sea	27.2	22.4 - 28.8	0.22	0.12 - 1.14	72.3	21.0 - 97.6
	Aegean Sea	23.3	10.3 - 28.5	0.33	0.08 - 5.71	88.9	20.1 - 97.6
Taxon level model	26.4	19.0 - 28.8	0.23	0.10 - 2.83	69.3	22.1 - 97.2	
CALEDW	East Cape Verde	22.1	18.9 - 24.7	0.86	0.18 - 5.33	69.3	47.0 - 86.3
	West Cape Verde	22.2	19.1 - 24.8	1.08	0.18 - 6.67	69.5	43.6 - 86.3
	Taxon level model	22.0	17.8 - 24.9	0.59	0.16 - 5.33	63.5	40.7 - 86.3
Supra-taxon		21.7	44.4 - 28.6	0.30	0.08 - 2.78	55.0	20.6 - 96.6

3.5 Discussion

In this study we use migratory connectivity and ecological niche modelling (ENM) to analyse ecological divergence between three closely related taxa. Through the migratory connectivity analysis, we detected structure at a population level within each taxon, which was confirmed through the ENM, in both spatial distribution and environmental preferences. In addition, our results showed a clear spatial and ecological segregation in the non-breeding distributions of the three taxa. These results have important implications for population dynamics and structure, as well as management and conservation.

3.5.1 Ecological segregation

Migratory connectivity and ENM results, showed a clear segregation among the three taxa: CALBOR and CALDIO clearly selected different non-breeding areas, which was reflected in the relatively high MC at a taxon level and visible in the geographical projections of the ENM. While optimal habitat for CALDIO was mainly between the 30°N and 30°S, with a clear tropical distribution, CALBOR had most of its optimal habitat outside these limits, with a clear sub-tropical distribution (Fig. 12). Although CALEDW shared non-breeding areas with CALBOR, the two segregated environmentally, with CALEDW using warmer, but also more productive and dynamic waters, which suggested ecological segregation between them, rather than a spatial segregation.



Despite the clear non-breeding allopatry between CALBOR and CALDIO, there were a few “contact zones” between them: in the main non-breeding grounds, both taxa co-occurred in the transition area between the Sahelian and Saharan upwellings as well as along the Namib coast. In addition, CALDIO also frequented subtropical regions of the Brazil current and the North Atlantic region during their migratory movements, sharing them with CALBOR. The existence of these “contact zones” favours potential events of dispersion: a bird of one population following in the return migration from common non-breeding grounds a bird of another population from the same or a different taxon (Friesen et al. 2007). Indeed, there are records of CALBOR individuals

ringed as chicks in Atlantic colonies and found years later breeding in Mediterranean colonies (Lo Valvo and Massa 1988; Thibault and Bretagnolle 1998; Gómez-Díaz et al. 2009; Gabirot et al. 2016) including some cases of successful inter-taxa pairing (Martínez-Abraín et al. 2002, J. Navarro pers. comm.). In contrast to the occasional dispersal of CALBOR individuals to CALDIO breeding sites, there are no reported dispersal events in the opposite direction (CALDIO birds breeding in CALBOR sites), and although CALEDW and CALBOR completely overlapped in their non-breeding areas, there are only a handful of documented cases of CALEDW individuals seen in CALBOR colonies, all of them in the Canary and Salvagem Islands (Fagundes et al. 2012; Copete et al. 2015; Gil-Velasco et al. 2018), but none of them with confirmed breeding. Finally, when we modelled the non-breeding distribution of the three taxa together, the model had a worse fit than when modelled separately. This is an indication that the three taxa occupied different ecological niches with little overlap among them, which made any attempt at modelling the ecological niche of the three species together less accurate, which can be interpreted as an evidence of lineage divergence (Raxworthy et al. 2007).

The Almería – Orán oceanographic front (AOOF; Fig. 6), which serves as ecological divide between Atlantic and Mediterranean distributions of many marine taxa (Pascual et al. 2017), also appears to be the divide between CALBOR and CALDIO breeding distributions, with CALBOR breeding in the Atlantic, West of the front, and CALDIO breeding inside the Mediterranean, East of the front (Gómez-Díaz et al. 2009). However, colonies of both taxa are located nearby in both sides of AOOF, with at least one mixed colony (Gómez-Díaz et al. 2006; Zidat et al. 2017). This

partial sympatry, together with the breeding synchrony between the two taxa, opens the possibility of hybridization and introgression producing gene flow among them (Brown et al. 2015), which suggests that reproductive isolation might not yet be complete. Indeed, genetic studies in the mixed colony of Chafarinas, based on both mitochondrial DNA (mtDNA, Gómez-Díaz et al. 2009) and microsatellites (Zidat et al. 2017), reported the existence of introgressed individuals which could indicate relatively-recent hybridisation events. Our study includes tracks from three colonies in the AOOFF area: Terreros, Palomas, and Chafarinas. Projections of the non-breeding habitat for CALDIO from Palomas and CALBOR from Chafarinas (Figs. A10 & A11) showed that the optimal habitat for individuals breeding in these two colonies included areas of suitable habitat for both taxa. Unfortunately, the non-breeding habitat for CALBOR from Terreros and the CALDIO from Chafarinas could not be modelled because of our small sample size (see Chapter 1). However, most of the non-breeding locations of CALBOR from Terreros fell within the Sahelian and Saharan upwellings (Fig. A13), which are predominantly used by CALDIO populations (Fig. 7). The non-breeding locations of CALDIO from Chafarinas occupied mainly tropical non-breeding areas of the Eastern Atlantic coast, typical of their taxon, but also the Brazil coast and North Atlantic pelagic area, which are mainly used by CALBOR populations. Although the small sample size for these two colonies, makes it difficult to make proper inference, these results suggest that, although the two taxa have clearly segregated non-breeding ecological niches, there is some locally-restricted genetic mixture in the small area of the AOOFF, where both taxa still breed in sympatry (as had already been reported in Gómez-Díaz et al. 2006, Flood and

Gutiérrez 2019), producing an intermediate phenotype between that of CALBOR and CALDIO for individuals breeding in this area.

Our results suggested that the three *Calonectris* taxa adapted to local ecological niches in both the non-breeding and breeding distribution. Even though the three taxa breed synchronically, local adaptation coupled with the strong philopatry and breeding site fidelity possibly resulted in a process of ecological divergence, together with the divergence of other traits previously reported in other studies (morphology, vocalisations, odour, Zidat et al. 2017) conducive to species differentiation.

3.5.2 Spatial structure (within a given taxon)

The differences in migratory connectivity at a colony and population levels for the three taxa evidenced that, while there was a high degree of spread and mixing between colonies (Finch et al. 2017), birds from a given population of CALBOR or CALDIO selected a subset of non-breeding areas among all those used for all populations of each taxon, thus pointing out some degree of population structuring along the colonies within each taxa. Unfortunately, since we have only sampled one colony of CALEDW in each of its two breeding populations, we could not study its non-breeding distribution structure at the level of population.

Our results have structured CALBOR populations in three main groups according to common non-breeding areas: (1) Salvages and Berlengas populations, (2) the Azorean populations, and (3) the Canarian populations (Figs. 8 & A10). For CALDIO, our results showed that non-breeding habitat from the Eastern populations differed from that of

Central and Western populations in that they used almost exclusively the northernmost non-breeding areas (Figs. 9 & A11). For CALEDW, although we only had data from one colony from each population, the use of non-breeding areas of the two sampled colonies was almost identical, hinting to a lack of structure in non-breeding distributions (Figs. Y5 & A12). There are studies showing a size gradient within CALBOR populations, with a slight increase in size with latitude, (Gómez-Díaz et al. 2006; Gómez-Díaz et al. 2009), and within CALDIO populations, with a more substantial decrease in size with longitude (Massa and Lo Valvo 1986; Gómez-Díaz et al. 2006), but no structure has been found in any other phenotypical markers for neither of them (e.g., acoustic and chemical markers; Bretagnolle and Lequette 1990, Gabriot et al. 2016). Although there is genetic structure defined for CALBOR, the results suggest a different spatial population structure to the one we reported here (Genovart et al. 2013; Munilla et al. 2016). For CALDIO, Gómez-Díaz et al. (2009) found that genetic structure among colonies indicated divergence between Eastern and Western Mediterranean populations. Although no genetic evidences are available for CALEDW, since there are no phenotypical differences between colonies (neither in morphology and foraging areas [González-Solís pers. comm.], nor choice of non-breeding area, nor preferred environmental variables) it seems reasonable to assume that there is no population structuring in this locally-restricted taxon endemic of Cape Verde.

The presence of structure (i.e., high MC) at a population level but not at a colony level demonstrated that the spatial structure in the breeding distributions of CALBOR and CALDIO is maintained during the non-breeding distribution only at a population level. This has obvious

implications in population dynamics, but also in conservation (Esler 2000). Populations were defined as groups of colonies with overlapping 100 km radius buffers, which hints to the possibility of a distance-based structure (i.e., neighbouring colonies have more similar behaviour than distant colonies because there is more gene flow among them). The negative correlation between pairwise similarity of non-breeding habitat and between-colony distance provides further support for this theory, since it demonstrates that non-breeding habitat is more similar the closer the corresponding breeding colonies are. This “isolation by distance” model had already been proposed as driver of the differentiation between CALBOR and CALDIO, and as a possible driver of population structure within CALDIO (Gómez-Díaz et al. 2009). Certainly, understanding what constitutes a population, and how they relate to each other demographically and genetically is key to quantify the effect that local (nest availability, predation at colony) or global (climate change, food availability) perturbations in the breeding distribution might have on the population dynamics (Matthiopoulos et al. 2005). In addition, understanding the degree of migratory connectivity among populations can also advance our understanding of the effects of perturbations in the non-breeding distribution: in situations of low migratory connectivity (high mixing among populations and high population spread), any perturbation in a given non-breeding region is likely to cause a global effect to the entire species, while the effect would be more localised in a species with high connectivity. Finally, the degree of intra-population spread in the non-breeding distribution can affect the ability of the population to respond to non-breeding range shifts and reductions in a context of climate change (Finch et al. 2017).

3.5.3 Concluding remarks

To our knowledge, this is the first time that non-breeding distributions have been used to identify the degree of ecological divergence in a species complex of migratory birds, extending the previous attempts of applying ENM to support lineage divergence in resident species (Raxworthy et al. 2007; Rissler and Apodaca 2007). Our evidence on major habitat segregation among the three taxa during the non-breeding period increase the list of phenotypical differences supporting the divergence between CALBOR and CALDIO that have been documented so far: morphological (Gómez-Díaz et al. 2009), acoustic (Bretagnolle and Lequette 1990), chemical (Zidat et al. 2017), and now ecological (this study). In addition, we showed that although CALBOR and CALEDW partly overlap in their non-breeding distribution, their environmental preferences differ, which confirms the ecological segregation. We also provide evidence of intermediate habitat preferences in the non-breeding distributions of birds breeding in the contact zone between CALBOR and CALDIO, which is consistent with previous evidences of hybridization and introgression (Gómez-Díaz et al. 2009). Finally, our analyses of migratory connectivity and modelling of non-breeding habitats detected the presence of structure within the CALBOR and CALDIO taxa, present at a population level, rather than at a colony level, thus calling for a separate management and conservation of these populations.

3.6 References

Ambrosini R, Møller AP, Saino N. 2009. A quantitative measure of migratory connectivity. *J Theor Biol.* 257(2):203–211. doi:10.1016/j.jtbi.2008.11.019.

- Baddeley A, Rubak E, Turner R. 2015. Spatial Point Patterns: methodology and applications with R. New York: Chapman and Hall/CRC. <https://www.taylorfrancis.com/books/9781482210217>.
- Bauer S, Lisovski S, Hahn S. 2016. Timing is crucial for consequences of migratory connectivity. *Oikos*. 125(5):605–612. doi:10.1111/oik.02706. <http://doi.wiley.com/10.1111/oik.02706>.
- Bivand R, Rundel C. 2019. rgeos: Interface to Geometry Engine - Open Source ('GEOS'). <https://cran.r-project.org/package=rgeos>.
- Bretagnolle V, Lequette B. 1990. Structural variation in the call of the Cory's shearwater (*Calonectris diomedea*, Aves, Procellariidae). *Ethology*. 85:313–323.
- Brown RM, Techow NMSM, Wood AG, Phillips RA. 2015. Hybridization and Back-Crossing in Giant Petrels (*Macronectes giganteus* and *M. halli*) at Bird Island, South Georgia, and a Summary of Hybridization in Seabirds. Peter H-U, editor. *PLoS One*. 10(3):e0121688. doi:10.1371/journal.pone.0121688. [accessed 2018 Apr 5]. <http://dx.plos.org/10.1371/journal.pone.0121688>.
- Burg TM, Croxall JP. 2001. Global relationships amongst black-browed and grey-headed albatrosses: analysis of population structure using mitochondrial DNA and microsatellites. *Mol Ecol*. 10(11):2647–2660. doi:10.1046/j.0962-1083.2001.01392.x. <http://doi.wiley.com/10.1046/j.0962-1083.2001.01392.x>.
- Calenge C. 2011. Home Range Estimation in R: the adehabitatHR Package. *R vignette*.:1–61. doi:10.1111/j.1365-2656.2006.01186.x.
- Cohen EB, Hostetler JA, Hallworth MT, Rushing CS, Sillett TS, Marra PP. 2017. Quantifying the strength of migratory connectivity. *Methods Ecol Evol*.:1–12. doi:10.1111/2041-210X.12916. [accessed 2017 Nov 9]. <http://onlinelibrary.wiley.com/doi/10.1111/2041-210X.12916/abstract>.
- Copete JL, Lorenzo JA, Amengual E, Bigas D, Fernández P, López-Velasco D, Rodríguez G, García-Tarrasón M. 2015. Observaciones de Aves Raras en España, 2012 y 2013. *Ardeola*. 62(2):453–508.
- Dias MP, Granadeiro JP, Phillips RA, Alonso H, Catry P. 2011. Breaking the routine: individual Cory's shearwaters shift winter destinations between hemispheres and across ocean basins. *Proc R Soc B*. 278(1713):1786–

1793. doi:10.1098/rspb.2010.2114.
<http://rspb.royalsocietypublishing.org/cgi/doi/10.1098/rspb.2010.2114>
 .
- Ekstrom PA. 2004. An advance in geolocation by light. *Mem Natl Inst Polar Res.* 58:210–226.
- Esler D. 2000. Applying Metapopulation Theory to Conservation of Migratory Birds. *Conserv Biol.* 14(2):366–372. doi:10.1046/j.1523-1739.2000.98147.x. papers://5ad192e1-39fa-4ac6-a9e7-61ea4bf4b48e/Paper/p948.
- Fagundes I, Pérez C, Moniz F. 2012. Cape Verde Shearwater in Selvagens in April 2012. *Dutch Bird.* 34(5):310–311.
- Fielding AH, Bell JF. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ Conserv.* 24(1):38–49. doi:10.1017/S0376892997000088.
- Finch T, Butler SJ, Franco AMA, Cresswell W. 2017. Low migratory connectivity is common in long-distance migrant birds. Chapman J, editor. *J Anim Ecol.* 86(3):662–673. doi:10.1111/1365-2656.12635. <http://doi.wiley.com/10.1111/1365-2656.12635>.
- Fišer C, Robinson CT, Malard F. 2018. Cryptic species as a window into the paradigm shift of the species concept. *Mol Ecol.* 27(3):613–635. doi:10.1111/mec.14486. <http://doi.wiley.com/10.1111/mec.14486>.
- Flood RL, Gutiérrez R. 2019. Status of Cory’s Shearwater in the western Mediterranean. *Dutch Bird.* 41(3):159–165.
- Friesen VL. 2015. Speciation in seabirds: why are there so many species...and why aren’t there more? *J Ornithol.* doi:10.1007/s10336-015-1235-0. <http://link.springer.com/10.1007/s10336-015-1235-0>.
- Friesen VL, Burg TM, McCoy KD. 2007. Mechanisms of population differentiation in seabirds. *Mol Ecol.* 16(9):1765–1785. doi:10.1111/j.1365-294X.2006.03197.x. <http://doi.wiley.com/10.1111/j.1365-294X.2006.03197.x>.
- Gabirot M, Raux L, Dell’Ariccia G, Bried J, Ramos R, González-Solís J, Buatois B, Crochet PA, Bonadonna F. 2016. Chemical labels differ between two closely related shearwater taxa. *J Avian Biol.* 47(4):540–

551. doi:10.1111/jav.00853.

- Genovart M, Thibault J-C, Igual JM, Bauzá-Ribot MDM, Rabouam C, Bretagnolle V. 2013. Population Structure and Dispersal Patterns within and between Atlantic and Mediterranean Populations of a Large-Range Pelagic Seabird. Votier S, editor. PLoS One. 8(8):e70711. doi:10.1371/journal.pone.0070711. <https://dx.plos.org/10.1371/journal.pone.0070711>.
- Gil-Velasco M, Rouco M, Ferrer J, García-Tarrasón M, García-Vargas FJ, Gutiérrez A, Hevia R, López F, López-Velasco D, Ollé À, et al. 2018. Observaciones de Aves Raras en España, 2016. Ardeola. 65(1):97–139. doi:10.13157/arla.65.1.2018.rb. <http://www.bioone.org/doi/10.13157/arla.65.1.2018.rb>.
- Gómez-Díaz E, González-Solís J, Peinado MA. 2009. Population structure in a highly pelagic seabird, the Cory's shearwater *Calonectris diomedea*: an examination of genetics, morphology and ecology. Mar Ecol Prog Ser. 382:197–209.
- Gómez-Díaz E, González-Solís J, Peinado MA, Page RDM. 2006. Phylogeography of *Calonectris* shearwaters using molecular and morphometric data. Mol Phylogenet Evol. 41:322–332.
- González-Solís J, Croxall JP, Oro D, Ruiz X. 2007. Trans-equatorial migration and mixing in the wintering areas of a pelagic seabird. Front Ecol Environ. 5(6):297–301. doi:10.1890/1540-9295(2007)5[297:TMAMIT]2.0.CO;2. [accessed 2017 Jul 8]. [http://doi.wiley.com/10.1890/1540-9295\(2007\)5\[297:TMAMIT\]2.0.CO;2](http://doi.wiley.com/10.1890/1540-9295(2007)5[297:TMAMIT]2.0.CO;2).
- Granadeiro JP, Burns MD, Furness RW. 1999. Food provisioning to nestling shearwaters: why parental behaviour should be monitored? Anim Behav. 57:662–671.
- Granadeiro JP, Dias MP, Rebelo R, Santos CD, Catry P. 2006. Numbers and Population Trends of Cory's Shearwater *Calonectris diomedea* at Selvagem Grande, Northeast Atlantic. Waterbirds. 29(1):56–60. doi:10.1675/1524-4695(2006)29[56:NAPTOC]2.0.CO;2.
- Gu Z, Gu L, Eils R, Schlesner M, Brors B. 2014. Circlize implements and enhances circular visualization in R. Bioinformatics. 30(19):2811–2812.

doi:10.1093/bioinformatics/btu393.

- Hanski I, Gilpin M, editors. 1997. *Metapopulation Biology*. Elsevier.
<https://linkinghub.elsevier.com/retrieve/pii/B9780123234452X50007>.
- Hazevoet CJ. 1995. *The birds of the Cape Verde Islands. An annotated check list*. London, UK: British-Ornitologist Union.
- Hostetler JA, Hallworth MT. 2018. *MigConnectivity: Estimate Strength of Migratory Connectivity for Migratory Animals*. R package.
- International B. 2019. *Birdlife International Seabird Tracking Database | Tracking Ocean Wanderers*. [accessed 2018 Dec 13].
<http://www.seabirdtracking.org/>.
- Lisovski S, Hahn S. 2012. *GeoLight - processing and analysing light-based geolocator data in R*. *Methods Ecol Evol.* 3(6):1055–1059.
doi:10.1111/j.2041-210X.2012.00248.x.
- Louzao M, Pinaud D, Péron C, Delord K, Wiegand T, Weimerskirch H. 2011. *Conserving pelagic habitats: Seascape modelling of an oceanic top predator*. *J Appl Ecol.* 48(1):121–132. doi:10.1111/j.1365-2664.2010.01910.x.
- Marmion M, Parviainen M, Luoto M, Heikkinen RK, Thuiller W. 2009. *Evaluation of consensus methods in predictive species distribution modelling*. *Divers Distrib.* 15(1):59–69. doi:10.1111/j.1472-4642.2008.00491.x. <http://doi.wiley.com/10.1111/j.1472-4642.2008.00491.x>.
- Martínez-Abraín A, Sánchez A, Oro D. 2002. *Atlantic Cory's Shearwaters Breeding in a Colony of Mediterranean Cory's Shearwaters. Waterbirds.* 25(2):221–224. doi:10.1675/1524-4695(2002)025[0221:ACSBIA]2.0.CO;2.
- Massa B, Lo Valvo M. 1986. *Biometrical and biological considerations on the Cory's shearwater *Calonectris diomedea**. *NATO ASI Ser. G* 12:293–313.
- Matthiopoulos J, Harwood JL, Thomas L. 2005. *Metapopulation consequences of site fidelity for colonially breeding mammals and birds*. *J Anim Ecol.* 74(4):716–727. doi:10.1111/j.1365-2656.2005.00970.x.
- McConnell BJ, Chambers C, Fedak MA. 1992. *Foraging ecology of southern*

elephant seals in relation to the bathymetry and productivity of the Southern Ocean. *Antarct Sci.* 4:393–398.

Mougin J-L, Granadeiro JP, Jouanin C, Roux F. 1999. Philopatry and faithfulness to nest site in Cory's Shearwaters *Calonectris diomedea* at Selvagem Grande. *Ostrich.* 70(3–4):229–232. doi:10.1080/00306525.1999.9634242. <https://www.tandfonline.com/doi/full/10.1080/00306525.1999.9634242>

Mougin J-L, Jouanin C, Roux F, Zino F. 2000. Fledging weight and juvenile survival of Cory's shearwaters *Calonectris diomedea* on Selvagem Grande. *Ring. Migr.* 20:107–110.

Munilla I, Genovart M, Paiva VH, Velando A. 2016. Colony Foundation in an Oceanic Seabird. *PLoS One.* 11(2):e0147222. doi:10.13140/RG.2.1.2966.0248.

Navarro J, González-Solís J. 2007. Experimental increase of flying costs in a pelagic seabird: effects on foraging strategies, nutritional state and chick condition. *Oecologia.* 151(1):150–160.

Paiva VH, Geraldes P, Rodrigues I, Melo T, Melo J, Ramos JA. 2015. The foraging ecology of the endangered Cape Verde shearwater, a sentinel species for marine conservation off West Africa. *PLoS One.* 10(10):e0139390. doi:10.1371/journal.pone.0139390. <http://dx.plos.org/10.1371/journal.pone.0139390>.

Pascual M, Rives B, Schunter C, Macpherson E. 2017. Impact of life history traits on gene flow: A multispecies systematic review across oceanographic barriers in the Mediterranean Sea. *PLoS One.* 12(5):1–20. doi:10.1371/journal.pone.0176419.

Quillfeldt P, Moodley Y, Weimerskirch H, Cherel Y, Delord K, Phillips RA, Navarro J, Calderón L, Masello JF. 2017. Does genetic structure reflect differences in non-breeding movements? A case study in small, highly mobile seabirds. *BMC Evol Biol.* 17(1):160. doi:10.1186/s12862-017-1008-x. [accessed 2017 Nov 30]. <http://bmcevolbiol.biomedcentral.com/articles/10.1186/s12862-017-1008-x>.

R Core Team. 2019. R: A language and environment for statistical computing.

<https://www.r-project.org/>.

- Ramos R, Granadeiro JP, Nevoux M, Mougin J-L, Dias MP, Catry P. 2012. Combined spatio-temporal impacts of climate and longline fisheries on the survival of a trans-equatorial marine migrant. *PLoS One*. 7(7):1–12. doi:10.1371/journal.pone.0040822.
- Raxworthy CJ, Ingram CM, Rabibisoa N, Pearson RG. 2007. Applications of Ecological Niche Modeling for Species Delimitation: A Review and Empirical Evaluation Using Day Geckos (*Phelsuma*) from Madagascar. Weins J, editor. *Syst Biol*. 56(6):907–923. doi:10.1080/10635150701775111. <https://academic.oup.com/sysbio/article/56/6/907/1654899>.
- Reyes-González JM, Zajková Z, Morera-Pujol V, De Felipe F, Militão T, Dell’Ariccia G, Ramos R, Igual JM, Arcos JM, González-Solís J. 2017. Migración y ecología espacial de las poblaciones españolas de pardela cenicienta. Migra M nº 3 del programa, editor. Madrid: SEO/BirdLife.
- Rissler LJ, Apodaca JJ. 2007. Adding More Ecology into Species Delimitation: Ecological Niche Models and Phylogeography Help Define Cryptic Species in the Black Salamander (*Aneides flavipunctatus*). Weins J, editor. *Syst Biol*. 56(6):924–942. doi:10.1080/10635150701703063. <https://academic.oup.com/sysbio/article/56/6/924/1655548>.
- Sangster G, Collinson JM, Crochet P-A, Knox AG, Parkin DT, Votier SC. 2012. Taxonomic recommendations for British birds: eighth report. *Ibis (Lond 1859)*. 154(4):874–883. doi:10.1111/j.1474-919X.2012.01273.x. <http://doi.wiley.com/10.1111/j.1474-919X.2012.01273.x>.
- Schmitt S, Pouteau R, Justeau D, de Boissieu F, Birnbaum P. 2017. ssdm: An r package to predict distribution of species richness and composition based on stacked species distribution models. *Methods Ecol Evol*. 8(12):1795–1803. doi:10.1111/2041-210X.12841.
- Segurado P, Araújo MB, Kunin WE. 2006. Consequences of spatial autocorrelation for niche-based models. *J Appl Ecol*. 43(3):433–444. doi:10.1111/j.1365-2664.2006.01162.x. <http://doi.wiley.com/10.1111/j.1365-2664.2006.01162.x>.
- Spalding MD, Fox HE, Allen GR, Davidson N, Ferdaña ZA, Finlayson M, Halpern BS, Jorge MA, Lombana A, Lourie SA, et al. 2007. Marine

- Ecoregions of the World: A Bioregionalization of Coastal and Shelf Areas. *Bioscience*. 57(7):573–583. doi:10.1641/B570707. <https://academic.oup.com/bioscience/article-lookup/doi/10.1641/B570707>.
- Thibault J-C, Bretagnolle V. 1998. A Mediterranean breeding colony of Cory's shearwater *Calonectris diomedea* in which individuals show behavioural and biometric characters of the Atlantic subspecies. *Ibis* (Lond 1859). 140:253–258.
- Thibault J-C, Bretagnolle V, Rabouam C. 1997. Cory's shearwater. In: *Birds of the Western Palearctic Update*. Oxford: Oxford University Press. p. 75–98.
- Thuiller W, Lafourcade B, Engler R, Araújo MB. 2009. BIOMOD - A platform for ensemble forecasting of species distributions. *Ecography* (Cop). 32(3):369–373. doi:10.1111/j.1600-0587.2008.05742.x.
- Tigano A, Damus M, Birt TP, Morris-Pocock JA, Artukhin YB, Friesen VL. 2015. The arctic: Glacial refugium or area of secondary contact? Inference from the population genetic structure of the thick-billed murre (*Uria lomvia*), with Implications for management. *J Hered*. 106(3):238–246. doi:10.1093/jhered/esv016.
- Lo Valvo M, Massa B. 1988. Considerations on a specimen of Cory's shearwater ringed at Selvagem Grande and recovered in the central Mediterranean. *Bocagiana*. 124:1–5.
- Wakefield ED, Phillips RA, Matthiopoulos J. 2009. Quantifying habitat use and preferences of pelagic seabirds using individual movement data: a review. *Mar Ecol Prog Ser*. 391:165–182.
- Warham J. 1996. *The Behaviour, Population Biology and Physiology of the Petrels*. London: Academic Press Ltd. <https://linkinghub.elsevier.com/retrieve/pii/B9780127354156X50005>.
- Warham J. 1997. *The Behaviour, Population Biology and Physiology of the Petre*. London: Academic Press Inc. https://www.cambridge.org/core/product/identifier/S0032247400014480/type/journal_article.
- Webster MS, Marra PP, Haig SM, Bensch S, Holmes RT. 2002. Links between worlds: unraveling migratory connectivity. *Trends Ecol Evol*. 17:76–83.

- Wilson RP, Ducamp JJ, Rees WG, Culik BM, Niekamp K. 1992. Estimation of location: global coverage using light intensity. In: Wildlife telemetry: remote monitoring and tracking of animals. p. 131–134.
- Worton BJ. 1989. Kernel Methods for Estimating the Utilization Distribution in Home-Range Studies. *Ecology*. 70(1):164–168.
- Zidat T, Dell’Ariccia G, Gabirot M, Sourrouille P, Buatois B, Celerier A, Bonadonna F, Crochet P-A. 2017. Reproductive isolation maintains distinct genotypes, phenotypes and chemical signatures in mixed colonies of the two European *Calonectris* shearwaters (*Procellariiformes: Procellariidae*). *Zool J Linn Soc*. 181(3):711–726. doi:10.1093/zoolinnean/zlx002. <http://academic.oup.com/zoolinnean/article/181/3/711/3811184>.

4

Environmental and behavioural drivers of foraging segregation in a long-range seabird: insights from Bayesian spatial models within the INLA framework

4.1 Abstract

The foraging distributions of colonially breeding seabirds performing central place foraging depend, among other factors, on the at-sea distribution of resources, but also on the presence of conspecifics. At sea, prey is usually patchily distributed and inconspicuous from the surface, and to find it seabirds are thought to rely on memory and environmental cues, but also on information conveyed by conspecifics, either at sea (local enhancement) or at the colony (information centre hypothesis).

However, these events of information transfer are difficult to detect with commonly used modelling techniques. We present an improvement on the existent species distribution modelling techniques, based on point process modelling through continuously indexed Gaussian random fields within the INLA (Integrated Nested Laplace Approximation) framework. We use this method to model the segregation of the foraging areas of Cory's shearwaters (*Calonectris borealis*) GPS-tracked simultaneously from three neighbouring breeding colonies in the northeast Atlantic. First, an analysis of the overlap of core foraging areas found significant segregation among the foraging distributions of the three colonies, both in the areas surrounding the colonies and in the distant foraging grounds. Then, the point process models provided evidences that this segregation was not driven by environmental preferences only, detecting small-scale clustering indicative of local enhancement events, and a different spatial structure in the distribution of each colony that could be indicative of information transfer at the colonies shaping the foraging distributions. We provide a powerful and flexible modelling tool that had not yet been applied to tracking data, and demonstrate, for the first time to our knowledge, the information transfer-mediated spatial partitioning of foraging distributions in a long-ranging pelagic seabird.

4.2 Introduction

A fundamental goal in spatial ecology is understanding patterns in species distribution and abundance. For highly mobile predators, the search for resources is one of the main determinants of movement and, thus, of space use and distribution (Matthiopoulos 2003). In central place

foragers, the search for resources is highly conditioned by the need to commute between foraging areas and the central place (e.g. colony, nest, roost or burrow), as opposed to performing random, constant movements through the home range (Orians and Pearson 1979). This causes the energy costs of foraging to increase with distance between foraging and resting places (Papastamatiou et al. 2018), which makes central place foragers highly dependent on their prey search capabilities. For this reason behavioural and morphological characteristics that optimise foraging efficiency are essential, particularly in environments where the resources are unpredictable and patchily distributed (Trevail et al. 2019).

In the marine environment, while foraging areas at a large and medium scale (e.g. continental shelves, upwellings, oceanographic features) can be predictable and thus detectable following cues and memory, small scale foraging patches (e.g. schools of fish) are much more unpredictable (Boyd et al. 2016). This is particularly true for seabird species feeding on small epipelagic fish, which are distributed in spatially and temporally dynamic patches, inconspicuous from the surface. In addition to foraging in unpredictable environments, seabirds act as central place foragers during the breeding which makes foraging mechanisms are crucial for their survival (Trevail et al. 2019). At a large scale, birds can rely on memory to return to the area where they have previously found food (Matthiopoulos 2003, Weimerskirch 2007), and at a medium scale information on the location of highly productive areas can be obtained from oceanographic features such as sea surface temperature and thermal and chlorophyll-a fronts, or from olfactory cues (Dell'Araccia et al. 2014, Scales et al. 2014, van Eeden et al. 2016, Soldatini et al. 2019). In

addition, most seabirds breed colonially (96% of the species worldwide; Coulson 2002), which suggests that coloniality may have an adaptive function: it causes a high number of individuals foraging in common grounds, providing the perfect conditions for obtaining foraging cues from conspecifics (Buckley 1997). The information transfer can happen at sea: if individuals feeding on a food patch (i.e. a school of fish) are more conspicuous than the prey itself, they can provide cues for conspecifics on the presence of food, generating large foraging groups in a short amount of time in a process called local enhancement (Thiebault et al. 2014). The information transfer can also happen in the colonies: the information centre hypothesis proposes that colonially breeding animals can use public information provided by conspecifics, either willingly or inadvertently, to locate resources or favourable habitats (Ward and Zahavi 1972, Weimerskirch et al. 2010). Despite the advantages coloniality might have for information transfer, it also has consequences for competition, both within and between colonies. According to the hinterland model (Cairns 1989), (1) due to inter-colony competition colonies within foraging range of each other will have non-overlapping foraging areas, and; (2) due to intra-colony competition larger colonies will have larger foraging areas, since larger foraging areas should indicate more food availability, and thus capacity to sustain a large number of breeders. A modification of Cairn's model (termed density dependent hinterland model, DDH; Wakefield et al. 2013), which also contemplates the information transfer both at sea and at the colonies, has served to develop models and explanations about among colony segregations and colony sizes in different seabird species (Opperl et al. 2015, Jovani et al. 2015, Bolton et al. 2018). However, it makes

three assumptions that, although possibly suitable for seabird species with short foraging ranges, are hard to fulfil for long-ranging ones: (1) prey is distributed uniformly in the areas surrounding the colony, (2) distance is a limiting factor, so the probability of finding a foraging bird will decrease linearly with distance, and (3) all colonies at a relevant distance are accounted for. Thus, although the transfer of information has been proved in short ranging species of sulids and cormorant (Weimerskirch et al. 2010, Wakefield et al. 2013, Jones et al. 2018), the subject has been scarcely tackled in long ranging pelagic species such as shearwaters (but see Ceia et al. 2014, Genovart et al. 2017).

Understanding the drivers of seabird distributions is not easy, especially in long-ranging species, since the resolution at which environmental and geographical scales can be a limitation itself, and there is also a trade-off between the range of the study area and the resolution at which we can observe the ecological processes driving the distributions (Fritz et al. 2003, Benhamou 2014). Common species distribution modelling techniques have usually been employed, which involve summarising the location data as presences or counts in grid cells, and then relating each cell value to a set of environmental variables. However, the use of this methodology has three main drawbacks: first, this type of models are scale-dependent, since the results can vary depending on the cell size (Renner and Warton 2013); second, the information regarding each individual location is lost, since they are aggregated in each cell, which makes it impossible to study the clustering or segregation at scales smaller than the cell size, a key aspect for models aiming to capture bird aggregations caused by local enhancement (Aarts et al. 2012); lastly, the value at each cell (presence or count) is related to the values of the

environmental variables at the same cell, but the spatial autocorrelation is not being taken into account (i.e. “everything is related to everything else, but near things are more related than distant things”; Tobler 1970). Thus, although used to model spatial distribution of species, these models are not truly spatial models, and do not allow a detailed analysis of the spatial structure of the data.

A better framework for modelling spatial presence only data is point process methodology (Renner and Warton 2013). Analysing spatial data as point processes allows us to model the animal presences themselves, instead of aggregating data in cells. In addition, point processes can be modelled as a combination of fixed and random covariates, including a spatially structured field that can capture any structure in the original point pattern that is not explained by covariates. All these characteristics make point process methodology a very suitable tool for modelling species distributions (Illian and Burslem 2017). However, fitting most point process models is not possible in a frequentist framework, and Bayesian methods have been developed to fit the most complex models (Bakka et al. 2018). The statistical complexity of point process methodology, the difficulty of developing the models in the Bayesian framework, and their computational cost, have kept this type of models outside the ecologists’ toolbox (Illian and Burslem 2017). In recent years, the development of the Stochastic Partial Differential Equation (SPDE) approach within the Integrated Nested Laplace Approximation (INLA; Rue et al. 2009) framework has allowed to efficiently fit point process models in continuous space in a computationally efficient and accessible way (for non-experts) through the R packages `R-INLA` (Rue et al. 2009) and `inlabru` (Bachl et al.

2019). These continuously indexed models provide more flexibility and avoid the loss of information that is inevitable in the gridded approach and, crucially, can incorporate a spatial field able to capture unobserved spatial structure in the data, which could aid in the identification behavioural drivers of distributions, such as the transfer of information (Lindgren et al. 2011). These models have been used successfully to model spatial processes in the fields of sociology, economy, and also ecology (see examples in Bakka et al. 2018), but to date and to our knowledge, it has never been used to model foraging distributions from tracking data.

Calonectris shearwaters perform long, targeted foraging trips to remote productive areas, which indicates strong prey detecting capabilities. Indeed, it has been proven that they follow environmental (Ramos et al. 2013, Péron et al. 2018) and chemical cues (Dell’Ariccia et al. 2014) to find food. In addition, for both Cory’s (*C. borealis*) and Scopoli’s (*C. diomedea*) shearwaters, there is some evidence of foraging area segregation among nearby colonies (i.e. within foraging range of each other) that is unrelated to environmental variables (Ramos et al. 2013, Ceia et al. 2014, Reyes-González et al. 2017, Genovart et al. 2018). This segregation could be a mechanism to avoid inter-colony competition, and could be mediated by the transfer of information between conspecifics at the colonies or waters surrounding the colonies (Wakefield et al. 2013), although this behaviour is very difficult to prove empirically.

Combining GPS tracking data with remote sensing environmental data, and using continuously indexed point process models within the INLA framework, we aim to (1) test the overlap between the foraging areas of

Cory's shearwaters breeding in three neighbouring colonies, and (2) disentangle behavioural and environmental drivers of their foraging distribution. Similarly to what has been proven for *Calonectris* shearwaters from other colonies, we expect to find the effects of information transfer as a higher degree of segregation among the foraging distributions of the three species than what would be expected at random. In addition, we expect to find that, although environmental variables explain the distribution of birds at sea at large scales, smaller-scale structure to remain unexplained and captured in the spatial field, allowing us to detect the effects of information transfer, both at sea and at the colony.

4.3 Methods

4.3.1 Study species and area

The Cory's shearwater is a medium-sized procellariiform species that breeds colonially in islands and islets of the northeast Atlantic Ocean (Reyes-González et al. 2017). Females lay a single egg, and both partners share incubation and rearing duties (Thibault et al. 1997). As many Procellariiforms, chick rearing adults perform short foraging trips around the colony for chick-provisioning and fewer, longer trips for self-provisioning (Granadeiro et al. 1998). The study was conducted on three colonies of Cory's shearwaters in the Canary Islands. The Canary Islands are a volcanic archipelago 95 km off the northwest coast of Africa (in its nearest point). It holds ca. 14% of the world population of Cory's shearwaters in 13 colonies spread across all its islands and islets (Table A4). We sampled birds from breeding colonies on the islet of Montaña Clara (29.29° N, 13.53° W), in the Timanfaya National Park (Lanzarote;

28.99° N, 13.83° W) and in the Veneguera ravine (Gran Canaria; 27.84° N, -15.79°W) during the chick-rearing period of the 2015 breeding season (Table 11). Birds from the three colonies and other neighbouring Cory's shearwater colonies use the nearby waters of the Canary Current as foraging grounds (Fig. 13 and Table A4; Ramos et al. 2013), sharing it with many other species of seabirds and other top predators (e.g. Valdés and Déniz-González 2015, Grecian et al. 2016).

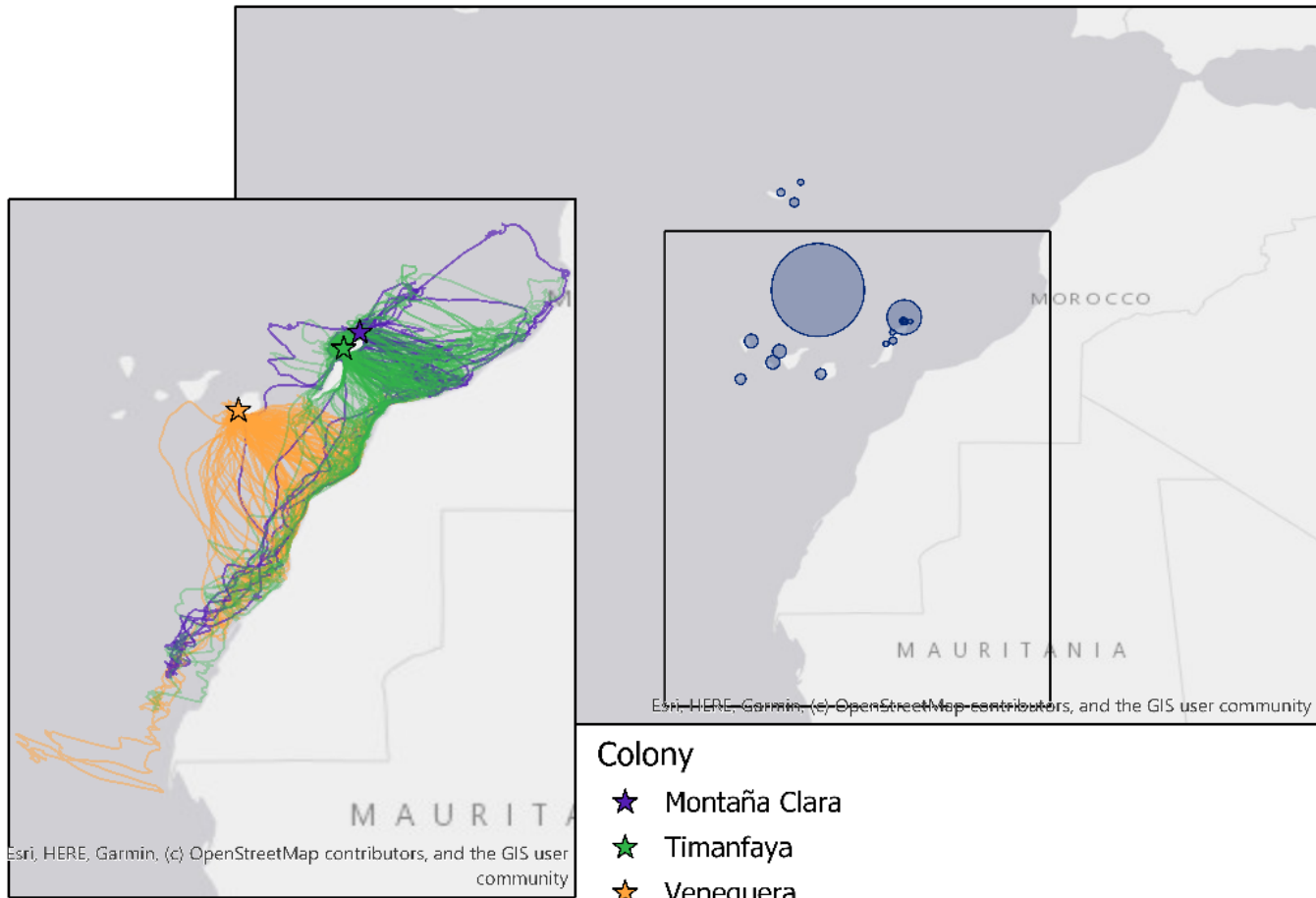


Table 11

Summary of main parameters for the three colonies of Montaña Clara, Timanfaya and Veneguera where we tracked breeding Cory's shearwaters (*C. borealis*). There is more than one trip for each bird, the number of positions is indicated in orders of magnitude, and the % Foraging indicates the proportion of total positions that are assigned to the foraging behaviour for each trip. Max. distance indicates the maximum distance from the colony reached (in km), duration indicates the total duration of the trip from the moment the bird leaves the colony until it returns (in hours) and total distance indicates the total distance traveled by the bird in each trip.

Colony	Birds	Trips	Positions	% Foraging	Max distance (km)	Duration (hours)	Total distance (km)
Montaña Clara	10	34	$\sim 8 \times 10^3$	39.3 (30.0, 66.1)	173.6 (23.4, 787.8)	38.2 (12.9, 260.9)	476.5 (80.1, 2645.6)
Timanfaya	29	171	$\sim 3 \times 10^4$	38.0 (18.2, 68.8)	121.7 (14.8, 567)	20.5 (10.1, 201.8)	270.7 (42.9, 2182.8)
Veneguera	30	110	$\sim 2 \times 10^4$	33.6 (17.4, 65.2)	227.5 (24.7, 456.9)	40.2 (12.9, 156.4)	550 (80.5, 1508.2)

Figure 13 (previous page)

Foraging tracks of Cory's shearwaters (*C. borealis*) breeding in the colonies of Montaña Clara, Timanfaya and Veneguera, in the Canary Islands obtained by GPS devices during the 2015 chick rearing period. Birds forage mainly in the Canary Current, over the continental shelf of the Western Sahara.

Blue circles in the background map indicate location and breeding population size of all the other colonies foraging in the Canary Current according to Ramos et al. 2013

4.3.2 *Environmental data*

Environmental variables that could have an effect in seabird foraging distribution were obtained from the appropriate on-line services as monthly means for the months of the chick-rearing season (July – September) and the three previous months (April –June), since environmental variables affecting marine productivity might have a lagged effect on top predators’ foraging distribution (Wakefield et al. 2009). The monthly means were averaged to obtain one data layer, and all layers were rescaled to the coarsest resolution. Sources of all the considered variables, units, and processing details can be found in Table 12. Prior to entering the covariates in the model, we checked their pairwise Pearson correlations and retained only variables with a correlation < 0.5 . Next, we performed preliminary generalised additive models (function `gam()` from `mgcv` package in R; Wood 2011) to pre-select the covariates that a priori had an effect on the point intensity. After this initial selection, from all variables considered (Table 12) we kept slope ($^{\circ}$, SLO) and distance to the colony (m, DIST) as static variables and sea surface temperature ($^{\circ}\text{C}$, SST), SST gradient ($\%$, SSTgrad), SST anomaly ($^{\circ}\text{C}$, SSTan), Chlorophyll A concentration ($\text{mg}\cdot\text{m}^{-3}$, CHLA), current velocity ($\text{m}\cdot\text{s}^{-1}$, CURR), and wind velocity ($\text{m}\cdot\text{s}^{-1}$, WIN) as dynamic variables (Fig. A14). When necessary, values were log-transformed to reduce skewness, and all covariates were scaled to have mean 0 and standard deviation of 1 before entering the models.

Table 12

Environmental variables considered for the spatial models, their resolution, processing to final product and source were they were obtained

Variable	Acronym	Units	Original resolution	Final resolution	Processing	Source
Bathymetry	BAT	m	652, 748 (x, y) m	2.6, 2.9 (x, y) km		doi:10.7289/V5C8276M
Slope	SLO	rad	652, 748 (x, y) m	2.6, 2.9 (x, y) km	Calculated raster::terrain	with
Distance to shelf break (200m)	DISTshelf	m	652, 748 (x, y) m	2.6, 2.9 (x, y) km	Calculated raster::distance	with following Louzao et al 2009
Distance to closest coastline	DISTcoast	m	652, 748 (x, y) m	2.6, 2.9 (x, y) km	Calculated raster::distance	with
Distance to colony	DIST	m	652, 748 (x, y) m	2.6, 2.9 (x, y) km	Calculated raster::distance	with
Sea Surface temperature	SST	°C	1/100, 1/100 (x, y) °	2.6, 2.9 (x, y) km		https://bit.ly/2mGLJNC
SST gradient	SSTgrad	°C	1/100, 1/100 (x, y) °	2.6, 2.9 (x, y) km	Calculated from SST with raster::terrain and log transformed	
SST Anomaly	SSTan	°C	1/100, 1/100 (x, y) °	2.6, 2.9 (x, y) km		https://bit.ly/2kZsY7E
Chlorophyll-A concentration	CHLA	mg/m ³	4, 4 (x, y) km	2.6, 2.9 (x, y) km	Log transformed	https://bit.ly/2mBBWbM

Variable	Acronym	Units	Original resolution	Final resolution	Processing	Source
CHLA Gradient	CHLAgrad	rad	4, 4 (x, y) km	2.6, 2.9 (x, y) km	Calculated from ChlA concentration with raster::terrain	
Mixed Layer Thickness	MLT	m	1/12, 1/12 (x, y) °	2.6, 2.9 (x, y) km		https://bit.ly/2nb0CIu
Sea Surface Height	SSH	m	1/12, 1/12 (x, y) °	2.6, 2.9 (x, y) km		https://bit.ly/2nb0CIu
Wind velocity	WINDs	m/s	1/4, 1/4 (x, y) °	2.6, 2.9 (x, y) km	Calculated from the x and y components as $\sqrt{(x^2 + y^2)}$	https://bit.ly/2IH8Nvw
Wind direction	WINDd	°	1/4, 1/4 (x, y) °	2.6, 2.9 (x, y) km	Calculated with raster::atan2	https://bit.ly/2IH8Nvw
Current velocity	CURRs	m/s	1/12, 1/12 (x, y) °	2.6, 2.9 (x, y) km	Calculated from the x and y components as $\sqrt{(x^2 + y^2)}$ and log transformed	https://bit.ly/2nb0CIu
Current direction	CURRi	°	1/12, 1/12 (x, y) °	2.6, 2.9 (x, y) km	Calculated with raster::atan2	https://bit.ly/2nb0CIu
Ekman transport intensity	EKM	Kg·m ⁻¹ ·s ⁻¹	1, 1 (x, y) °	2.6, 2.9 (x, y) km	Calculated from the x and y components as $\sqrt{(x^2 + y^2)}$	https://bit.ly/2mGnMpE
Salinity	SAL	g·kg ⁻¹	1/12, 1/12 (x, y) °	2.6, 2.9 (x, y) km		https://bit.ly/2nb0CIu

4.3.3 Tracking data

We deployed GPS data-loggers (Perthold Engineering LCC; $42 \times 23 \times 16$ mm, 17 g) to adult breeding birds during the chick-rearing phase, attached to the mantle feathers with TESA © tape. Deployed loggers represented ca. 2.25% of the animals weight (below 4.0%; Passos et al. 2010). The devices were programmed to store a location every five min. Upon downloading, tracks were split in foraging trips, and points on land removed using custom-made R routines. We also filtered out all trips of a duration shorter than three hours, to remove from the final dataset short trips to rest in waters surrounding the colony. Since we had more than one trip per individual, we tested the effect of individual site fidelity in the spatial distribution of birds following the methods developed in chapter one and found a significant effect of the individual site fidelity on the foraging distribution of the three colonies. Thus, we randomly selected one trip per individual using a custom-made random selection function. From the filtered dataset we calculated trip metrics (maximum distance from the colony, duration, distance travelled) for each colony. Then, we classified at sea locations of birds in four behavioural modes (resting, relocating, intensive search and extensive search) using the R package **EMbC** (Garriga et al. 2016), and we performed the rest of the analysis only with foraging locations (i.e. intensive and extensive search). Even after selecting only foraging positions some correlation remained among consecutive locations. To address that issue we used the function **rthin()** from the package **spatstat** (Baddeley et al. 2016) to perform independent random thinning, producing a point pattern with the same density surface as the original one but only 10% of the original points. With the thinned dataset we constructed one model for each colony.

4.3.4 *Overlap analysis*

To assess the overlap between the foraging areas of the three colonies we used only the foraging positions of the filtered dataset (before thinning). First, for each colony, we calculated the average scale of the Area Restricted Search (ARS) behaviour with the functions developed in Lascelles et al. 2016. Second, we calculate kernel utilisation distributions (Worton 1989). We used the function `kernelUD()` from the package `adehabitathr`, with the average ARS scale as smoothing factor (Lascelles et al. 2016), and a value of 300 for the grid (for more information see the package vignettes). Third, we extracted the 50% (core area) and 95% (home range) contours and plotted them over the study area. Lastly, with the function `kerneloverlaphr()` we calculated pairwise overlaps between the core areas and home ranges of the three colonies, using the conditional estimate of the Bhattacharyya affinity (BA; Fieberg and Kochanny 2005). This measure goes from 0 (no overlap) to 0.5 or 0.95 for core areas and home ranges respectively (total overlap).

To assess whether the overlap among colonies was significantly different to what would be expected at random, we created a null distribution of BA by randomly reassigning a colony to the 69 trips of the dataset and calculating the overlap between core areas and home ranges of these randomised groups. We run this 10,000 times, and calculated a p-value as the proportion of randomly generated BA values that were smaller than the observed BA values.

4.3.4 Hierarchical Bayesian modelling

A spatial point pattern is a dataset containing locations of recorded events in space (i.e. animal presences). In this context, points themselves are not the subject of the analyses. Instead, we are interested in the mechanism, the ecological process that caused the points to be distributed in space in the way they are. This underlying mechanism is called “point process” and is the subject of point process modelling (Baddeley et al. 2016). Point processes are defined by their intensity function $\lambda(\mathbf{s})$, which describes how the points distribute within the study area. The Log-Gaussian Cox Processes (LGCP) are a particular type of point processes that can be modelled as a Gaussian random field. Formally, a random field is a family of random variables $Z(\mathbf{x})$ with values at every location \mathbf{x} of the study area (Illian et al. 2007). Practically, a Gaussian random field can be constructed as a linear combination of environmental covariates. In addition, a Gaussian random field may include a spatially structured field that can capture any structure in the original point pattern that is not explained by the other covariates.

4.3.4.1 Model definition

For each colony, the at-sea locations of birds were modelled as a spatial LGCP over an area A defined as the convex hull of the locations of the three colonies together (for comparability), with a buffer of 50 km around it to avoid any location falling too close to the boundary. In these models, the number of events (at-sea locations) in the region A is characterised by an intensity function $\lambda(\mathbf{s})$, with the function

$$\Lambda(A) = \int_A \lambda(\mathbf{s}) d\mathbf{s} \quad (1)$$

where $A(A)$ is the expected number of events in the area A and $\lambda(\mathbf{s}) \geq 0$ is the local intensity at location \mathbf{s} (i.e. $\lambda(\mathbf{s})$ represents the local intensity in a small –tending to 0– area around the location \mathbf{s}). The particular characteristic of LGCP that makes them appropriate for modelling events in relation to environmental variables is that their intensity can be modelled as a Gaussian model with a log-link, incorporating fixed (linear) effects, random (non-linear) effects and, crucially, a spatial random effect that can capture any structure in the data not explained by the covariates. Thus, the intensity is defined, for each location \mathbf{k} , as

$$\log \lambda(\mathbf{s}_k) = \beta_0 + \sum_{i=1}^I \beta_i x_i(\mathbf{s}_k) + f(u_1) + f(u_2) + \dots + w(\mathbf{s}_k) + \varepsilon(\mathbf{s}_k) \quad (2)$$

where β_0 corresponds to the intercept, x_i are the linear effects with their corresponding coefficients β_i , $f(u_i)$ are the non-linear effects, $w(\mathbf{s}_k)$ is the spatially structured random effect, a Gaussian Markov Random field (GMRF), and $\varepsilon(\mathbf{s}_k)$ is an unstructured random field or error term (akin to the nugget effect in a geostatistical model). Practically, due to computational limitations and identifiability problems between the structured and unstructured random effects (Sørbye et al. 2019) we could not include the unstructured random field in our model. The GMRF is defined as

$$w(\mathbf{s}) \sim N(\mathbf{0}, \Sigma) \quad (3)$$

where Σ is the covariance matrix with a Matérn structure defined as

$$\Sigma = \text{Cov}[w(\mathbf{s}_i), w(\mathbf{s}_j)] \quad (4)$$

4.3.4.2 Mesh generation

To efficiently represent the GMRF using the SPDE approach we constructed a mesh following the finite element method described in Simpson et al. (2011). The mesh was constructed covering all the area A and extended outwards in a 200 km buffer to avoid edge effects in our model. The minimum edge length of the triangles was 20 km in the inner boundary and 50 km in the outer boundary, and the cut-off (minimum length of segments forming the border) was set at 10 km. To avoid predicting intensity over land we artificially assigned value of 0 to all mesh nodes that were not on water. Further details and R code for mesh construction can be found in the Annex 3.

4.3.4.3 Covariates and model selection

All non-correlated covariates were included in the model as both fixed and random, in order to separate linear and non-linear effects for each, easing interpretation (Lombardo et al. 2019). The covariates retained with both fixed and random components are represented in the results as the sum of the fixed component coefficient and the effect of the random component: where the median and 95% CI fall above 0 the covariate has an overall positive effect on point intensity and, conversely, where median and 95% CI fall under 0 the covariate has an overall negative effect in point intensity. Due to the random component, this effect might be stronger (farther from 0) or weaker (closer to 0) at certain covariate values.

The fixed effect coefficients were assigned the default prior on the precision (0 for the intercept and 0.01 for the covariate coefficients), and the random effects were modelled as random walks of second order (rw2)

grouped into 20 classes, and scaled so the variance equals 1 for all of them, for easy interpretation (Sørbye and Rue 2014). This type of models has one hyperparameter, the precision, which controls the size of the effect, and how much the variable is allowed to deviate from the mean (i.e. how wiggly the effect is). INLA allows us to specify a prior on this hyperparameter.

The spatial effect was modelled as an SPDE model with two hyperparameters to which we set priors, the range (ρ) and the standard deviation (σ). The range represents the distance at which two locations are no longer correlated (i.e. the typical distance between peaks and troughs), which controls the smoothness of the spatial field; and the standard deviation represents how much the spatial field is allowed to vary locally (i.e. the magnitude of the peaks and troughs).

From the full model with all covariates as fixed and random effects and the spatially structured random field, we sequentially removed the random effects of covariates that appeared to have a linear effect on the intensity, and the fixed effects of covariates when the 95% credible interval (CI) for the coefficient overlapped zero, until we arrived at the most parsimonious model.

4.3.4.4 Prior values specification

We set priors for the following hyperparameters of the model: the precision of each of the random effects (rw2 models), and the range and standard deviation of the spatial field (SPDE model). To facilitate their specification we used Penalised Complexity (PC) priors, a newly developed framework that allows easily interpretable and controllable priors (Simpson et al. 2014). PC priors are weakly informative (allowing

the posterior of each hyperparameter to be mainly controlled by the data) and penalise model complexity by “pulling” the model towards its simplest realisation (the “base” model): for the rw2 models, the base model has infinite precision, which would produce a model with no random effect, essentially a fixed effect (Simpson et al. 2014). For the SPDE models the base model has infinite range and zero standard deviation (i.e. a completely flat spatial field, absence of spatial structure). To set the priors, we inform the model of “how far it is allowed to deviate” from those base models using the following specifications:

- For the rw2 precision the prior is set on the standard deviation (σ), which is inverse to the square root of the precision (τ) so: $\sigma = \frac{1}{\sqrt{\tau}}$. We provide the upper tail quantile σ_0 and the probability P_σ so that:

$$\text{Prob}(\sigma > \sigma_0) = P_\sigma, \quad \sigma_0 > 0, \quad 0 < P_\sigma < 1 \quad (5)$$

which translates to: “the probability that the σ of the model is larger than σ_0 is P_σ ”. For example, if $\sigma_0 = 3$ and $P_\sigma = 0.01$, the probability of σ of being larger than 3 is 1%. In this way, we are limiting the rw2 model’s σ to stay between 0 and 3 with a 99% probability. Since the precision (τ) is inversely related to σ , we are basically limiting the precision to values between infinite (the base model) and $\frac{1}{\sigma^2}$

- The prior on the spatial field range (ρ) is set providing the lower tail quantile ρ_0 and the probability P_ρ so that

$$\text{Prob}(\rho > \rho_0) = P_\rho, \quad \rho_0 > 0, \quad 0 < P_\rho < 1 \quad (6)$$

or “the probability that the true range is smaller than ρ_0 is P_ρ ”. For example, if $\rho_0 = 500$ and $P_\rho = 0.01$, the probability of the true range of the spatial field being smaller than 500 km is 1%. In this way we are limiting the range to values between infinite (the base model) and 500, i.e. we are saying that the smaller that the range could possibly be is 500 with a probability of 99%.

- The prior on the spatial field standard deviation is set providing the upper tail quantile σ_0 and the probability P_σ so that

$$\text{Prob}(\sigma > \sigma_0) = P_\sigma, \quad \sigma_0 > 0, \quad 0 < P_\sigma < 1 \quad (7)$$

or “the probability that the standard deviation is larger than σ_0 is P_σ ”. For example, if $\sigma_0 = 0.5$ and $P_\sigma = 0.01$, the probability of the standard deviation of the spatial field being larger than 0.5 is 1%. In this way, we are effectively limiting the standard deviation to values between 0 (the base model) and 0.5 with a 99% probability.

Priors have to be carefully specified, but there is no absolute rule for it. For that reason, a large part of the modelling process with this type of models is deciding the priors to use. Below we outline the decision process for each of the priors involved in our model:

- The rw2 precision prior was originally set to the value of the standard deviation of the covariate being modelled. In all cases this proved to be too restrictive, giving a random effect that was completely flat, so we repeated the model multiplying that value

by a sequence of increasing values until we arrived at a satisfactory level of non-linearity.

- The SPDE range prior was chosen according to the following process: we began with a value of 20 km (as small as the length of the mesh triangles side) to indicate that the range could be between the minimum that the model is able to detect and infinite, allowing the posterior range to be heavily affected by the data. We ran the model with this prior once, without any covariate but the spatial effect, and used the median range of the posterior of that model as a prior.
- The SPDE standard deviation prior was initially set as 0.2. However, since the priors for the two hyperparameters of the spatial effect have a great influence in the posterior of the model, we performed a sensitivity test: once we had selected a final model, we repeated that model with a series of values for σ and range, and compared results to see the differences in the spatial effect and the predicted intensity, and from that, we selected the most convincing model and used those parameters to run the definitive model (Annex 4). The parameters chosen for the definitive models are specified in Table 13.

Table 13

Summary of components of all models fitted to GPS location data for each colony.

		Included	Priors of the spatial effect hyperparameters	Non-linear covariates	Linear Covariates
Montaña Clara	Model 0	Yes	$\rho_0 = 80 \text{ km}; \sigma_0 = 0.2$	No	No
	Model 1	No		No	CHLA, SLO, SST, DIST
	Model 2	No		CHLA, SST, DIST	CHLA, SLO, SST, DIST
	Model 3	Yes	$\rho_0 = 80 \text{ km}; \sigma_0 = 0.2$	CHLA, SST, DIST	CHLA, SLO, SST, DIST
Timanfaya	Model 0	Yes	$\rho_0 = 90 \text{ km}; \sigma_0 = 0.3$	No	No
	Model 1	No		No	CHLA, SSTgrad, SLO, DIST, SST
	Model 2	No		CHLA, SSTgrad, DIST	CHLA, SSTgrad, SLO, DIST, SST
	Model 3	Yes	$\rho_0 = 90 \text{ km}; \sigma_0 = 0.3$	CHLA, SSTgrad, DIST	CHLA, SSTgrad, SLO, DIST, SST
Veneguera	Model 0	Yes	$\rho_0 = 100 \text{ km}; \sigma_0 = 0.2$	No	No
	Model 1	No		No	CHLA, SLO, SST, DIST
	Model 2	No		CHLA, SST, DIST	CHLA, SLO, SST, DIST
	Model 3	Yes	$\rho_0 = 100 \text{ km}; \sigma_0 = 0.2$	CHLA, SST, DIST	CHLA, SLO, SST, DIST

4.3.4.5 Prediction

To obtain the intensity predicted by the model, we used the `inla.posterior.sample()` function to draw 1,000 posterior samples from linear predictor of the model. These posterior samples contain all necessary parameters to reconstruct a modelled prediction of the point intensity as well as the covariate fixed and random effects and the spatial field effect. Code and more information on the process can be found in the Annex 5.

4.3.4.6 Characterization of spatial random effect

The spatial field captures the spatial autocorrelation structure in the point intensity that has not been captured by the covariates. From the posterior samples obtained above we can obtain a median value of the effect of the spatial field at each mesh node and map it on our study area using the functions provided in the **R-INLA** package (Annex 5). We plotted the spatial field in the scale of the predictor (log scale) to obtain a surface centred around 0, with peaks of positive values representing areas where the observed point intensity was higher than our covariates were able to capture (hotspots) and negative troughs representing areas where the observed point intensity was lower than our covariates were able to capture (coldspots). The median of the spatial field posterior range represents the typical distance between these peaks and troughs (the smoothness of the surface).

4.3.4.7 Performance of models

Since common model validation and evaluation processes (i.e. AIC, cross validation) are not appropriate for LGCP models (Baddeley et al. 2005) we devised procedures to evaluate the performance of the models. First

of all, to assess the benefits of increasing the complexity of the linear predictor, we ran three models for each colony: the simplest model only with fixed effects (Model 1), a more complex model with fixed and random effects, and the most complex model with a spatial field, fixed and random effects in the linear predictor (Model 3). Second, to ensure the spatial field is not capturing all the structure of the data, confounding the effect of the environmental covariates, we ran a null model (Model 0) only with an intercept and a spatial field (i.e. without covariates), to compare the posterior spatial field of this null model with that of Model 3. Third, we calculated the L-function of our observed point pattern. The L function is a second order characteristic of point patterns that provides information on the aggregation of points by using a measure related to the number of points inside a radius r from the focal point (Illian et al. 2007). We simulated 1,000 point patterns from our fitted model using the function `sample.lgcp()` from the package `inlabru`, and generated an envelope (95% interval) with the function `envelope()` from the package `spatstat`. We then plotted the L-function of our observed point pattern (L_{obs}) and the envelope of our 1,000 simulated patterns. If L_{obs} is inside the envelope the model from which the envelope was generated has captured the structure of the observed pattern. If L_{obs} is above the envelope the observed point pattern is more clustered than the model can explain, and if L_{obs} is below the envelope the observed pattern is more dispersed. Lastly, from the 1,000 simulated point patterns we extracted the median and 95% CI of the total number of points simulated for each and compared that to the number of points of the observed point pattern, to see how well our model can predict the total abundance of our pattern.

4.4 Results

4.4.1 Overlap analysis

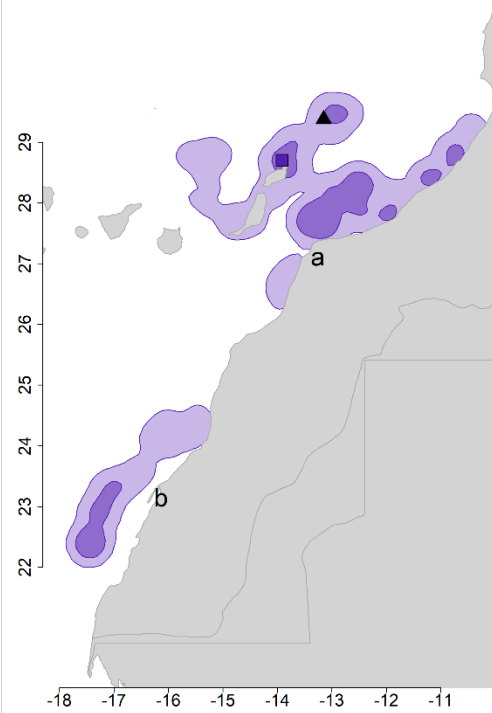
From the original dataset (Table 11), after the process of selecting one track per individual, the behavioural classification and the random thinning, we obtained point patterns of 153 points for Montaña Clara, 393 points for Timanfaya and 281 points for Veneguera. For the three colonies, there was a core foraging area in the waters surrounding the colony (from now on adjacent foraging areas), and several other areas not adjacent to the colony, and most of them over the African continental shelf (from now on distant foraging areas). The largest core area and home range were those of the birds breeding in Montaña Clara, and the smallest those of Veneguera (Fig. 14; Table 14).

Table 14

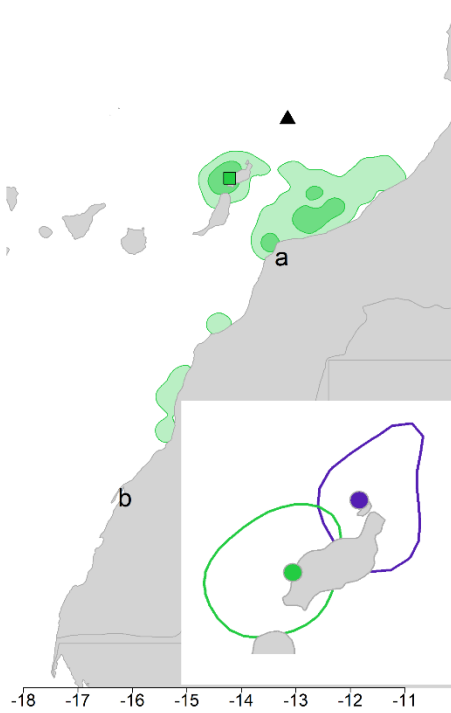
Measure of overlap (BA) between core areas (lower diagonal) and home ranges (upper diagonal) calculated from the foraging positions of birds breeding in the three sampled colonies. In parenthesis, we give the p-value obtained from the number of null BA values that were lower than the observed BA value for each pairing. Also shown, in italics, the size of the core areas (bottom row) and home ranges (left column) for each colony.

	Montaña Clara	Timanfaya	Veneguera	<i>Home range size (km²)</i>
Montaña Clara	–	0.64 (0.01)	0.16 (0.00)	<i>54019</i>
Timanfaya	0.20 (0.02)	–	0.83 (0.10)	<i>31369</i>
Veneguera	0.00 (0.00)	0.13 (0.00)	–	<i>42643</i>
<i>Core area size (km²)</i>	<i>8127</i>	<i>5230</i>	<i>6131</i>	–

MONTAÑA CLARA



TIMANFAYA



VENEGUERA

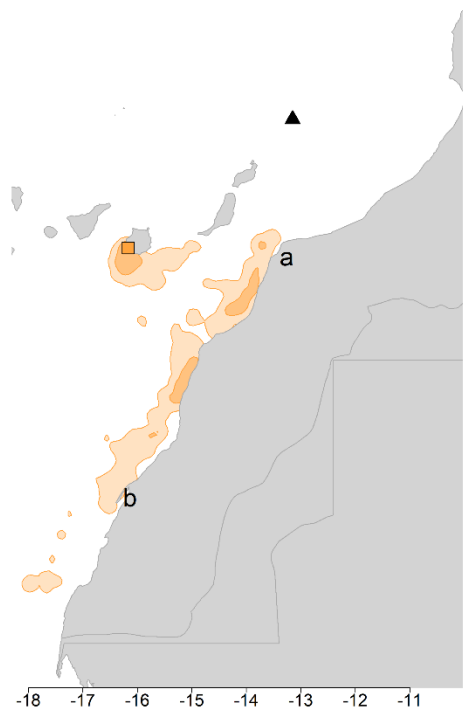


Figure 14 (previous page)

Core areas (darker colours) and home ranges (lighter colours) from the foraging positions of birds breeding in Montaña Clara (purple), Timanfaya (green) and Veneguera (tan). Darker circles mark the location of the corresponding colony. The black triangle marks the location of the Concepcion Seamount. Also shown are the locations of Cape Juby (a) and the Dakhla Peninsula (b). The inset in the middle plot shows a zoom of the overlap between the adjacent core areas of the Montaña Clara and Timanfaya colonies

Overall, there was less overlap between core areas than between home ranges: the maximum core areas overlap was between Timanfaya and Montaña Clara, and the maximum home ranges overlap between Timanfaya and Veneguera (Table 14). All overlaps were significantly smaller than expected at random (at an α level of 0.05) except for the overlap of the home ranges of Veneguera and Timanfaya (Fig. 2; Table 4).

4.4.2 Hierarchical Bayesian modelling

The final models (Model 3) of the three colonies included random and fixed effects for the covariates DIST and CHLA. Random effect for SSTgrad was included for the Timanfaya and Veneguera models. SLO was included with a random effect for the Montaña Clara and Timanfaya models, and with a fixed effect for the Veneguera model. SST was included as random effect in the Montaña Clara, and as fixed in the Timanfaya and Veneguera models (Table 4).

4.4.3 Model evaluation

Model fit improved with increasing complexity for the three colonies. The predicted intensities reflected the observed point pattern better when random effects for the covariates and a spatial field had been

included as linear predictors (Fig. A15). The L-functions and their envelopes also showed a better ability of the models with a spatial field to capture clustering of points than the other two models. However, there was clustering not captured at scales <10 km even in the spatial model, as shown by the observed pattern being represented above the envelope (Fig. A16). Contrary to that, the predicted total abundances were less accurate in the more complex models for the three species (Table A5). This seems a consequence of our models under-estimating the clustering of the point pattern, which caused in turn the under-estimation of the total abundance of points.

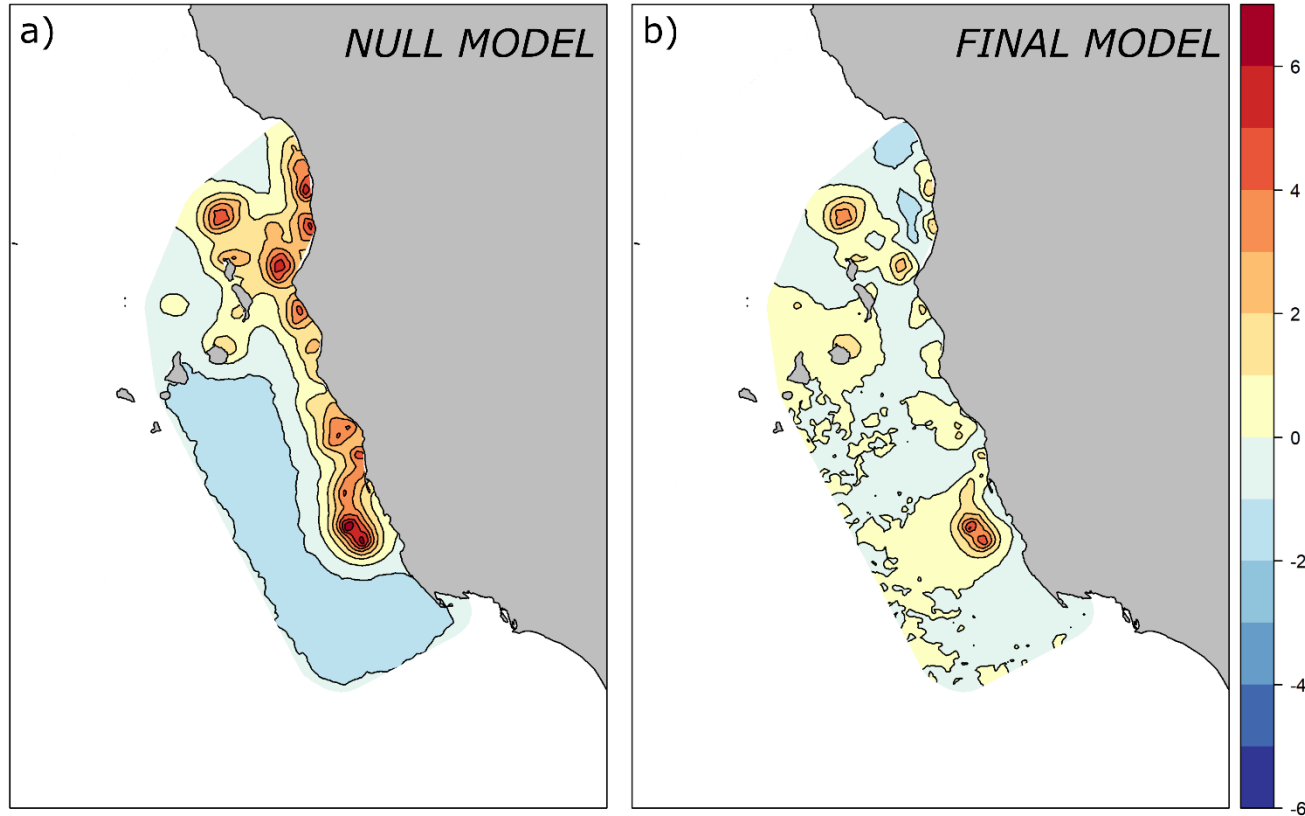
4.4.4 Environmental covariates

For the three colonies, comparison of final and null models showed differences in relative intensity of the spatial fields, revealing the relevant contribution of the covariates to capturing part of the spatial structure of the point patterns (Fig. 15). For all three colonies, CHLA and SLO had an overall positive effect on point intensity, and DIST and SST had a negative effect. The effect of SSTgrad was positive in the Timanfaya model, but negative in the Veneguera model (Fig. 16).

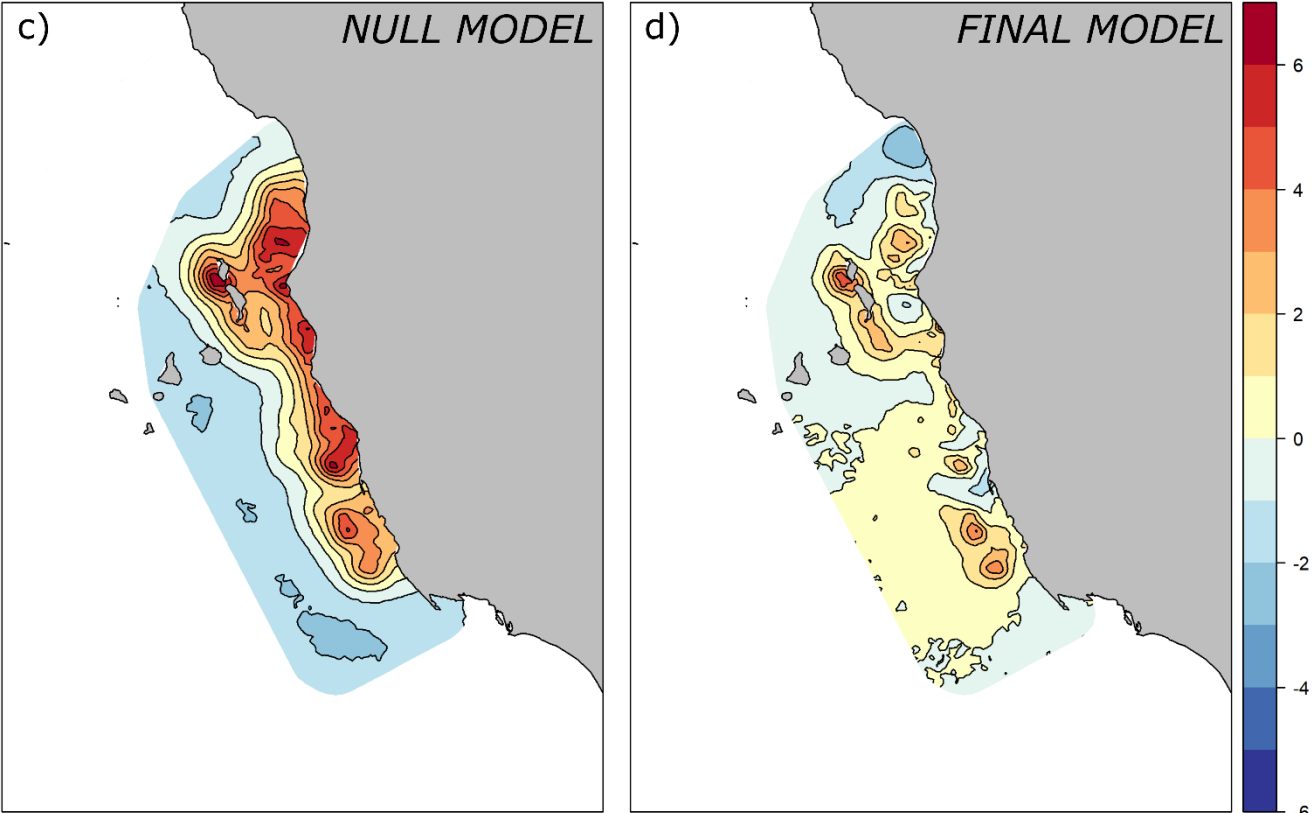
Figure 15 (next pages)

Spatial field of the null model and the final model for the GPS locations of the birds breeding in Montaña Clara (a, b), Timanfaya (c, d) and Veneguera (e, f). The colour scale represents the intensity of the effect in the scale of the linear predictor (i. e. in the logarithmic scale). Positive values show areas where observed intensity is higher than predicted by the covariates, and negative values show areas where observed intensity is lower than predicted by the covariates

MONTAÑA CLARA



TIMANFAYA



VENEGUERA

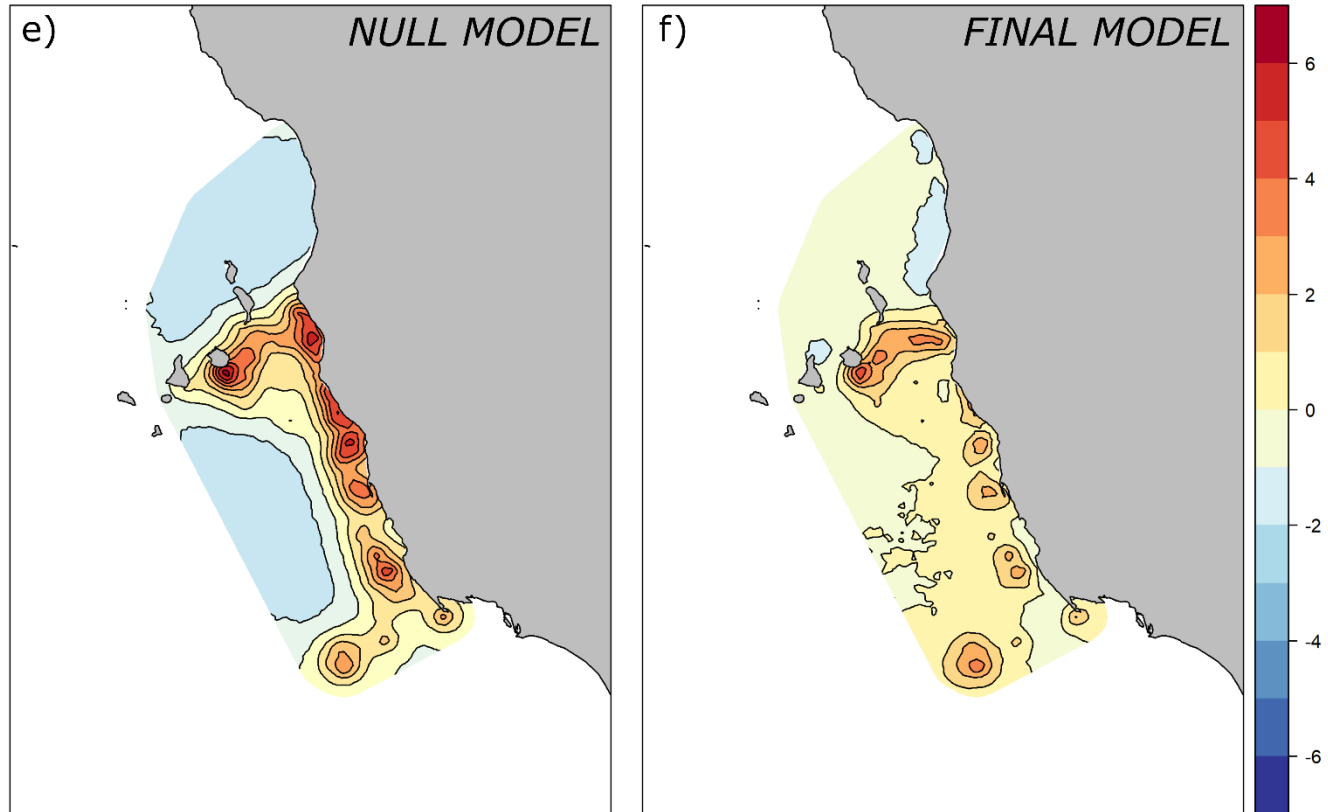
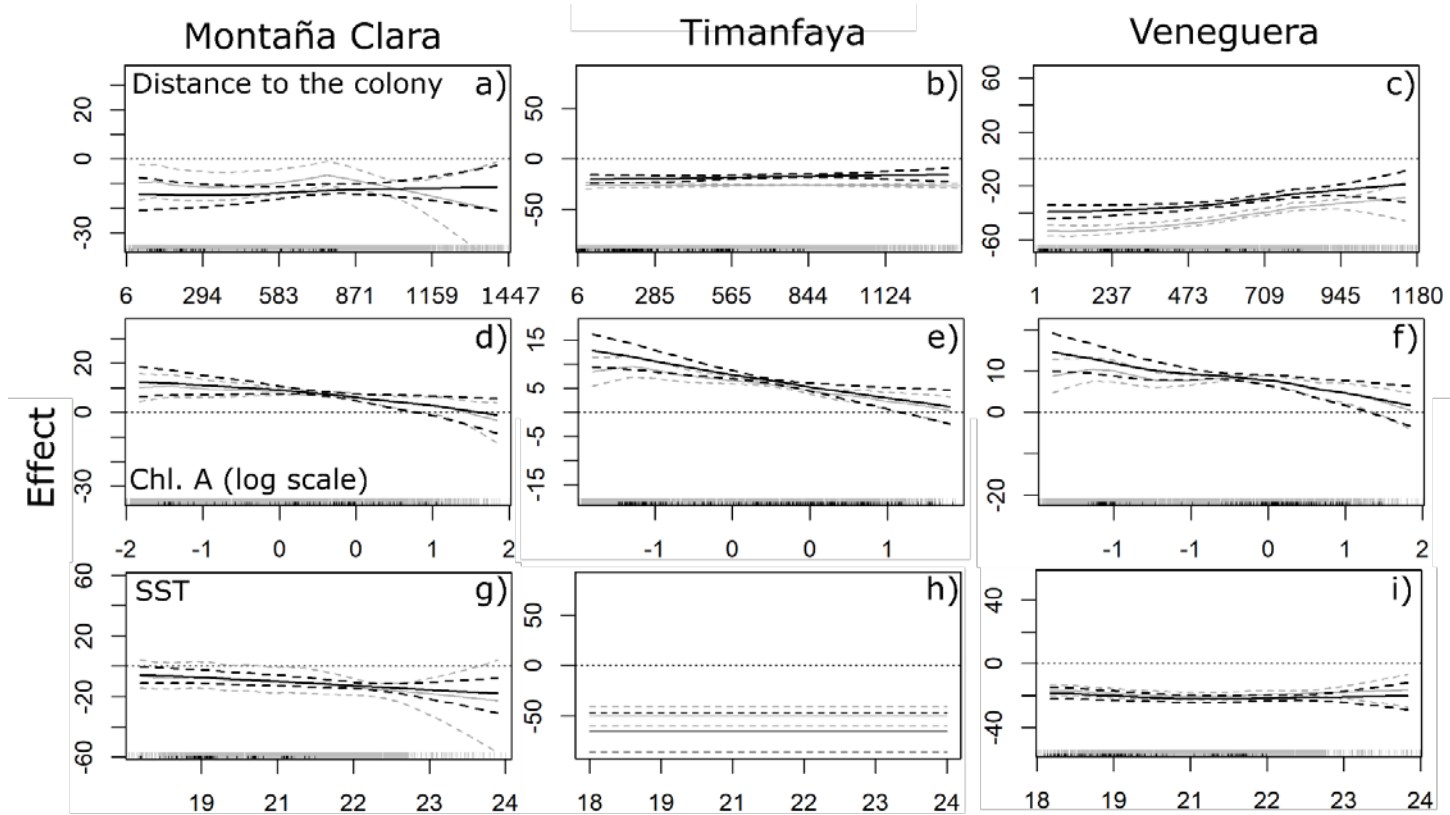


Figure 16



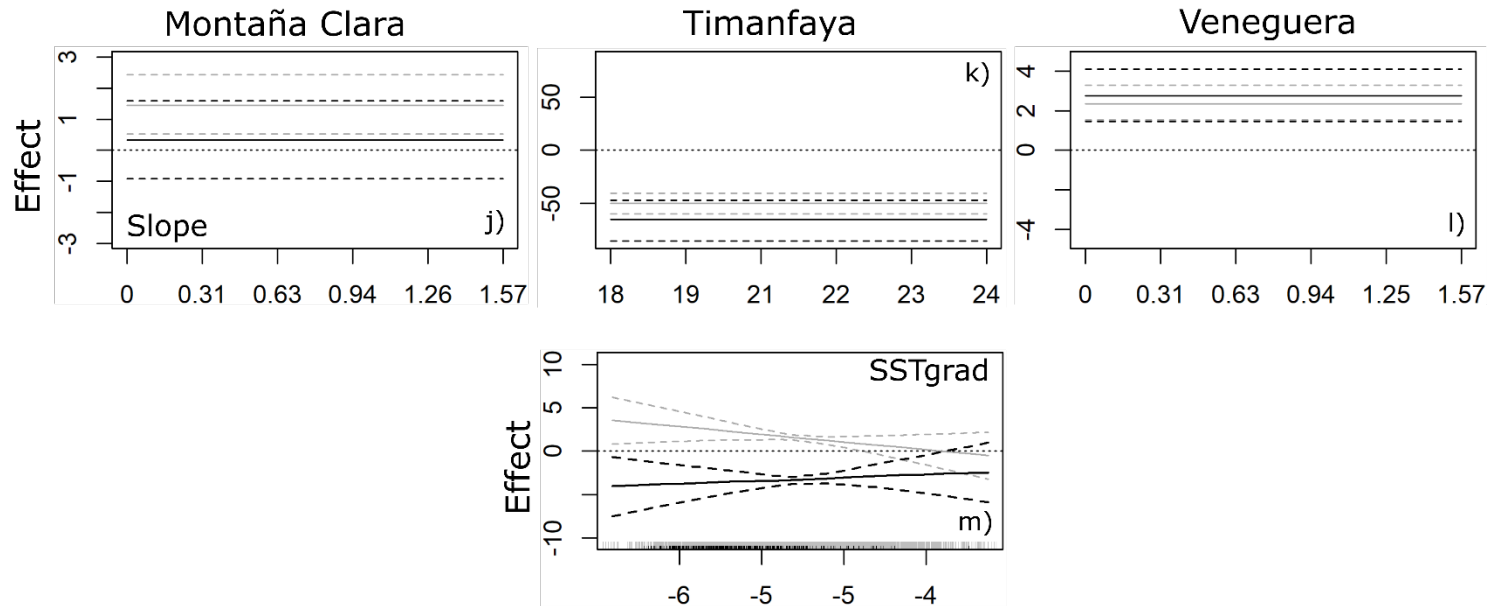


Figure 16

size of the non-linear effects of the covariates for the spatial model (black lines) and non-spatial model (gray lines) for the colony of Montaña Clara (left), Timanfaya (centre) and Veneguera (right). The solid line represents the median, and the dashed lines the 95% CI. The gray rug represents the values for the covariate at which the effect is evaluated (integration points) and the black rug represents the values for the covariate at which there are GPS locations. The slope was included as linear variable for the three models, and the SST for the Timanfaya models, but is represented here as well as a constant for clarity.

4.4.5 *Spatial effects*

Median posterior ranges of the spatial effect were between 103.6 and 115.8 km, and posterior σ between 1.5 and 1.7 (Table A5). Plots of the spatial effect for the three colonies showed some spatial structure that had not been captured by the environmental covariates (Fig. 15 b, d & f).

4.5 Discussion

Our kernel overlap analysis provides clear evidence of segregation among the foraging distributions of Cory's shearwaters breeding in three nearby colonies, both in adjacent and distant areas, and our spatial models prove the existence of small-scale clustering positions and different spatial structure in the foraging distributions of the three colonies. In combination, these results are strongly suggestive of the presence of both local enhancement and information transfer at the colony in the colonies studied.

Since the colonies of Montaña Clara and Timanfaya are less than 50 km apart, we expected their adjacent foraging areas to overlap substantially. However, this overlap was avoided by birds from each colony using mostly adjacent areas in opposite directions. That is, birds from Montaña Clara used waters to the northeast of the colony and birds from Timanfaya waters to the southwest. Although it has been suggested that this type of very short scale segregation can be caused by the orientation of colonies and the direction birds take when flying out of the breeding sites (Ceia et al. 2014), the Montaña Clara colony is oriented to the southeast, discarding this possibility. It is possible, however, that birds fly northeast after leaving the colony to target the

productive waters surrounding the Concepcion Seamount (100 km Northeast of Montaña Clara), which is not used by birds of any of the other two colonies sampled in this study, but is known to be used by birds of the nearby colonies of Alegranza and Salvage islands (Ramos et al. 2013). The observed segregation between the adjacent foraging areas of the two colonies, thus, could certainly be explained within the framework of Wakefield's DDH model: birds avoided intra-colony competition by expanding outwards, and when the foraging areas overlapped birds avoided inter-colony competition by searching waters with less conspecifics present, causing inter-colony segregation. This segregation could be perpetuated and reinforced by inexperienced birds following experienced ones to the colony-specific foraging areas. The absence of overlap between the adjacent foraging areas of Veneguera and the colonies of Montaña Clara and Timanfaya is not surprising, since the distance between Veneguera and the other two colonies (~ 200 km) is similar to the distance from Veneguera to the continental shelf.

In the spatial models, distance to the colony had a clear, negative effect on point intensity for every colony. Since flight is an energetically expensive activity, breeding birds minimise their energy expenditure (e.g. foraging distance) while maximising food intake (MacArthur and Pianka 1966), which should result in a decrease in the number of foraging birds with distance to the colony. However, for the three colonies, and particularly for Veneguera, the non-linear component of the covariate allowed us to detect that this negative effect was weaker at larger distances, reflecting a bird abundance peak far from the colony that other covariates could not explain. In addition, the spatial fields for the three models showed one common hotspot for the three colonies, between

the Dakhla Peninsula and Cap Blanc (Fig. 15 b, d & f). The existence of common foraging ground suggests that a geographical or environmental driver not included in our modelling is affecting birds of the three colonies equally, and it promotes the aggregation of foraging birds in that area. In this regard, although bathymetry was not included in our models because it was highly correlated with other covariates, the shelf off the coast in that area is the widest in the Canary Current, so it is conceivable that birds from the three colonies concentrated in such highly productive waters (Longhurst et al. 1995).

At large scales, the distant core foraging areas for the three species were distributed over the cold, productive waters of continental shelf and the shelf break, between the 22 and 29° N of latitude, well within the area of influence of the productive Canary Current upwelling (Longhurst 1998, Davenport et al. 2002). In our spatial models this was illustrated by a positive relationships of point intensity to CHLA and negative to SST for all three models, as found in previous studies (Navarro and González-Solís 2009, Ramos et al. 2013). However, at medium scales birds from Montaña Clara and Timanfaya used mostly waters to the North of Cape Juby, while birds from Veneguera used mostly waters to the South. Although it could seem that the distant core areas of birds from Montaña Clara and Timanfaya occupied the same general area, the overlap between them was significantly smaller than expected at random. Environmentally, this segregation was reflected by the different preferences for the meso-scale features included in the models (SSTgrad and slope; Acha et al. 2015). Thus, our models evidence the utility of environmental variables to explain large scale foraging distributions of seabirds, like their use of the highly productive waters of the Canary

Current, although they also underline the need for caution when modelling and predicting foraging distributions of these animals at a medium scale (e.g., shelf-break and thermal fronts), as the use of averaged values of highly dynamic fronts might be inappropriate at these smaller scales (Acha et al. 2015).

Procellariiforms, such as Cory's shearwaters are highly mobile, able to forage at >1,000 km (Oppel et al. 2018) At these geographical scales, trying to study their distributions in accordance to the DDH model is not possible since the assumptions the model make are not met: at such scales, prey at sea is not uniformly distributed, but aggregated in highly productive areas promoted by certain geographic and environmental features (Carroll et al. 2017), which violates assumption 1. This makes the relationship of distance to number of foraging birds not linear, also violating assumption 2, as it can be seen from our models. Lastly, our sampled colonies were far from being the only relevant populations in the area, violating assumption 3 from the DDH model: the largest Cory's shearwater colony in the Canary Islands (Alegranza Islet) lays barely 10 km to the north of Montaña Clara, and the nearby population of Salvages Islands holds the largest population of Cory's shearwaters in the world. Birds from all these colonies, together with individuals from many other species, are known to also use the Canary Current for foraging (Ramos et al. 2013, Grémillet et al. 2015, Grecian et al. 2016). In these distant foraging areas, the effects of inter-colony competition are expected to be buffered by the increasing foraging costs and the elevated productivity of the area, which can explain the ability of this area to sustain a large number of foraging individuals (Grecian et al. 2016, Bolton et al. 2018). Such an elevated number of foraging animals

present in the limited area of the Canary Current upwelling could influence their foraging distributions through local enhancement. This behaviour has been identified in many colonial seabird species including shearwaters (Buckley 1997), therefore it is conceivable that local enhancement was responsible for some of the small-scale clustering that the L-functions and their envelopes evidenced in our data, since it is precisely at those short scales that the effects of local enhancement would be evident. However, foraging groups formed through local enhancement develop quickly and can disappear just as quickly if the number of predators is high enough to deplete the prey patch in a short time, or if the prey display escape behaviour (Lewis et al 2001). Thus, the development of spatiotemporal models able to model distributions at a smaller spatial scale could provide further evidences of this type of aggregations.

Another factor that could be affecting at sea distribution of foraging birds is public information obtained at the colony. The instances of information transfer identified so far in seabird species rely on birds obtaining cues from returning individuals, following departing individuals, or through “compass rafts” that form near the colony (Weimerskirch et al. 2010, Wakefield et al. 2013, Jones et al. 2018). It is complicated to prove how information gathered in this way could inform the bird’s decision on where to forage in the distant Canary Current, but a combination of tracking environmental cues, coupled to the different position of each colony with respect to the continental shelf, and the avoidance of inter-colony competition, could define different optimal foraging areas which maximise foraging efficiency for each colony, and these different areas could be reinforced and perpetuated by

unexperienced birds from each colony following experienced ones to the foraging grounds. Whatever the mechanisms, the spatial fields of the models for the three colonies showed hotspots of unexplained point intensity (i.e. abundance of birds) in different regions of the study area for each of the sampled colonies (Fig. 3 d-f), difficult to explain unless there is a transfer of public information.

4.5.1 Conclusions

Our results prove the applicability of continuously indexed GMRF models to the analysis of animal distribution from tracking data. The ability to include covariates as fixed and random effects, and particularly the inclusion of a spatial field, provided very significant improvements in model fit and predictions. However, there was still some small-scale aggregation (at scales <10 km particularly for the Montaña Clara and Timanfaya models) that our models did not have enough resolution to explain. Future developments within the INLA framework can help dealing with this issue by allowing for a second spatial field with a smaller mesh and range to be added to the linear predictor, or by developing a spatial field model that is able to account for both local and large scale variability with a SPDE model (as has already been developed for the gridded approach; Sørbye et al. 2019). Despite these limitations, we have demonstrated the significant spatial segregation of at-sea foraging distributions of birds breeding in three neighbouring colonies, both in the waters surrounding the colony and in distant foraging grounds, have confirmed the environmental preference of foraging shearwaters for the cool productive waters of the canary current, and have suggested a mechanism through which the transfer of

information, both at sea and in the colony, can shape the distributions of foraging long-ranging seabirds.

4.6 References

- Aarts, G., J. Fieberg, and J. Matthiopoulos. 2012. Comparative interpretation of count, presence-absence and point methods for species distribution models. *Methods in Ecology and Evolution* 3:177–187.
- Acha, E. M., A. R. Piola, O. Iribarne, and H. Mianzan. 2015. *Ecological Processes at Marine Fronts: Oases in the Ocean*. Springer.
- Bachl, F. E., F. K. Lindgren, D. L. Borchers, and J. B. Illian. 2019. inlabru: an R package for Bayesian spatial modelling from ecological survey data. *Methods in Ecology and Evolution* 10:760–766.
- Baddeley, A., E. Rubak, and R. Turner. 2016. *Spatial Point Pattern: Methodology and Applications with R*. CRC Press, Boca Raton.
- Baddeley, A., R. Turner, J. Møller, and M. Hazelton. 2005. Residual analysis for spatial point processes (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67:617–666.
- Bakka, H., H. Rue, G. A. Fuglstad, A. Riebler, D. Bolin, J. B. Illian, E. Krainski, D. P. Simpson, and F. K. Lindgren. 2018. Spatial modeling with R-INLA: A review. *Wiley Interdisciplinary Reviews: Computational Statistics*:1–24.
- Benhamou, S. 2014. Of scales and stationarity in animal movements. *Ecology Letters* 17:261–272.
- Bolton, M., G. Conolly, M. Carroll, E. D. Wakefield, and R. Caldow. 2018. A review of the occurrence of inter-colony segregation of seabird foraging areas and the implications for marine environmental impact assessment. *Ibis*.
- Boyd, C., D. Grünbaum, G. L. Hunt, A. D. S. E. Punt, H. Weimerskirch, and S. Bertrand. 2016. Effectiveness of social information used by seabirds searching for unpredictable and ephemeral prey. *Behavioral Ecology* 27:1223–1234.
- Buckley, N. J. 1997. *Spatial-Concentration Effects and the Importance of Local*

- Enhancement in the Evolution of Colonial Breeding in Seabirds. *The American Naturalist* 149:1091–1112.
- Cairns, D. K. 1989. The regulation of seabird colony size: a hinterland model. *American Naturalist* 134:141–146.
- Carroll, G., M. Cox, R. Harcourt, B. J. Pitcher, D. Slip, and I. Jonsen. 2017. Hierarchical influences of prey distribution on patterns of prey capture by a marine predator. *Functional Ecology* 31:1750–1760.
- Ceia, F. R., V. H. Paiva, R. S. Ceia, S. Hervías, S. Garthe, J. C. Marques, and J. A. Ramos. 2014. Spatial foraging segregation by close neighbours in a wide-ranging seabird. *Oecologia* 177:431–440.
- Coulson, J. C. 2002. Colonial breeding in seabirds. Pages 87–113 *in* E. A. Schreiber and J. Burger, editors. *Biology of Marine Birds*. CRC Press.
- Davenport, R., S. Neuer, P. Helmke, J. Perez-Marrero, and O. Llinás. 2002. Primary productivity in the northern Canary Islands region as inferred from SeaWiFS imagery. *Deep-Sea Research Part II: Topical Studies in Oceanography* 49:3481–3496.
- Dell’Ariccia, G., A. Celerier, M. Gabirot, P. Palmas, B. Massa, and F. Bonadonna. 2014. Olfactory foraging in temperate waters: Sensitivity to dimethylsulfide by shearwaters in the Atlantic Ocean and Mediterranean Sea. *The Journal of experimental biology* 217:1701–1709.
- van Eeden, R., T. Reid, P. Ryan, and L. Pichegru. 2016. Fine-scale foraging cues for African penguins in a highly variable marine environment. *Marine Ecology Progress Series* 543:257–271.
- Fieberg, J., and C. O. Kochanny. 2005. Quantifying home-range overlap: the importance of the utilization distribution. *Journal of Wildlife Management* 69:1346.
- Fritz, H., S. Said, and H. Weimerskirch. 2003. Scale-dependent hierarchical adjustments of movement patterns in a long-range foraging seabird. *Proceedings of the Royal Society B* 270:1143–1148.
- Garriga, J., J. R. Palmer, A. Oltra, and F. Bartumeus. 2016. Expectation-Maximization Binary Clustering for Behavioural Annotation. *Plos One* 11:e0151984.
- Genovart, M., J. Bécares, J. Igual, A. Martínez-Abraín, R. Escandell, A.

- Sánchez, B. Rodríguez, J. M. Arcos, and D. Oro. 2018. Differential adult survival at close seabird colonies: The importance of spatial foraging segregation and bycatch risk during the breeding season. *Global Change Biology* 24:1279–1290.
- Genovart, M., D. Doak, J. M. Igual, S. Sponza, J. Kralj, and D. Oro. 2017. Varying demographic impacts of different fisheries on three Mediterranean seabird species. *Global Change Biology* 23:3012–3029.
- Granadeiro, J. P., M. Nunes, M. P. Silva, and R. W. Furness. 1998. Flexible foraging strategy of Cory's shearwater, *Calonectris diomedea*, during the chick-rearing period. *Animal Behaviour* 56:1169–1176.
- Grecian, W. J., M. J. Witt, M. J. Attrill, S. Bearhop, P. H. Becker, C. Egevang, R. W. Furness, B. J. Godley, J. González-Solís, D. Grémillet, M. Kopp, A. Lescroël, J. Matthiopoulos, S. C. Patrick, H.-U. Peter, R. A. Phillips, I. J. Stenhouse, and S. C. Votier. 2016. Seabird diversity hotspot linked to ocean productivity in the Canary Current Large Marine Ecosystem. *Biology Letters* 12.
- Grémillet, D., C. Péron, P. Provost, and A. Lescroël. 2015. Adult and juvenile European seabirds at risk from marine plundering off West Africa. *Biological Conservation* 182:143–147.
- Illian, J. B., and D. F. R. P. Burslem. 2017. Improving the usability of spatial point process methodology: an interdisciplinary dialogue between statistics and ecology. *AStA Advances in Statistical Analysis* 101:1–26.
- Illian, J. B., A. Penttinen, H. Stoyan, and D. Stoyan. 2007. *Statistical Analysis and Modelling of Spatial Point Patterns*. John Wiley & Sons, Ltd, Chichester, UK.
- Jones, T. B., S. C. Patrick, J. P. Y. Arnould, M. A. Rodríguez-Malagón, M. R. Wells, and J. A. Green. 2018. Evidence of sociality in the timing and location of foraging in a colonial seabird. *Biology Letters* 14:20180214.
- Jovani, R., B. G. Lascelles, L. Z. Garamszegi, R. Mavor, C. B. Thaxter, and D. Oro. 2015. Colony size and foraging range in seabirds. *Oikos*:n/a-n/a.
- Lascelles, B. G., P. R. Taylor, M. G. R. Miller, M. P. Dias, S. Oppel, L. Torres, A. Hedd, M. Le Corre, R. A. Phillips, S. A. Shaffer, H. Weimerskirch, and C. Small. 2016. Applying global criteria to tracking data to define important areas for marine conservation. *Diversity and Distributions*

22:422–431.

- Lindgren, F., H. Rue, and J. Lindström. 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73:423–498.
- Lombardo, L., T. Opitz, and R. Huser. 2019. Numerical Recipes for Landslide Spatial Prediction Using R-INLA. Pages 55–83 *in* H. Reza Pourghasemi and C. Gokceoglu, editors. *Spatial Modeling in GIS and R for Earth and Environmental Sciences*. Elsevier.
- Longhurst, A. R. 1998. *Ecological geography of the sea*. 2nd edition. Academic Press, San Diego.
- Longhurst, A., S. Sathyendranath, T. Platt, and C. Caverhill. 1995. An estimate of global primary production in the ocean from satellite radiometer data. *Journal of Plankton Research* 17:1245–1271.
- MacArthur, R. H., and E. R. Pianka. 1966. On Optimal Use of a Patchy Environment. *American Naturalist*, The 100:603–609.
- Matthiopoulos, J. 2003. The use of space by animals as a function of accessibility and preference. *Ecological Modelling* 159:239–268.
- Morera-Pujol, V., P. Catry, M. C. Magalhães, C. Péron, J. M. Reyes-González, J. P. Granadeiro, T. Militão, M. P. Dias, D. Oro, G. Dell’Omo, M. Müller, V. H. Paiva, B. Metzger, V. C. Neves, J. Navarro, G. Karris, S. Xirouchakis, J. G. Cecere, A. Zamora-López, M. G. Forero, F. De Felipe, Z. Zájková, M. Cruz-Flores, D. Grémillet, J. González-Solís, and R. Ramos. 2019. Detecting recurrent sources of variability in animal tracking studies. *Ecological Applications*.
- Navarro, J., and J. González-Solís. 2009. Environmental determinants of foraging strategies in Cory’s shearwaters *Calonectris diomedea* breeding on the Canary Island, NE Atlantic. *Marine Ecology Progress Series* 378:259–267.
- Oppel, S., A. Beard, D. Fox, E. Mackley, E. Leat, L. Henry, E. Clingham, N. Fowler, J. Sim, J. Sommerfeld, N. Weber, S. Weber, and M. Bolton. 2015. Foraging distribution of a tropical seabird supports Ashmole’s hypothesis of population regulation. *Behavioral Ecology and Sociobiology*.

- Oppel, S., M. Bolton, A. P. B. Carneiro, M. P. Dias, J. A. Green, J. F. Masello, R. A. Phillips, E. Owen, P. Quillfeldt, A. Beard, S. Bertrand, J. Blackburn, P. D. Boersma, A. Borges, A. C. Broderick, P. Catry, I. Cleasby, E. Clingham, J. Creuwels, S. Crofts, R. J. Cuthbert, H. Dallmeijer, D. Davies, R. Davies, B. J. Dilley, H. A. Dinis, J. Dossa, M. J. Dunn, M. A. Efe, A. L. Fayet, L. Figueiredo, A. P. Frederico, C. Gjerdrum, B. J. Godley, J. P. Granadeiro, T. Guilford, K. C. Hamer, C. Hazin, A. Hedd, L. Henry, M. Hernández-Montero, J. Hinke, N. Kokubun, E. Leat, L. M. Tranquilla, B. Metzger, T. Militão, G. Montrond, W. Mullié, O. Padget, E. J. Pearmain, I. L. Pollet, K. Pütz, F. Quintana, N. Ratcliffe, R. A. Ronconi, P. G. Ryan, S. Saldanha, A. Shoji, J. Sim, C. Small, L. Soanes, A. Takahashi, P. Trathan, W. Trivelpiece, J. Veen, E. D. Wakefield, N. Weber, S. Weber, L. Zango, F. Daunt, M. Ito, M. P. Harris, M. A. Newell, S. Wanless, J. González-Solís, and J. Croxall. 2018. Spatial scales of marine conservation management for breeding seabirds. *Marine Policy* 98:37–46.
- Orians, G. H., and N. E. Pearson. 1979. On the theory of central place foraging. Pages 155–177 *in* D. J. Horn, R. D. Mitchell, and G. R. Stairs, editors. *Analysis of ecological systems*. Ohio State University Press, Columbus.
- Papastamatiou, Y. P., Y. Y. Watanabe, U. Demšar, V. Leos-Barajas, D. Bradley, R. Langrock, K. Weng, C. G. Lowe, A. M. Friedlander, and J. E. Caselle. 2018. Activity seascapes highlight central place foraging strategies in marine predators that never stop swimming. *Movement Ecology* 6:9.
- Passos, C., J. Navarro, A. Giudici, and J. González-Solís. 2010. Effects of extra mass on the pelagic behavior of a seabird. *The Auk* 127:100–107.
- Péron, C., M. Authier, and D. Grémillet. 2018. Testing the transferability of track-based habitat models for sound marine spatial planning. *Diversity and Distributions*:1–16.
- Ramos, R., J. P. Granadeiro, B. Rodríguez, J. Navarro, V. H. Paiva, J. Bécares, J. M. Reyes-González, I. Fagundes, A. Ruiz, J. M. Arcos, J. González-Solís, and P. Catry. 2013. Meta-population feeding grounds of Cory's shearwater in the subtropical Atlantic Ocean: implications for the definition of Marine Protected Areas based on tracking studies. *Diversity and Distributions* 19:1284–1298.

- Renner, I. W., and D. I. Warton. 2013. Equivalence of MAXENT and Poisson Point Process Models for Species Distribution Modeling in Ecology. *Biometrics* 69:274–281.
- Reyes-González, J. M., Z. Zajková, V. Morera-Pujol, F. De Felipe, T. Militão, G. Dell’Araccia, R. Ramos, J. M. Igual, J. M. Arcos, and J. González-Solís. 2017. Migración y ecología espacial de las poblaciones españolas de pardela cenicienta. Page (M. nº 3 del programa Migra, Ed.). SEO/BirdLife, Madrid.
- Rue, H., S. Martino, and N. Chopin. 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 71:319–392.
- Scales, K. L., P. I. Miller, C. B. Embling, S. N. Ingram, E. Pirotta, and S. C. Votier. 2014. Mesoscale fronts as foraging habitats: composite front mapping reveals oceanographic drivers of habitat use for a pelagic seabird. *Journal of The Royal Society Interface* 11:20140679.
- Simpson, D. P., J. B. Illian, F. K. Lindgren, S. H. Sørbye, and H. Rue. 2011. Going off grid: Computationally efficient inference for log-Gaussian Cox processes. *Biometrika*:1–26.
- Simpson, D. P., H. Rue, T. G. Martins, A. Riebler, and S. H. Sørbye. 2014. Penalising model component complexity: A principled, practical approach to constructing priors:1–42.
- Soldatini, C., Y. V. Albores-Barajas, A. Ramos-Rodríguez, A. Munguia-Vega, E. González-Rodríguez, C. Catoni, and G. Dell’Omo. 2019. Tracking reveals behavioural coordination driven by environmental constraints in the Black-vented Shearwater *Puffinus opisthomelas*. *Population Ecology*:1–13.
- Sørbye, S. H., J. B. Illian, D. P. Simpson, D. Burslem, and H. Rue. 2019. Careful prior specification avoids incautious inference for log-Gaussian Cox point processes. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 68:543–564.
- Sørbye, S. H., and H. Rue. 2014. Scaling intrinsic Gaussian Markov random field priors in spatial modelling. *Spatial Statistics* 8:39–51.
- Thibault, J.-C., V. Bretagnolle, and C. Rabouam. 1997. Cory’s shearwater.

- Pages 75–98 *Birds of the Western Palearctic Update*. Oxford University Press, Oxford.
- Thiebault, A., R. Mullers, P. Pistorius, M. A. Meza-Torres, L. Dubroca, D. B. Green, and Y. Tremblay. 2014. From colony to first patch: processes of prey searching and social information in Cape Gannets. *Auk* 131:595–609.
- Tobler, W. R. 1970. A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography* 46:234.
- Trevail, A. M., J. A. Green, J. Sharples, J. A. Polton, J. P. Y. J. P. Y. Arnould, and S. C. Patrick. 2019. Environmental heterogeneity amplifies behavioural response to a temporal cycle. *Oikos* 128:1–12.
- Valdés, L., and I. Déniz-González. 2015. Oceanographic and biological features in the Canary Current Large Marine Ecosystem. Intergovernmental Oceanographic Commission - UNESCO, Paris.
- Wakefield, E. D., T. W. Bodey, S. Bearhop, J. Blackburn, K. Colhoun, R. Davies, R. G. Dwyer, J. A. Green, D. Grémillet, A. L. Jackson, M. J. Jessopp, A. Kane, R. H. W. Langston, A. Lescroël, S. Murray, M. Le Nuz, S. C. Patrick, C. Péron, L. M. Soanes, S. Wanless, S. C. Votier, and K. C. Hamer. 2013. Space partitioning without territoriality in gannets. *Science* 341:68–70.
- Wakefield, E. D., R. A. Phillips, and J. Matthiopoulos. 2009. Quantifying habitat use and preferences of pelagic seabirds using individual movement data: a review. *Marine Ecology Progress Series* 391:165–182.
- Ward, P., and A. Zahavi. 1972. The importance of certain assemblages of birds as “information-centres” for food-finding. *Ibis* 115:517–534.
- Weimerskirch, H. 2007. Are seabirds foraging for unpredictable resources? *Deep-Sea Research Part II: Topical Studies in Oceanography* 54:211–223.
- Weimerskirch, H., S. Bertrand, J. Silva, J. C. Marques, and E. Goya. 2010. Use of social information in seabirds: Compass rafts indicate the heading of food patches. *PLoS ONE* 5:e9928.
- Wood, S. N. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73:3–36.

Worton, B. J. 1989. Kernel Methods for Estimating the Utilization Distribution in Home-Range Studies. *Ecology* 70:164–168.

5

Discussion

In this thesis, we discuss the drawbacks and main sources of variability associated with the inference of higher-level (i.e. colony, population and species) distributions from individual tracking data, and we provide analytical tools to detect this variability as well as to model spatially explicit data, and apply them to (a) detect and discuss the implications of the migratory connectivity present in three pelagic seabirds, (b) study the ecological segregation among the three species in their non-breeding season, (c) demonstrate the existence of spatial segregation among foraging areas of three neighbouring colonies, and (d) discuss the role of information transfer in the at-sea distribution of foraging seabird. These results are highly relevant for the fields of spatial and seabird ecology and for conservation, since they provide tools applicable to individual movement data regardless of species and tracking device and

provide insights into the drivers and implications of movement for pelagic seabirds.

5.1 Addressing sources of variability in the inference from individual tracking data

The study of space and habitat use from a multi-colony point of view is affected by several sources of bias or variability, introduced by inferring the space use of higher levels (population or species) from the individual tracks of a few individuals (Giuggioli and Bartumeus 2012). To assess the effects of individual site fidelity and environmental variability on the inference of space use at higher levels, in **chapter one**, we developed a test, applicable to any individual tracking dataset. Regardless of species tracked and type of device used, this test detects either significant effects of site fidelity or environmental variability at an individual level in the population-level distribution. In addition, when designing multi-colony tracking studies, it is important to maximise not only the number of animals tracked, but also the extent of the sampled area (in case of spatially structured populations such as seabirds, the number of colonies sampled), since animals from different areas of the distribution of a given species might face different environmental conditions, and thus make a different use of space (Börger et al. 2006). To that aim, we developed a function to test the ability of each sampled population to properly represent the entire distribution of the species. To demonstrate the applicability of these three functions, we used a multi-species, multi-colony dataset of 1,346 year-round tracks from 805 individuals of three *Calonectris* shearwaters breeding in 34 colonies. First, we used the

developed functions to demonstrate the lack of effect of both individual site fidelity and environmental variability in the distributions at a population level. Then, we calculated the species representativity level of each population for the three species. The most remarkable result was that obtained from the Cory's shearwater (*C. borealis*) non-breeding distribution (Fig. 17), where single-population distributions always underrepresented that of the entire species, providing evidence of the need to consider the above-mentioned issues when inferring species space use from individual tracking data. In addition, linear models demonstrated a spatial gradient in the degree of species representativeness of each population for both Cory's and Scopoli's shearwaters (*C. diomedea*; Fig. 18). For Cory's shearwater, the relationship with latitude was weak (although significant), and probably a by-product of the fact that the largest colonies are those to the north of the distribution, holding around 65% of the estimated world population (Fontaine et al. 2011): northern colonies were better at representing the entire species because they are, in fact, the majority of the species. For Scopoli's shearwaters, however, the relation between representativeness and distance to the species centre was strongly negative. This shows that colonies closer to the centre of mass of the breeding distribution were more representative of the entire species, proving that using data from only a small part of the entire breeding distribution will lead to an underrepresentation, or even an erroneous estimation, of the species non-breeding distribution (Webster et al. 2002)

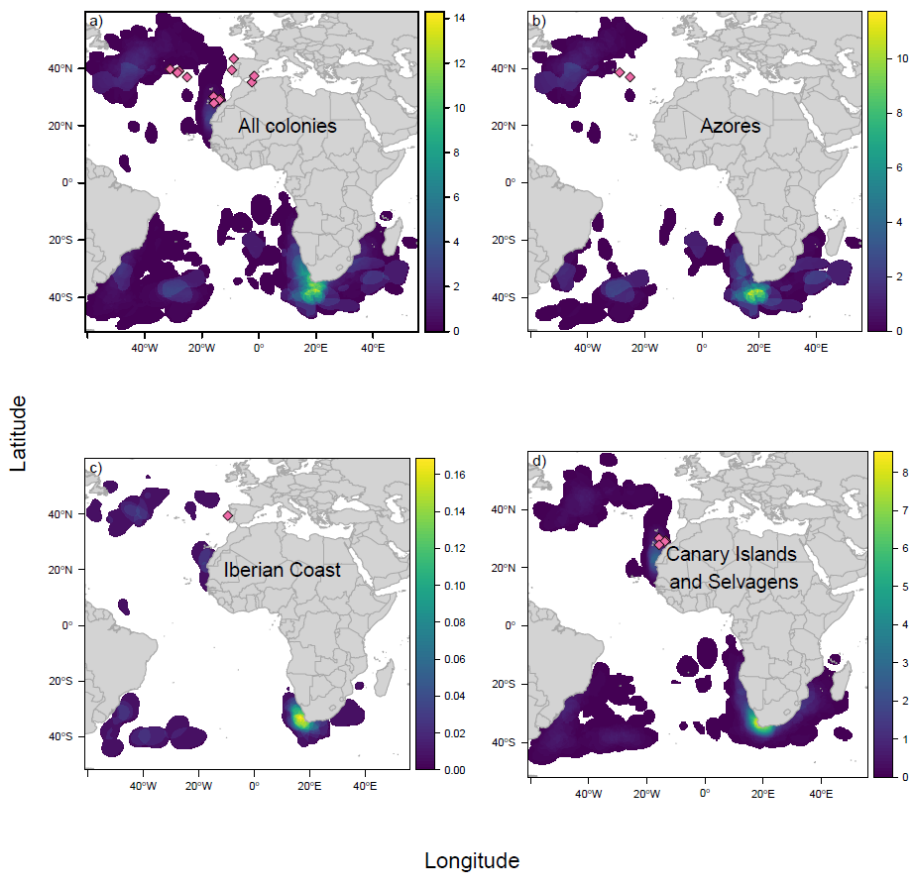


Figure 17 (previous page)

Non-breeding distributions of Cory's shearwaters corrected for sample effort and weighted by population size. (a) Using data from all representative colonies pooled together to represent the species' non-breeding distribution, and using data only from the representative colonies in (b) Azores, (c) the Iberian coast, and (d) Canary and Salvages Islands. Pink diamonds in (a) show the location of all sampled colonies. In (b), (c) and (d), pink diamonds show representative colonies for each population. Scales show the number of individuals per $0.1^{\circ} \times 0.1^{\circ}$ cell.

According to these results, it is clear that using only individuals tracked from the Iberian Coast population would cause a gross underestimation of the breeding distribution of the entire species. Using only birds from the Azores population would not underestimate the total area so dramatically, but would miss the non-breeding area off the Northwest African coast entirely.

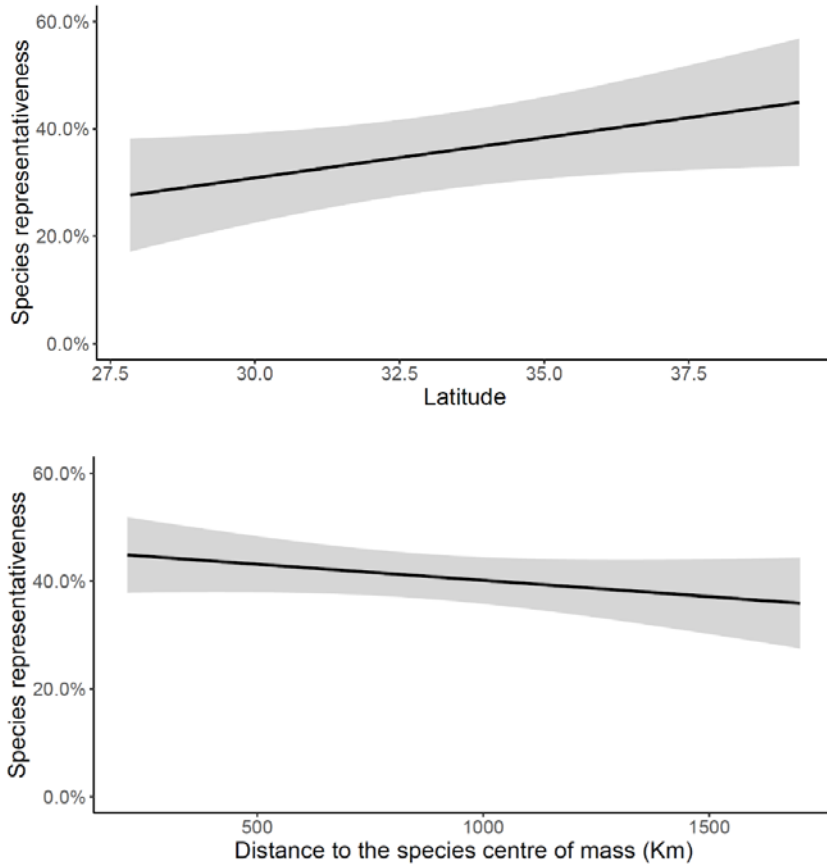


Figure 18

Species representativeness for each population was calculated as a measure of how well the non-breeding distribution of a single colony represents that of the entire species. In spatially structured species with high migratory connectivity, we would expect representativeness values to be relatively low for all colonies, since non-breeding distributions would not be shared among colonies.

Significant relationship between species representativeness values and colony latitude for Cory's shearwater (top), and representativeness value and distance to the species centre of mass for Scopoli's shearwater (bottom). The black line represents a linear fit and the grey shaded area represents the 95% CI.

In addition to the issues addressed in chapter one, the serial correlation between consecutive positions inherent to high-resolution tracking data can also lead to over- or underestimation of space use when used, for example, in species distribution models (Holdo and Roach 2013). In **chapter three**, we addressed this issue by using a spatial random thinning technique to subset the data while maintaining the spatial structure of the distribution at a population level. Besides serial correlation of tracking data, there is, by definition, spatial structure in distributional data (i.e. spatial autocorrelation). Although the aim of species distribution modelling techniques is to explain all this structure using a combination of different environmental variables, certain spatial structure usually remains unexplained, which reduces the prediction accuracy of any model (Aarts et al. 2008). In **chapter three**, we applied a spatially explicit point pattern model that includes a spatial field in the linear predictor, able to capture any spatial structure absent in the covariates (Fig 19). This field provides key information on the spatial structure of unobserved environmental covariates or behavioural process non-included in the final modelling. The information provided by this spatial field became essential while evaluating the accuracy of model predictions.

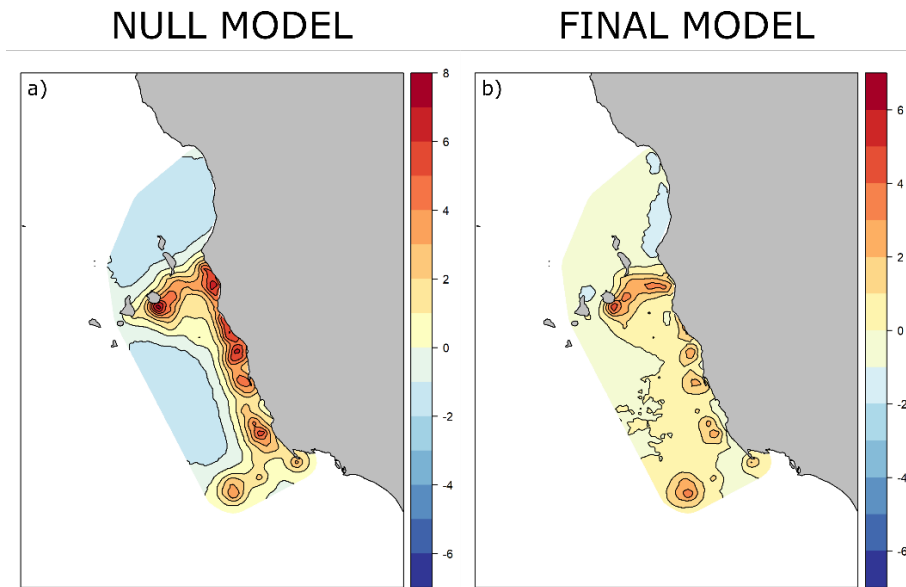


Figure 19

To be able to appreciate the relevance of the environmental variables in the spatial structure of the foraging locations, we ran first a null model, that contained no environmental variables and only a spatial field which captured all the structure present on the data. Comparing this spatial field to the one from the full model, only representing the structure not captured by the covariates, allowed us to see where the hotspots of foraging birds were explained away by the environmental variables and where those were not enough and some structure remained, indicating possible effects of local enhancement or the consequences of information transfer at the colony.

Representation of the spatial field capturing the spatial structure of the data in (a) the null model, and (b) in the final model including all the relevant covariates. The intensity of the spatial field is closer to 0 all across the study area in the final model, demonstrating the spatial structure captured by the included covariates. However, some structure remains that the covariates could not capture, and with the use of these type of models we can obtain an estimation of the shape of the “leftover” structure, which can provide information about the unobserved environmental or behavioural process that may be causing this structure.

5.2 Space and habitat use in pelagic seabirds

In **chapter two**, we applied those methods developed in chapter one to study the migratory connectivity and non-breeding distributions of the three shearwater species from a multi-colony point of view. The analyses showed, for Cory’s and Scopoli’s shearwaters, a pattern of population structure —indicated by positive values of migratory connectivity— present among distance-based groups of colonies (i.e. populations) but absent at colony level (Fig. 20). This suggests that, for both species, breeding colonies are structured in populations with a low degree of among-population mixing but high degree of within-population mixing in the non-breeding areas.

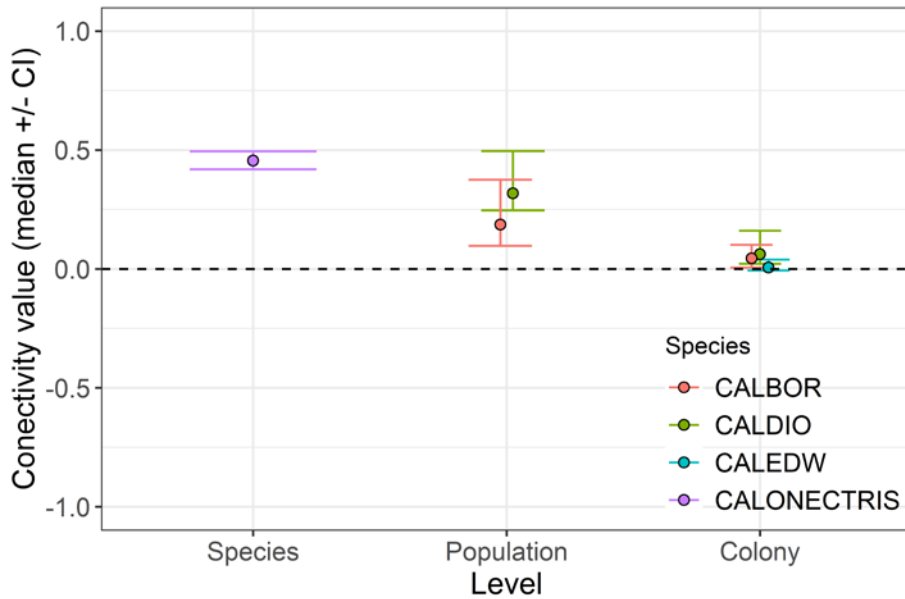


Figure 20 (previous page)

Migratory connectivity (MC) values are a measure of how well the structure among breeding populations is maintained in the non-breeding distributions. High values of MC indicate that populations that breed separated do not mix in the non-breeding areas either, while low values of MC indicate low high degree of mixing among populations in the non-breeding areas.

MC values at species (CALONECTRIS) level, at a population level for Cory's (CALBOR) and Scopoli's (CALDIO) shearwater and at a colony level for Cory's, Scopoli's and Cape Verde (CALEDW) shearwater. Closed circles represent median value, and whiskers represent 95% confidence interval).

This structuring in populations has relevant implications in the evolutionary history of these species, since it can promote processes of ecological divergence (Matthiopoulos et al. 2005). In addition, the migratory connectivity among the three closely-related shearwater species showed even higher levels than those computed among populations within each species, suggesting very little inter-species mixing in their non-breeding distributions.

This among-species segregation was confirmed with ensemble ecological niche models (EENM; Fig. 21), where the spatial segregation between Cory's and Scopoli's shearwaters, and between Cape Verde and Scopoli's shearwaters, were unmistakable, and the environmental preferences confirmed an ecological segregation between Cory's and Cape Verde shearwaters even though their non-breeding areas partially overlapped (Fig 22).

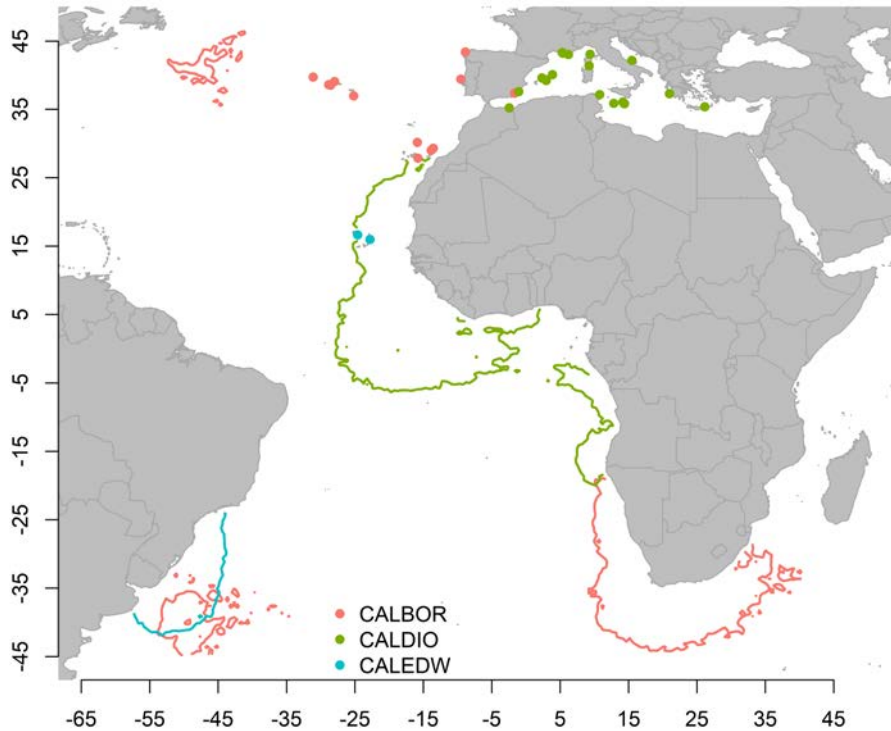


Figure 21

Using ensemble environmental niche models we obtained the non-breeding habitat suitability map for the three *Calonectris taxa*, and from the maps calculated the polygons that contain habitat with a suitability value $>80\%$. These polygons show no overlap between the non-breeding habitats of Cory's and Scopoli's shearwaters, and between Cape Verde and Scopoli's shearwaters, and only partial overlap between the non-breeding habitats of Cory's and Cape Verde shearwaters, indicating a clear spatial segregation among the non-breeding habits of the three species

Polygons delimiting the areas with habitat suitability of $>80\%$ for Cory's (pink), Scopoli's (green) and Cape Verde (blue) shearwaters, showing the complete segregation between Cory's and Scopoli's shearwater, and Cape Verde and Scopoli's shearwaters, and partial segregation of Cory's and Cape Verde shearwaters. Closed circles show the locations of the sampled colonies for each species.

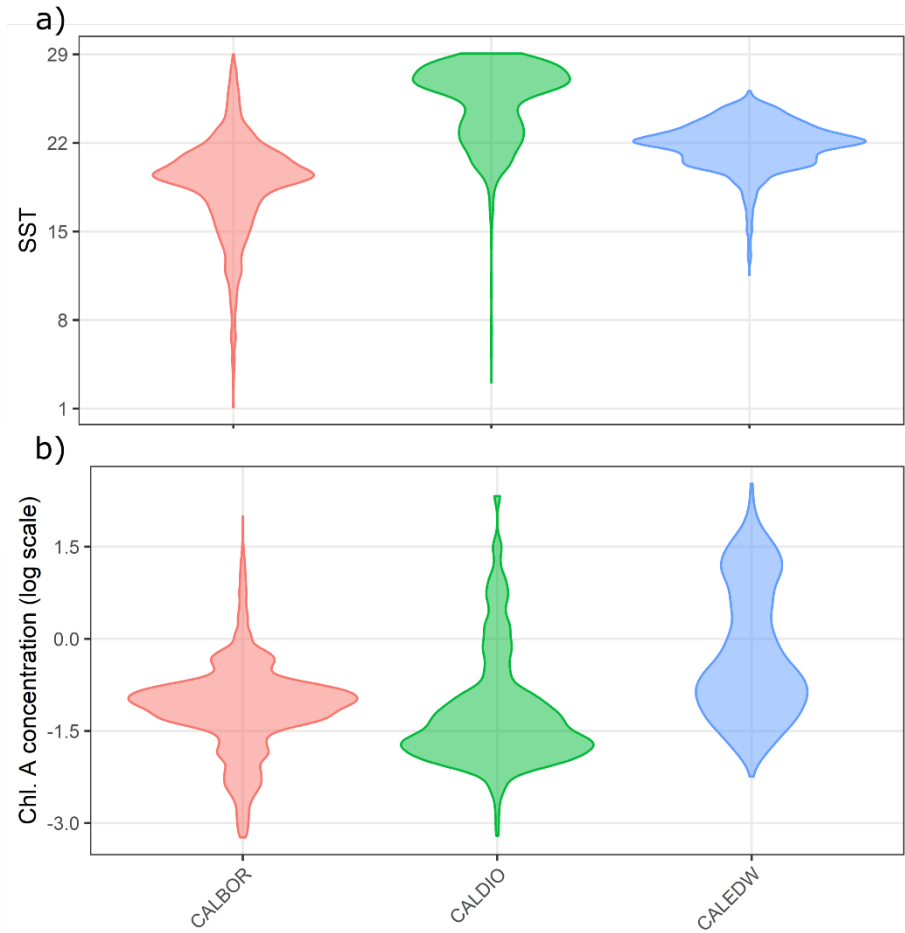


Figure 22

From the EENM we obtained, for the two most important variable, the preferred values for each taxa. Differences in environmental preferences demonstrated an ecological segregation between Cory's and Scopoli's shearwater, and also between Cory's and Cape Verde shearwaters.

Violin plots representing the values selected in the three environmental niche models at a taxon level for the three most important variables: (a) SST, (b) CHLA. The points represent the value of the environmental variable at each of the presences, and the width of the violin represents the density of points.

These results provide evidence of an ecological segregation in the non-breeding areas/season among the three species, which has been described as a signal of divergence in the lineage in migratory seabirds (Friesen et

al. 2007; Friesen 2015). Thus, in combination with genetic and genomic analyses, our results could provide robust evidences of the divergence within the *Calonectris* complex, and support the identification of these taxa as full species, statuses that are still under discussion (Gómez-Díaz and González-Solís 2007; Gómez-Díaz et al. 2009; Genovart et al. 2013; Zidat et al. 2017).

In **chapter three**, using a dataset with a much higher spatiotemporal resolution, we demonstrated the existence of spatial segregation in the usage of adjacent foraging areas of three neighbouring colonies of Cory's shearwaters (Fig. 23). we also proved the existence of spatial structure in the at-sea locations of birds breeding in these colonies that was unexplained by the environmental covariates retained in the linear predictor. These findings allowed me to discuss potential mechanisms of information transfer among long-ranging shearwaters, and, in general, the implications that such information transfer might have in the spatial distribution and ecology of seabirds. Local enhancement, which happens when information is transferred at the foraging places, has already been described for many species of seabirds including some shearwaters, but, mainly, from an observational point of view (Buckley 1997).

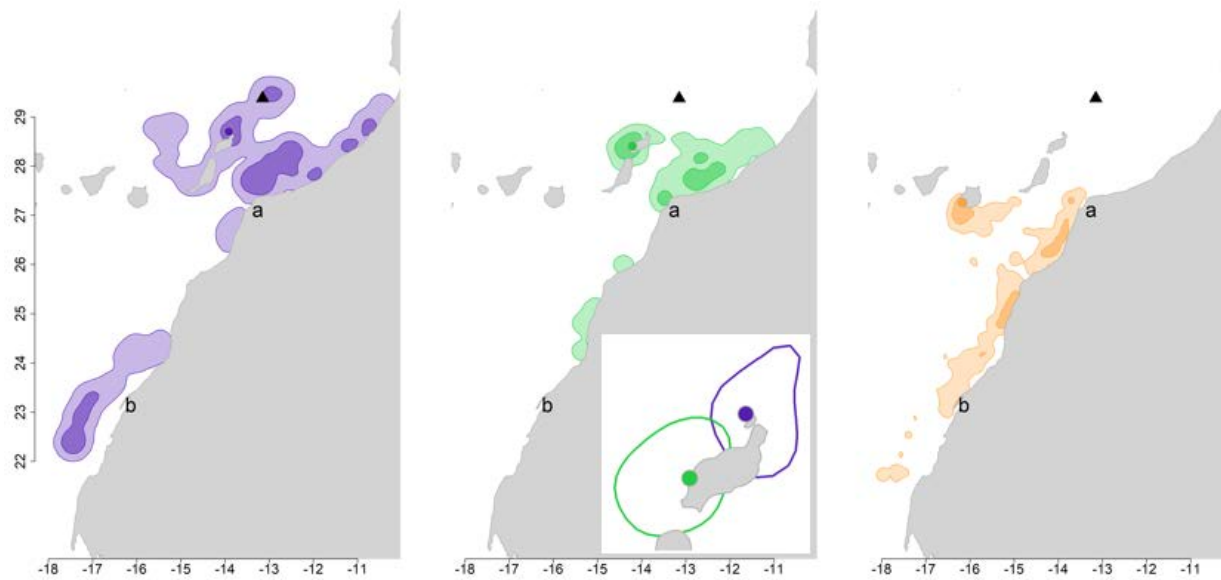


Figure 23

Segregation among the foraging distributions of Cory's shearwaters from three neighbouring colonies. Birds from Montaña Clara (purple) and Timanfaya (green) foraged mainly to the north of the colonies and the Canary Current, while birds from the Veneguera colony (tan) foraged mainly to the south. This segregation was more intense than what would be expected at random.

Core areas (darker colours) and home ranges (lighter colours) from the foraging positions of birds breeding in Montaña Clara, Timanfaya and Veneguera. Darker circles mark the location of the corresponding colony. The black triangle marks the location of the Concepcion Seamount. Also shown are the locations of Cape Juby (a) and the Dakhla Peninsula (b). The inset in the middle plot shows a zoom of the overlap between the adjacent core areas of the Montaña Clara and Timanfaya colonies.

The spatially-explicit models allowed me to detect clustering of foraging locations at short distances (<10 kilometres) through the use of L-functions and envelopes (Fig 24), but further development of point pattern spatiotemporal models should improve the detection of such aggregations, since local enhancement has an important temporal component that cannot be modelled with static methodologies (Yuan et al. 2017). The information transfer at or near the colonies has been formally tested in species with short foraging ranges (e.g. Wakefield et al. 2013, Jones et al. 2018), but never in long-ranging ones, although some studies demonstrating segregation in foraging areas have cited information transfer as a possible driver (reviewed in Bolton et al. 2018). The spatial field present in the linear predictor of our spatially-explicit models allowed us to detect a spatial structure in at-sea locations not captured by environmental covariates, but the improvement of spatiotemporal models to jointly model the distributions of different colonies (Bakka et al. 2018) will certainly represent a step forward towards the detection and formal description of the effects of information transfer in foraging distributions.

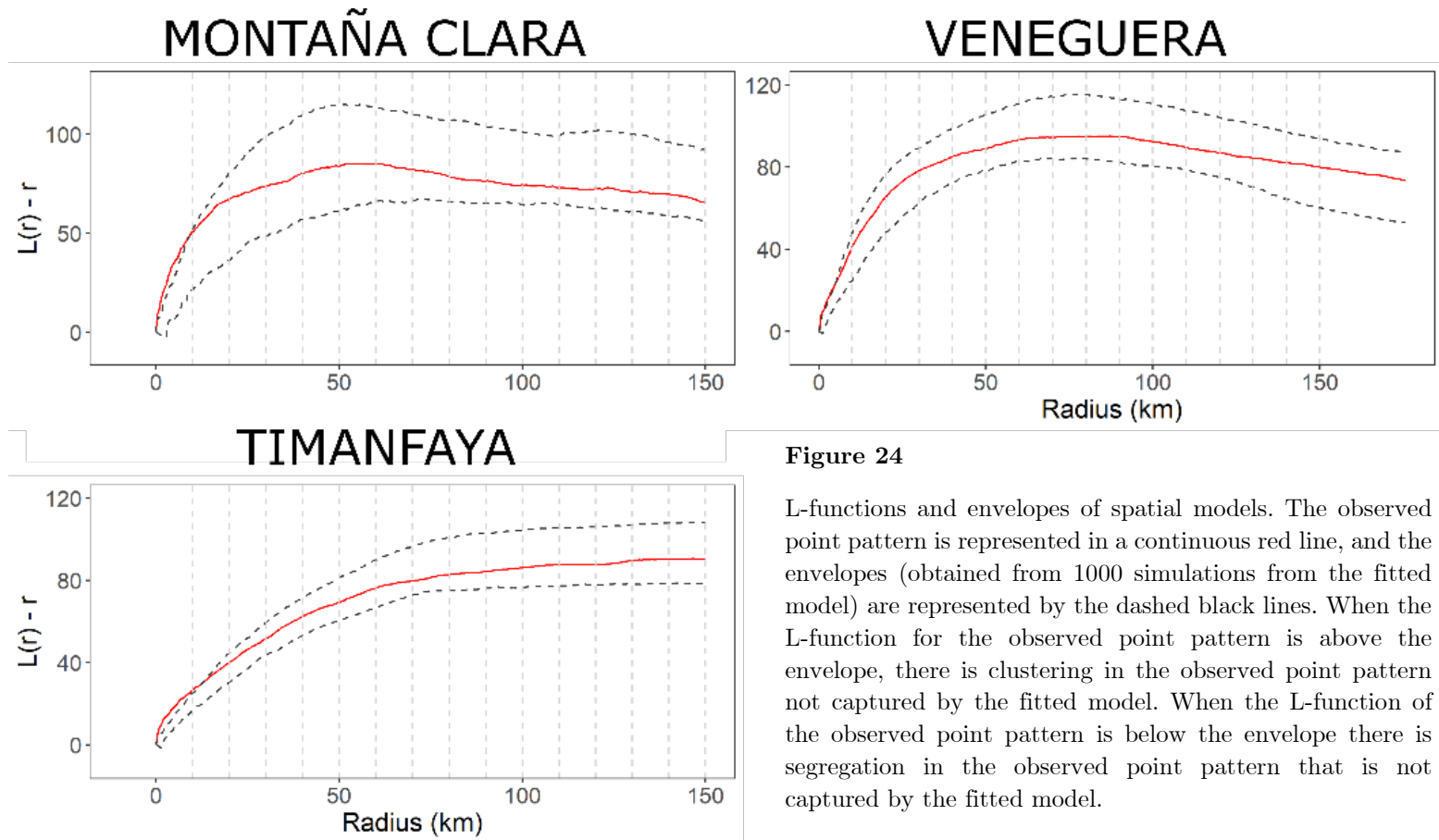


Figure 24

L-functions and envelopes of spatial models. The observed point pattern is represented in a continuous red line, and the envelopes (obtained from 1000 simulations from the fitted model) are represented by the dashed black lines. When the L-function for the observed point pattern is above the envelope, there is clustering in the observed point pattern not captured by the fitted model. When the L-function of the observed point pattern is below the envelope there is segregation in the observed point pattern that is not captured by the fitted model.

5.3 Implications for conservation of the findings of this work

The increasing availability of animal tracking data certainly has an impact in management and conservation policies, as it provides useful information for combating species decline (Sherley et al. 2017), it helps defining the limits of protected areas in an effective way (Hays et al. 2019), it serves as an indicator of ecosystems health (Kays et al. 2015), and has even been used to detect harmful human activities such as illegal fishing (Weimerskirch et al. 2017).

In **chapter one**, this thesis provides methods to remove biases and sources of variability in tracking data, which should contribute to increase the robustness and credibility of studies using individual tracking data to identify areas of ecological importance, delimit protected areas, or use the spatial distribution of mobile species to manage human activities that could be threatening biodiversity (e.g. temporal fishery closures). In addition, we test these tools in a multi-colony dataset spanning the entire breeding distribution of three species of pelagic seabirds, providing large-scale evidence of the relevance of offshore areas for the conservation of migratory species.

In **chapter two**, our findings on the spatial structure in the non-breeding distributions of *Calonectris* shearwaters both among and within species, obtained from the migratory connectivity and EENM, could not only aid in the design of offshore marine protected areas (MPA) in the temperate regions the species inhabit, but also offer a case study that should aid in the designation of other MPAs elsewhere based on the

distributions of other top predators. Results demonstrating population structure within species also have relevance in conservation, since they contribute to the understanding of how perturbations in the non-breeding regions can affect specific breeding populations, and what implications this can have in their population dynamics, which ultimately conditions the species demography.

Tracks of foraging trips have also contribute to the understanding of the effect of off-shore installations, or other anthropogenic impacts, such as interactions with fisheries, in different colonies (e.g. Genovart et al. 2018, Cook et al. 2018). In addition, this type of data has been used to establish the limits of MPAs both in national and internationally managed waters (Hays et al. 2019). Methods developed in **chapter three**, which analyse foraging distributions of marine fauna, can aid in the identification of marine hotspots, something that is essential in the development of management measures of intensely exploited areas such as the Canary Current where the studied species forage. In addition, our results proving the spatial segregation between the foraging distributions, both in the areas surrounding the colony and in the productive waters of the Canary Current, represent key insights into what consequences any anthropogenic perturbation (e.g. an off-shore wind farm or oil installation) could have for each of the sampled colonies. Extending our analysis to other breeding colonies with birds also foraging in the Canary Current could aid to the understanding of the consequences that these perturbations could have not only at a multi-colony level, but also at a species level.

5.4 References

- Aarts G, MacKenzie M, McConnell BJ, Fedak M, Matthiopoulos J. 2008. Estimating space-use and habitat preference from wildlife telemetry data. *Ecography (Cop)*. 31(1):140–160. doi:10.1111/j.2007.0906-7590.05236.x.
- Bakka H, Rue H, Fuglstad GA, Riebler A, Bolin D, Illian JB, Krainski E, Simpson DP, Lindgren FK. 2018. Spatial modeling with R-INLA: A review. *Wiley Interdiscip Rev Comput Stat.(May)*:1–24. doi:10.1002/wics.1443.
- Bolton M, Conolly G, Carroll M, Wakefield ED, Caldow R. 2018. A review of the occurrence of inter-colony segregation of seabird foraging areas and the implications for marine environmental impact assessment. *Ibis (Lond)* 185(9). doi:10.1111/ibi.12677. <http://doi.wiley.com/10.1111/ibi.12677>.
- 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. *J Anim Ecol*. 75(6):1393–1405. doi:10.1111/j.1365-2656.2006.01164.x.
- Buckley NJ. 1997. Spatial-Concentration Effects and the Importance of Local Enhancement in the Evolution of Colonial Breeding in Seabirds. *Am Nat*. 149(6):1091–1112. doi:10.1086/286040. <https://www.journals.uchicago.edu/doi/10.1086/286040>.
- Cook ASCP, Humphreys EM, Bennet F, Masden EA, Burton NHK. 2018. Quantifying avian avoidance of offshore wind turbines: Current evidence and key knowledge gaps. *Mar Environ Res*. 140(June):278–288. doi:10.1016/j.marenvres.2018.06.017. <https://linkinghub.elsevier.com/retrieve/pii/S014111361830179X>.
- Fontaine R, Gimenez O, Bried J. 2011. The impact of introduced predators, light-induced mortality of fledglings and poaching on the dynamics of the Cory's shearwater (*Calonectris diomedea*) population from the Azores, northeastern subtropical Atlantic. *Biol Conserv*. 144(7):1998–2011. doi:10.1016/j.biocon.2011.04.022. <http://dx.doi.org/10.1016/j.biocon.2011.04.022>.
- Friesen VL. 2015. Speciation in seabirds: why are there so many species...and

- why aren't there more? J Ornithol. doi:10.1007/s10336-015-1235-0.
<http://link.springer.com/10.1007/s10336-015-1235-0>.
- Friesen VL, Burg TM, McCoy KD. 2007. Mechanisms of population differentiation in seabirds. Mol Ecol. 16(9):1765–1785. doi:10.1111/j.1365-294X.2006.03197.x.
<http://doi.wiley.com/10.1111/j.1365-294X.2006.03197.x>.
- Genovart M, Bécarea J, Igual J, Martínez-Abraín A, Escandell R, Sánchez A, Rodríguez B, Arcos JM, Oro D. 2018. Differential adult survival at close seabird colonies: The importance of spatial foraging segregation and bycatch risk during the breeding season. Glob Chang Biol. 24(3):1279–1290. doi:10.1111/gcb.13997.
<https://onlinelibrary.wiley.com/doi/abs/10.1111/gcb.13997>.
- Genovart M, Thibault J-C, Igual JM, Bauzá-Ribot MDM, Rabouam C, Bretagnolle V. 2013. Population Structure and Dispersal Patterns within and between Atlantic and Mediterranean Populations of a Large-Range Pelagic Seabird. Votier S, editor. PLoS One. 8(8):e70711. doi:10.1371/journal.pone.0070711.
<https://dx.plos.org/10.1371/journal.pone.0070711>.
- Giuggioli L, Bartumeus F. 2012. Linking animal movement to site fidelity. J Math Biol. 64(4):647–656. doi:10.1007/s00285-011-0431-7.
<http://link.springer.com/10.1007/s00285-011-0431-7>.
- Gómez-Díaz E, González-Solís J. 2007. Geographic assignment of seabirds to their origin: combining morphologic, genetic, and biogeochemical analyses. Ecol Appl. 17(5):1484–1498. doi:10.1890/06-1232.1.
- Gómez-Díaz E, González-Solís J, Peinado MA. 2009. Population structure in a highly pelagic seabird, the Cory's shearwater *Calonectris diomedea*: an examination of genetics, morphology and ecology. Mar Ecol Prog Ser. 382:197–209.
- Hays GC, Bailey H, Bograd SJ, Bowen WD, Campagna C, Carmichael RH, Casale P, Chiaradia A, Costa DP, Cuevas E, et al. 2019. Translating Marine Animal Tracking Data into Conservation Policy and Management. Trends Ecol Evol. 34(5):459–473. doi:10.1016/j.tree.2019.01.009.
<https://linkinghub.elsevier.com/retrieve/pii/S0169534719300242>.

- Holdo RM, Roach RR. 2013. Inferring animal population distributions from individual tracking data: Theoretical insights and potential pitfalls. *J Anim Ecol.* 82(1):175–181. doi:10.1111/j.1365-2656.2012.02031.x.
- Jones TB, Patrick SC, Arnould JPY, Rodríguez-Malagón MA, Wells MR, Green JA. 2018. Evidence of sociality in the timing and location of foraging in a colonial seabird. *Biol Lett.* 14(7):20180214. doi:10.1098/rsbl.2018.0214. <http://rsbl.royalsocietypublishing.org/lookup/doi/10.1098/rsbl.2018.0214>.
- Kays R, Crofoot MC, Jetz W, Wikelski M. 2015. Terrestrial animal tracking as an eye on life and planet. *Science* (80-). 348(6240):aaa2478. doi:10.1126/science.aaa2478. [accessed 2018 Mar 20]. <http://www.sciencemag.org/cgi/doi/10.1126/science.aaa2478>.
- Matthiopoulos J, Harwood JL, Thomas L. 2005. Metapopulation consequences of site fidelity for colonially breeding mammals and birds. *J Anim Ecol.* 74(4):716–727. doi:10.1111/j.1365-2656.2005.00970.x.
- Sherley RB, Ludynia K, Dyer BM, Lamont T, Makhado AB, Roux JP, Scales KL, Underhill LG, Votier SC. 2017. Metapopulation Tracking Juvenile Penguins Reveals an Ecosystem-wide Ecological Trap. *Curr Biol.* 27(4):563–568. doi:10.1016/j.cub.2016.12.054. <http://dx.doi.org/10.1016/j.cub.2016.12.054>.
- Wakefield ED, Bodey TW, Bearhop S, Blackburn J, Colhoun K, Davies R, Dwyer RG, Green JA, Grémillet D, Jackson AL, et al. 2013. Space partitioning without territoriality in gannets. *Science* (80-). 341:68–70. <http://www.sciencemag.org/cgi/doi/10.1126/science.1236077>.
- Webster MS, Marra PP, Haig SM, Bensch S, Holmes RT. 2002. Links between worlds: unraveling migratory connectivity. *Trends Ecol Evol.* 17:76–83.
- Weimerskirch H, Filippi DP, Collet J, Waugh SM, Patrick SC. 2017. Use of radar detectors to track attendance of albatrosses at fishing vessels. *Conserv Biol.* doi:10.1111/cobi.12965. <http://doi.wiley.com/10.1111/cobi.12965>.
- Yuan Y, Bachl FE, Lindgren FK, Borchers DL, Illian JB, Buckland ST, Rue H, Gerrodette T. 2017. Point process models for spatio-temporal distance sampling data from a large-scale survey of blue whales. *Ann*

Appl Stat. 11(4):2270–2297. doi:10.1214/17-AOAS1078.
<https://projecteuclid.org/euclid.aoas/1514430286>.

Zidat T, Dell’Ariccia G, Gabirot M, Sourrouille P, Buatois B, Celerier A, Bonadonna F, Crochet P-A. 2017. Reproductive isolation maintains distinct genotypes, phenotypes and chemical signatures in mixed colonies of the two European *Calonectris* shearwaters (Procellariiformes: Procellariidae). *Zool J Linn Soc.* 181(3):711–726. doi:10.1093/zoolinnea/zlx002.
<http://academic.oup.com/zoolinnea/article/181/3/711/3811184>.

6

Conclusions

- **About addressing sources of variability in the inference from individual tracking data.**
 1. In the current scientific context of collaborative, multi-colony studies, it is important to understand the effect that sources of variability in individual tracking data such as individual site fidelity, temporal variability and spatial variability can have in the inference of space use at colony, multi-colony, population and species level. We have provided a protocol for an objective evaluation of these three main sources of variability that can affect the inference from individual tracking data to the distributions at a multi-colony level.
 2. The development of cheaper, smaller, and more precise tracking devices has boosted the use of individual tracking data for species distribution modelling. We have shown the applicability of a

spatially explicit, continuously indexed point process model to model species distributions from tracking data. These improve the commonly used species distribution models by removing the need to aggregate data in grids and allowing the incorporation of non-linear covariates and a spatial field to the linear predictor.

- **About the space and habitat use of pelagic seabirds at a multi-colony level.**

3. We have demonstrated the existence of spatial segregation among the non-breeding distributions of Cory's, Scopoli's, and Cape Verde shearwaters. This spatial segregation provides evidence of the ecological divergence within the *Calonectris* complex, further supporting the identification of each taxa as a full species.
4. We have also proved a higher degree of migratory connectivity between populations than between colonies, indicating the presence of spatial structure at a population level, in Cory's and Scopoli's shearwaters.
5. Using tracking data of three neighbouring colonies of Cory's shearwater, We have demonstrated the existence of spatial segregation among the foraging distributions of each colony both, in adjacent as well as in distant waters. This segregation was mediated by environmental drivers and information transfer, which occurred both at sea and at the breeding colonies.

- **About the implications for conservation of the findings of this work.**

6. Tracking data is being increasingly used to informing conservation strategies, such as identifying protected areas, defining extractive moratoria or implementing mitigation

strategies. The tools developed in this thesis offer robustness to studies that, using individual tracking data, aim to define conservation and management strategies of highly vagile species.

7. At a large (ocean) scale, we have shown. the relevance of certain pelagic areas for all the populations of *Calonectris* shearwaters, which can aid in the identification and delimitation of offshore marine protected areas (MPA) and offer guidelines for the identification of other MPAs based on tracking data of other top predators.
8. I have demonstrated the existence of structure both, among and within each of the three *Calonectris* species, in their non-breeding distribution, which contribute understanding the impacts anthropogenic perturbations could have for each of the populations and species.
9. Lastly, analyses of the foraging distributions of birds from three colonies of Cory's shearwater in the Canary Current provide insights into the potential consequences of anthropogenic activities, such as fisheries, off-shore wind farm or oil installation, may have at colony and population levels. Extending these analyses to other breeding colonies with individuals also foraging in the Canary Current could aid understanding the consequences these perturbations may have not only at a multi-colony level, but also at a species level.

Annex 1

Table A1

R functions developed specifically for this paper to test the effects of individual site fidelity, temporal and spatial variability in tracking data. The use of the functions and their arguments are described in detail.

Function name	Use	Arguments	Input format
indEffectTest()	Test the effect that individual site fidelity has in the distribution at a population level	Tracks	Tracking data. DataFrame (DF) or SpatialPolygonsDataFrame (SPDF), containing Latitude, Longitude, a unique identifier for each trip and all the grouping variables we are interested in comparing the within group and between group overlap
		tripID	Character string, name of the column in the data containing the unique identifier for each trip
		GroupVar	Character string, name of the column in the data the variable to use to group the trips and compare “within-group” and “between-group” effects. To test the effect of individual site fidelity we use individual identity as grouping variable, and to test the inter-annual differences we use year as grouping variable
		Method	Character string, one of the methods of measuring similarity between kernels offered by the adehabitatHR::kerneloverlap() function.
		Conditional	Logical (TRUE/FALSE), from the adehabitatHR::kerneloverlap() function. If set to TRUE (the default), the overlap is calculated between the two incomplete probability surfaces bound by the polygons calculated at the percentage indicated by UDLev. In this case, the maximum level of overlap

Function name	Use	Arguments	Input format
			will be equal to UDLev. If set to false the overlap is calculated between the two complete probability surfaces, and the UDLev argument becomes irrelevant. For more info check <code>?adehabitatHR::kerneloverlap()</code>
		UDLev	Numeric, Utilisation Distribution Level. Level of the KDE to be selected. For example, if 50 is used, the overlap is calculated between the smallest areas for which the probability to find the animal is equal to 0.50 (but see explanation for argument "conditional". From the <code>adehabitatHR::kerneloverlap()</code> function.
		Scale	Numeric, smoothing parameter for the kernel Utilisation Distribution calculation. It comes from the <code>h</code> argument from <code>adehabitatHR::kerneloverlap()</code> . The function uses projected data (projects it to Lambert Equal Area projection centred in the mean values of longitude and latitude), so this parameter can have a biological interpretation as the spatial scale at which we are interested in comparing distributions. Must be specified in km since the function will internally convert it to meters to match the units of the projection used.
		Grid	Numeric, size of the grid cells on which the UD will be estimated (from the <code>adehabitatHR::kerneloverlap()</code> function.
		plotIt	Logical (T/F). If TRUE, a boxplot of within and between group overlap values will be plotted

Function name	Use	Arguments	Input format
		Nboots	Numeric, number of iterations for the bootstrap K-S test. For more information, check ?Matching::boot.ks()
simulateDistribution()	simulate distribution from tracking data, avoiding differential sampling bias and correcting for different population sizes	Tracks PopulationInfo Scale Grid MultiFactor	Tracking data. DF or SPDF, containing Latitude, Longitude, Species and Population DF containing the name of all sampled populations within the study species (named as they appear in the tracking data set) and their size estimates (number of pairs, the function will multiply it by 2). Numeric, in km, the smoothing parameter for the estimation of the UD. Note that, whatever the projection of the original data, the function will reproject to a CRS in m and then multiply the "Scale" value by 1000 Numeric, size of the grid cells on which the UD will be estimated (from the adehabitatHR::kerneloverlap() function). Numeric. Number of points simulated for each individual of the colony. If it equals 1, the number of points simulated will be equal to the size estimate (i.e. from a colony of 100 breeding pairs we will obtain a data set of 200 simulated locations). If there are very small colonies in our data set, which might be inadequate for generating accurate kernel density estimates, a higher multiplication factor can be specified. However, higher values will generate very

Function name	Use	Arguments	Input format
bootstrapColony()	check a population's representativity of the distribution of the entire species	Tracks	large data sets, increasing the computation time of this function and the next one. Tracking data. DF or SPDF, containing Latitude, Longitude, Species and Population
		SimulatedDistributions	DF or SPDF containing the simulated distribution of the studied species, with at least the columns Longitude, Latitude, Population and Species. Can be directly the output of the simulateDistribution() function
		UDLev	Numeric, Utilisation Distribution Level. Level of the KDE to be selected. For example, if 50 is used, the overlap is calculated between the smallest areas for which the probability to find the animal is equal to 0.50 (but see explanation for argument "conditional". From the adehabitatHR::kerneloverlap() function.
		Scale	Numeric, in km, the smoothing parameter for the estimation of the UD. Note that, whatever the projection of the original data, the function will reproject to a CRS in m and then multiply the "Scale" value by 1000
		Grid	From the kernelUD function. A number giving the size of the grid on which the UD should be estimated (for more info check ?adehabitatHR::kernelUD)
		Iterations	Numeric, number of iterations for the bootstrap

Table A2

Summary of logger and deployment details for each colony, prior publication of the data used for this study, and permits and licences granted for the corresponding fieldwork

Colony	Type of logger	Model	Attachment method	Weight (in gr)	% of body mass	Prior publication	Permits/Licences
Sisargas	Migrate Technology	C250	Cable tie to leg-mounted darvic ring	2.5	0.33	Unpublished	Authorisation from the Autonomous Organism of National Parks and the Galician Regional Government (Xunta de Galicia)
Berlenga	Biotrack/ BAStrack	Mk5, Mk3	Cable tie to leg-mounted darvic ring	2.5 – 3.3	0.33 – 0.44	Missagia et al., 2015	No. 89/2011/CAPT
	Biotrack/ BAStrack	Mk5, Mk7, MK3005	Cable tie to metal ring	3.6	0.48	Catry, Dias, Phillips, & Granadeiro, 2011	107/2006, 116/2007, 333/2007/CAPT
Vila	Biotrack/ BAStrack	Mk4, Mk18, Mk19, Mk3005, MK4083	Cable tie to leg-mounted darvic ring	1.5 – 2.5	0.20 – 0.33	González-Solís, Croxall, Oro, & Ruiz, 2007	Lic. 61/2008/DRA; Lic. 66/2009/DRA; Lic. 02/2010/DRA; Lic.43/2011/DRA; Lic.63/2012/DRA
Pico	Biotrack/ BAStrack	Mk4, MK19	Cable tie to leg-mounted darvic ring	1.5 – 2.5	0.20 – 0.33	Unpublished	Lic. 61/2008/DRA; Lic. 66/2009/DRA; Lic. 02/2010/DRA; Lic.43/2011/DRA; Lic.63/2012/DRA

Colony	Type of logger	Model	Attachment method	Weight (in gr)	% of body mass	Prior publication	Permits/Licences
Graciosa	Biotrack/ BAStrack	MK4, MK19	Cable tie to leg-mounted darvic ring	1.5 – 2.5	0.20 – 0.33	Unpublished	Lic. 61/2008/DRA; Lic. 66/2009/DRA; Lic. 02/2010/DRA; Lic.43/2011/DRA; Lic.63/2012/DRA
Faial	Biotrack/ BAStrack	MK4, MK19	Cable tie to leg-mounted darvic ring	1.5 – 2.5	0.20 – 0.33	Unpublished	Lic. 61/2008/DRA; Lic. 66/2009/DRA; Lic. 02/2010/DRA; Lic.43/2011/DRA; Lic.63/2012/DRA
Corvo	Biotrack/ BAStrack	MK4, MK19	Cable tie to leg-mounted darvic ring	1.5 – 2.5	0.20 – 0.33	Unpublished	Lic. 61/2008/DRA; Lic. 66/2009/DRA; Lic. 02/2010/DRA; Lic.43/2011/DRA; Lic.63/2012/DRA
Selvagem	Biotrack/ BAStrack	Mk7, Mk19	Cable tie to metal ring	2.6 – 3.6	0.33 – 0.48	Dias, Granadeiro, & Catry, (2012, 2013)	107/2006, 116/2007, 07/2009S, 107/2010/CAPT, 4/2011S
Timanfaya	Migrate Technology	C250	Cable tie to leg-mounted darvic ring	2.5	0.33	Unpublished	RES-AUT I03/2015/JACOB GONZÁLEZ-SOLÍS

Colony	Type of logger	Model	Attachment method	Weight (in gr)	% of body mass	Prior publication	Permits/Licences
Montaña Clara	Biotrack/ BAStrack	Mk19, Mk3005	Cable tie to leg-mounted darvic ring	2.5	0.33	Unpublished	2011/0795, 2015/1170
	Migrate Technology	C250, C330	Cable tie to leg-mounted darvic ring	2.5 – 3.3	0.33 – 0.45		
Veneguera	Biotrack/ BAStrack	Mk13, Mk18, Mk19, Mk3005, Mk4,Mk9	Cable tie to leg-mounted darvic ring	1.5 – 2.5	0.20 – 0.45	González-Solís, Croxall, Oro, & Ruiz, 2007	84/2007, 2011/0795, 2015/1170
	Migrate Technology	C250,C330	Cable tie to leg-mounted darvic ring	2.5 – 3.3	0.33 – 0.45		
Chafarinas	Biotrack/ BAStrack	Mk13,Mk19	Cable tie to leg-mounted darvic ring	2.5	0.33	Unpublished	Ethics Committee of CSIC (Spanish Council of Scientific Research, (REF: 239/2015))
	Migrate Technology	C250	Cable tie to leg-mounted darvic ring	2.5	0.33		
Terreros	Migrate Technology	C330	Cable tie to leg-mounted darvic ring	3.3	0.45	Unpublished	Lic. date 21/04/2017 from Dirección General de Gestión del Medio Natural (Junta Andalucía)
Chafarinas	Biotrack/ BAStrack	Mk13, Mk19	Cable tie to leg-mounted darvic ring	2.5	0.35	Unpublished	Ethics Committee of CSIC (Spanish Council of

Colony	Type of logger	Model	Attachment method	Weight (in gr)	% of body mass	Prior publication	Permits/Licences
	Migrate Technology	C250	Cable tie to leg-mounted darvic ring	2.5	0.35		Scientific Research, (REF: 239/2015)
Palomas	Migrate Technology	C250, C330	Cable tie to leg-mounted darvic ring	2.5 – 3.3	0.35 – 0.46	Ramos, 2019	A.UP/2016/0083
Cala Morell	Migrate Technology	C250 – C330	Cable tie to leg-mounted darvic ring	2.5 – 3.3	0.35 – 0.46	Ramos, 2019	CEP-24/2015, CEP-30/2016
Pantaleu	Biotrack/ BAstrack	Mk13, Mk19, Mk3005	Cable tie to leg-mounted darvic ring	1.5 – 2.5	0.21 – 0.35	González-Solís, Croxall, Oro, & Ruiz, 2007	X32/2010; I40/2011; Q22/2012; P39/2013; M52/2014; F08/2015
Cabrera	Migrate Technology	C250	Cable tie to leg-mounted darvic ring	2.5	0.35	Unpublished	0160/15
Porquerolles	LOTEK	Lat2500	Cable tie to leg-mounted darvic ring	3.6	0.50	Unpublished	#A34 - 369, #A34 - 505
Riou	LOTEK	Lat2500	Cable tie to leg-mounted darvic ring	3.6	0.50	Péron & Grémillet, 2013	#A34 - 369, #A34 - 505

Colony	Type of logger	Model	Attachment method	Weight (in gr)	% of body mass	Prior publication	Permits/Licences
Frioul	LOTEK	Lat2500	Cable tie to leg-mounted darvic ring	3.6	0.50	Unpublished	#A34 - 369, #A34 - 505
Giraglia	LOTEK	Lat2500	Cable tie to leg-mounted darvic ring	3.6	0.50	Unpublished	#A34 - 369, #A34 - 505
Lavezzi	LOTEK	Lat2500	Cable tie to leg-mounted darvic ring	3.6	0.50	Péron et al., 2012	#A34 - 369, #A34 - 505
Linosa	Biotrack/ BAStrack	Mk3006	Cable tie to leg-mounted metal ring	3.6	0.5	Unpublished	Prot. 65887 (23/07/2007) Prot. 17233 (01/12/2010)
Gozo	Biotrack/ BAStrack	Mk3005	Cable tie to leg-mounted darvic ring	2.5	0.35	Unpublished	NP16/12 NP20/13, NP30/14, NP22/15
Malta	Biotrack/ BAStrack	Mk3005	Cable tie to leg-mounted darvic ring	2.5	0.35	Unpublished	NP16/12 NP20/13, NP30/14, NP22/15
Filfla	Biotrack/ BAStrack	Mk3005	Cable tie to leg-mounted darvic ring	2.5	0.35	Unpublished	NP16/12 NP20/13, NP30/14, NP22/15
Zembra	Migrate Technology	C250	Cable tie to leg-mounted darvic ring	2.5	0.35	Ramos, 2019	Authorisation by the Agence de Protection et d'Aménagement du Littoral (APAL)

Colony	Type of logger	Model	Attachment method	Weight (in gr)	% of body mass	Prior publication	Permits/Licences
Tremiti	Biotrack/ BAStrack	Mk3006				Unpublished	
Strofades	LOTEK	Lat250	Cable tie to leg-mounted darvic ring	3.6	0.50	Karris, 2014	2009-2014, Permissions from Management Agency of the National Marine Park of Zakynthos
Paximada	Migrate Technology	C250	Cable tie to leg-mounted darvic ring	2.5	0.35	Ramos, 2019	110553/1398 (19/05/2014)
Curral Velho	Biotrack/ BAStrack	Mk4, Mk9, Mk10, Mk13, Mk18, Mk19, Mk3005	Cable tie to leg-mounted darvic ring	1.5 – 2.5	0.32 – 0.53	González-Solís et al., 2009 Roscales, Gómez-Díaz,	01/2009, 002/10, 01/2011, 01/2012, 04/2013, 18/2014, 04/2016
	Migrate Technology	C250, C330	Cable tie to leg-mounted darvic ring	2.5 – 3.3	0.53 – 0.69	Neves, & González-Solís, 2011	
Raso	Biotrack/ BAStrack	Mk4, Mk9, Mk13, Mk14, Mk19	Cable tie to leg-mounted darvic ring	1.4 – 2.5	0.29 – 0.53		

Table A3

Number of individuals tracked repeatedly for each colony. In Veneguera and Curral Velho, there are animals tracked for up to 8 consecutive years, in Pantaleu up to 3 consecutive years.

Species	Population	Colony	Consecutive years tracked	No. birds
Cory's shearwater	Canary Islands and Selvagens	Veneguera	1	49
			2	29
			3	16
			4	6
			5	10
			6	4
			7	2
			8	5
Scopoli's shearwater	Balearic Is.	Pantaleu	1	23
			2	14
			3	9
Cape Verde shearwater	East Cape Verde	Curral Velho	1	18
			2	5
			3	3
			4	4
			5	2
			8	1

Table A4

Population estimates from all breeding colonies of Cory's, Scopoli's and Cape Verde shearwaters available in the published literature. n refers to the number of tracks obtained from each colony, Min. pop. and Max. pop. to minimum and maximum population estimates (in number of breeding pairs), and Lon. and Lat. to the longitude and latitude of the breeding colonies.

Species	Population	Colony	N	Min. pop	Max. pop	Lon	Lat	Reference
Cory's shearwater	Galicia	Sisargas	2	17	21	-8.84	43.36	Munilla, Genovart, Paiva, & Velando, 2016
		Cíes		21	27	-8.90	42.21	
		Coelleira		23	38	-7.63	43.76	
	Berlengas	Berlenga	37	310	310	-9.51	39.41	Lecoq, Geraldès, & Andrade, 2011
		Cerro da Velha		25	25	-9.50	39.42	
		Estela		10	15	-9.53	39.42	
		Estelão		5	5	-9.53	39.42	
		Manuel Jorge		5	5	-9.53	39.42	
		Farilhão Grande		500	550	-9.55	39.48	
		Farilhão da Cova		100	120	-9.55	39.47	
Farilhão do Nordeste		15	20	-9.54	39.48			
Farilhão dos Olhos		5	10	-9.55	39.47			
	Rabo de Asno		5	10	-9.55	39.47		

Species	Population	Colony	N	Min. pop	Max. pop	Lon	Lat	Reference	
East Azores	Vila - Santa Maria		72	400	400	-25.17	36.94	Ramírez, Geraldés, Meirinho, Amorim, & Paiva, 2008	
		São Miguel		23,826	23,826	-25.44	37.75		
Central Azores	Pico		15	16,000	16,000	-28.54	38.5		
	Graciosa		25	277	12,500	-27.96	39.06		
	Terceira			10,849	10,849	-27.20	38.73		
	Faial		23	5,713	5,713	-28.82	38.58		
	São Jorge			33,626	33,626	-28.02	38.64		
West Azores	Flores			2,869	33,487	-31.20	39.45		
	Corvo		10	33,487	33,487	-31.09	39.70		
Selvagens	Selvagem grande		21 9	29,540	29,540	-15.87	30.14	Granadeiro, Dias, Rebelo, Santos, & Cetry, 2006	
East Canary Islands	Timanfaya		15	350	4,407	-13.83	28.99	Arcos, Bécares, Rodríguez, & Ruiz, 2009	
	Islote de Lobos			1,000	1,000	-13.82	28.76		
	Costa Esquinzo			300	300	-14.01	28.67		
		Roque del Este			50	50	-13.34	29.28	Rodríguez, De León, Martín, Alonso, & Nogales, 2003
		Roque del Oeste			25	25	-13.53	29.31	
		Aleganza			10,000	12,000	-13.51	29.40	
		La Graciosa			300	300	-13.50	29.26	
		Montaña Clara		56	1,000	1,500	-13.53	29.29	

Species	Population	Colony	N	Min. pop	Max. pop	Lon	Lat	Reference
West Canary Islands		Veneguera	30 9	2,000	2,000	-15.79	27.85	Ramos et al., 2013
		Tenerife		3,000	3,000	-16.92	28.47	
		La Gomera		3,000	3,000	-17.10	28.17	
		La Palma		3,000	3,000	-17.69	28.75	
		El Hierro		2,000	2,000	-17.98	27.71	
Madeira		Madeira		1,100	1,900	-16.88	32.80	Geraldes, 2000
		Porto Santo		500	500	-16.34	33.08	Oliveira & Menezes, 2004
		Desertas		1,500	1,500	-16.51	32.53	
Chafarinas		Chafarinas	16	85	85	-2.43	35.18	Afán et al., 2014
Terreros		Terreros	2	30	35	-1.65	37.35	Arcos, Bécares, Rodríguez, & Ruiz, 2009
Scopoli's shearwater	NW Africa	Chafarinas	16	765	765	-2.43	35.18	Afán et al., 2014
		Rachgoun		100	150	-1.48	35.32	Defos du Rau et al., 2015
		Habibas		350	500	-1.13	35.72	
Palomas		Palomas	26	67	123	-1.04	37.57	Arcos, Bécares, Rodríguez, & Ruiz, 2009
Balearic Islands		Cala Morell	23	1,000	6,000	3.88	40.06	
		Pantaleu	10 3	200	200	2.35	39.58	
		Na Foradada	6	178	178	2.98	39.21	
		Na Pobra	9	178	178	2.97	39.20	
		Ibiza and islets		150	275	1.45	38.99	

Species	Population	Colony	N	Min. pop	Max. pop	Lon	Lat	Reference
		Formentera		85	105	1.44	38.71	
		Columbretes		64	64	0.69	39.90	
France		Porquerolles	7	180	370	6.22	43.00	Anselme & Durand, 2012
		Riou	27	280	300	5.39	43.18	
		Frioul	8	70	70	5.30	43.28	
Corsica - Sardinia		Giannutri		50	200	11.10	42.25	Baccetti et al., 2009
		Isolotto d'Ercole		1	30	11.21	42.38	
		Argentarola		35	50	11.08	42.42	
		Palmaiola		4	10	10.47	42.87	
		Cerboli		50	85	10.55	42.86	
		Pianosa		30	50	10.07	42.59	
		La Scola		60	100	10.11	42.58	
		Giraglia	24	38	40	9.41	43.02	Anselme & Durand, 2012
		Lavezzi	18	300	400	9.26	41.34	
		Cerbicale		79	115	9.36	41.55	Baccetti et al., 2009
		Capo Figari		3	20	9.66	40.99	
		Figarolo		30	100	9.64	40.98	
		Tavolara		10	50	9.72	40.91	
		Barrettini		50	100	9.40	41.28	
		Budelli		40	80	9.35	41.28	
		Camere E		40	60	9.59	41.07	

Species	Population	Colony	N	Min. pop	Max. pop	Lon	Lat	Reference
		Camere W		50	90	9.58	41.06	
		Caprera		20	200	9.46	41.18	
		Carpa		80	120	9.36	41.29	
		Corcelli		2	20	9.40	41.30	
		Maddalena		1	50	9.42	41.26	
		Mortorio		40	60	9.60	41.07	
		Nibani		3	30	9.57	41.13	
		Paduleddi		1	5	9.37	41.29	
		Razzoli		30	200	9.35	41.30	
		Santa Maria		50	200	9.37	41.29	
		Spargi		90	150	9.35	41.24	
		Spargiotto		120	180	9.32	41.25	
		Capo Caccia		300	1,600	8.16	40.56	
		Foradada		50	200	8.15	40.57	
		Piana di Alghero		100	200	8.14	40.60	
		Pan di Zucchero		300	400	8.40	39.33	
		Cavoli		5	30	9.53	39.09	
		Isola Rossa di Teulada		1	50	8.72	38.91	
		Toro		500	1,000	8.41	38.86	
		Vacca		1	50	8.45	38.94	

Species	Population	Colony	N	Min. pop	Max. pop	Lon	Lat	Reference
Isole Ponzane		Faraglione di Calzone Muto		1	5	12.97	40.88	
		Gavi		5	10	13.00	40.93	
		Palmarola		100	150	12.86	40.93	
		Ponza		60	100	12.96	40.90	
		Santo Stefano Ponziane		5	10	13.45	40.79	
		Ventotene		25	40	13.43	40.80	
		Zannone		24	30	13.06	40.97	
		Isole Eolie		Alicudi		9	12	14.35
Filicudi				2	20	14.56	38.57	
Lipari				2	20	14.94	38.48	
Panarea				2	20	15.06	38.64	
Salina				10	10	14.84	38.56	
Vulcano				3	6	14.96	38.40	
Ustica				15	20	13.19	38.71	
Sicilian channel		Marettimo		20	50	12.07	37.97	
		Favignana		20	50	12.30	37.93	
		Levanzo		20	50	12.33	38.00	
		Pantelleria		500	5,000	12.00	36.77	
		Lampedusa		20	50	12.58	35.51	
		Lampione		50	70	12.56	35.51	

Species	Population	Colony	N	Min. pop	Max. pop	Lon	Lat	Reference
		Linosa	90	5,000	5,000	12.87	35.87	Giacomo Dell'Omo pers. comm.
		Gozo	4	2,300	2,350	14.19	36.05	Anselme & Durand, 2012
		Malta	13	1,550	2,000	14.43	35.83	
		Comino		20	25	14.34	36.01	
		Filfla	12	200	200	14.41	35.79	
		Zembra	26	113,720	176,750	10.80	37.12	Defos du Rau et al., 2015
		la Galite		250	500	8.93	37.52	
East Algeria		Sriginia Island		90	130	6.89	36.94	Telailia, Boutabia, Bensouilah, & Houhamdi, 2014
		Kef Amor		24	39	7.33	37.08	Samraoui & Samraoui, 2008
Middle Adriatic		Tremiti	8	100	160	15.51	42.14	Baccetti et al., 2009
		San Domino		200	240	15.49	42.11	
		Vis		750	1,050	16.15	43.05	Defos du Rau et al., 2015
		Lastovo		140	175	16.86	42.75	
Diapontia Islands		Diapontia		60	100	19.40	39.86	Fric, Portolou, Manolopoulos, & Kastritis, 2012
Ionian Sea		Strofades	9	5,500	5,500	21.00	37.15	Karris, Xirouchakis, Grivas, Sfenthourakis, & Giokas, 2017
West Crete		Kythira		10	50	22.98	36.22	Fric, Portolou, Manolopoulos, & Kastritis, 2012
		Gramvousa		535	1,000	23.58	35.61	
Aegean Sea		Fournoi		800	1,000	26.48	37.57	
		Mykonos		150	250	25.33	37.45	

Species	Population	Colony	N	Min. pop	Max. pop	Lon	Lat	Reference
		Paros & Antiparos		100	160	25.13	37.07	
		Northern dodecanese		125	250	26.85	37.14	
		Cyclades		165	275	25.61	36.94	
		Anafi		400	700	25.76	36.37	
		Paximada - Dyonisiades	16	710	1,010	26.18	35.35	
		Limnos		50	150	25.21	39.90	
		Sporades		30	80	23.89	39.19	
		Skyros		81	156	24.56	38.91	
Cape Verde shearwater	East Cape Verde	Curral Velho - Boavista	71	33	155	-22.78	15.97	Birdlife International
		Ilhéu Rabo de Junco - Sal		20	20	-22.99	16.70	Militão pers comm.
	West Cape Verde	Branco		1,000	1,000	-24.67	16.66	Hazevoet, 1995
		Raso	29	6,500	6,500	-24.59	16.62	Paiva et al., 2015
		Brava		1,000	10,000	-24.73	14.82	Hazevoet, 1994

Table A5

Posterior distribution of the final model (Model 3) hyperparameters for the three colonies, and observed and predicted abundances by the three models, for the three colonies.

Colony	Montaña Clara		Timanfaya		Veneguera		
	Median	CI	Median	CI	Median	CI	
Posterior range (km)	88.1	(62.2, 129.7)	118.4	(82.6, 172.5)	132.6	(91.9, 193.7)	
Posterior sigma	1.3	(1.0, 1.6)	1.5	(1.2, 1.8)	1.3	(1.06, 1.65)	
Obs. abundance		153		393		281	
Predicted abundance	Model 1	105	(87, 124)	304	(274, 339)	203	(173, 229)
	Model 2	127	(104, 149)	351	(314, 390)	223	(195, 254)
	Model 3	145	(124, 170)	390	(331, 411)	259	(229, 292)

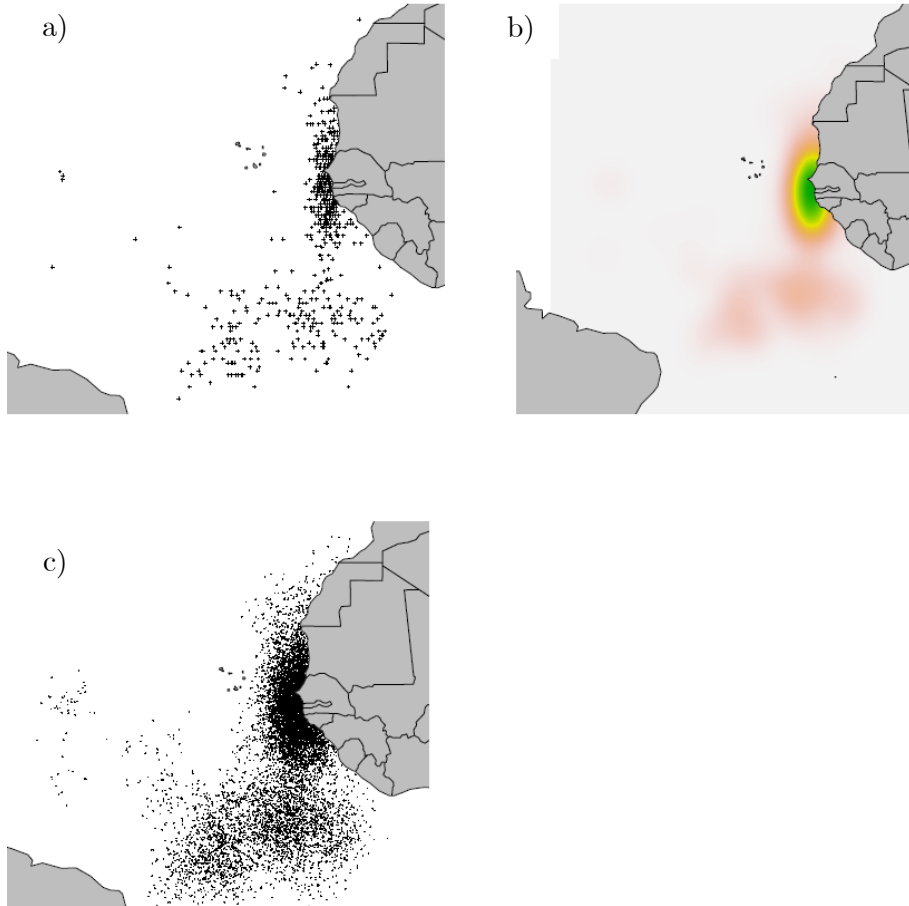


Figure A1

Example illustrating how the `simulateDistribution()` function works: from the wintering positions of all sampled Scopoli's shearwaters breeding in Strofades Island, Greece, (a) it generates the Kernel Density Estimate (KDE, b). Then it simulates one location per individual breeding in Strofades (since we have selected `MultiFactor = 1`) generated from the KDE of the tracking data (c).

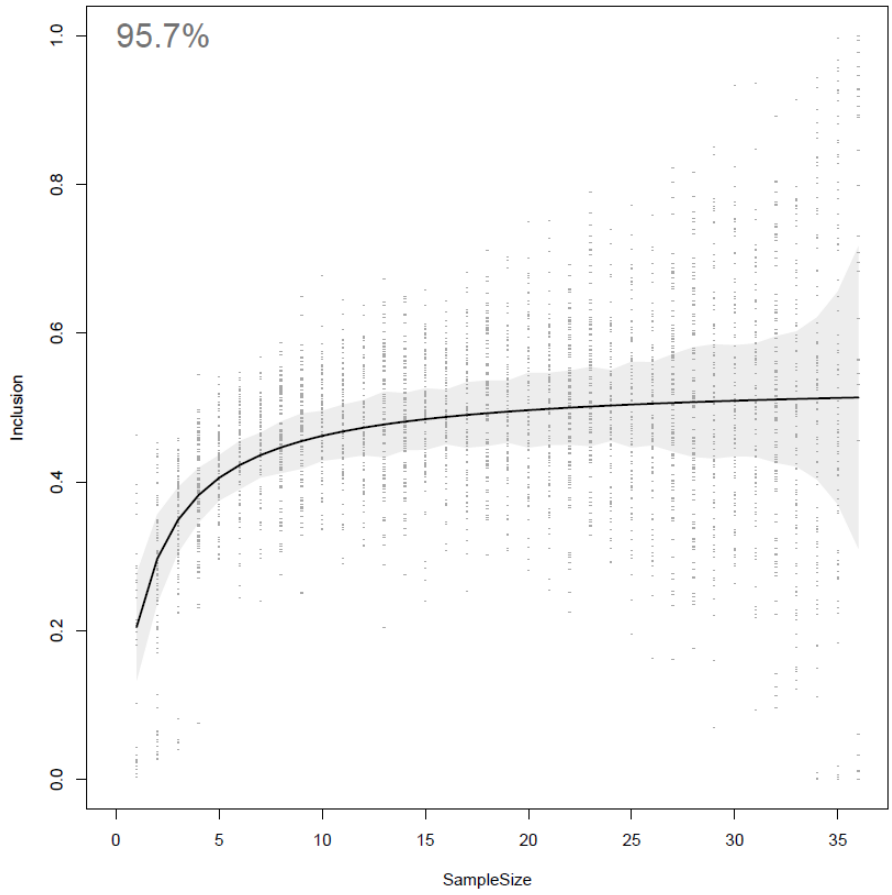


Figure A2

Example illustrating how the colonyBootstrap() function works: from a subsample of a sequentially increasing number of tracks (sample size, x axis) the function calculates the KDE. Then, it calculates the % of points from the simulated Cory’s shearwater distribution that are included in the 50% UD, for each sample size (Inclusion, y axis). This process is repeated 100 times for each sample size, with the tracks selected at random from the total available tracks of the colony. Each of the 100 results is represented as a grey circle. A non-linear regression model is fitted to this data (solid black line) and the value of inclusion at which the asymptote is reached is calculated. The grey area indicates the 95% confidence interval (CI) of the non-linear model.

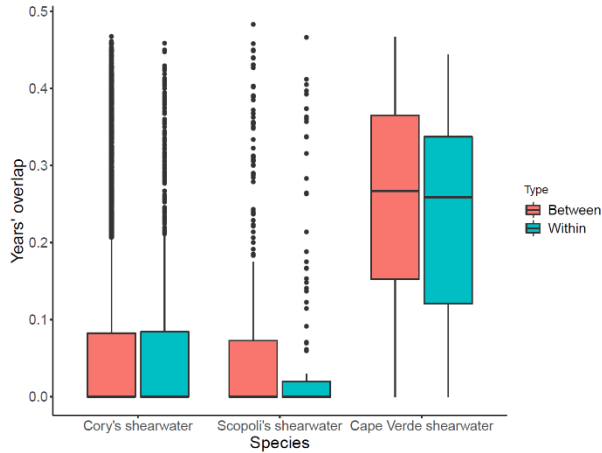


Figure A3:

Between (pink) and within (blue) individual overlap for the three tested colonies. Boxes indicate interquartile range, and the horizontal line represents the median. The whiskers represent the 25 and 75% quantiles and the closed circles represent values outside those quantiles.

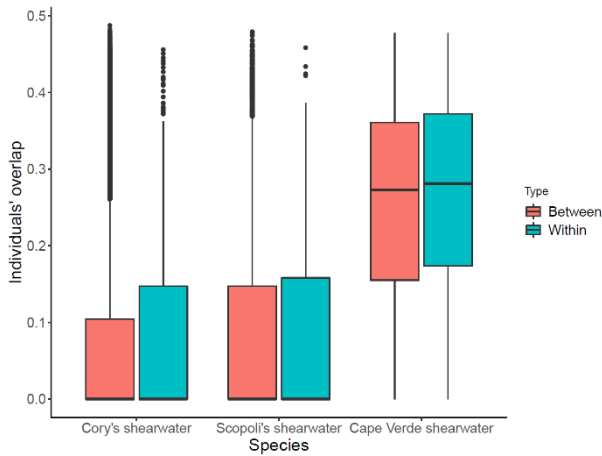


Figure A4:

Between (pink) and within (blue) year overlap for the three tested colonies. Boxes indicate interquartile range, and the horizontal line represents the median. The whiskers represent the 25 and 75% quantiles and the closed circles represent values outside those quantiles.

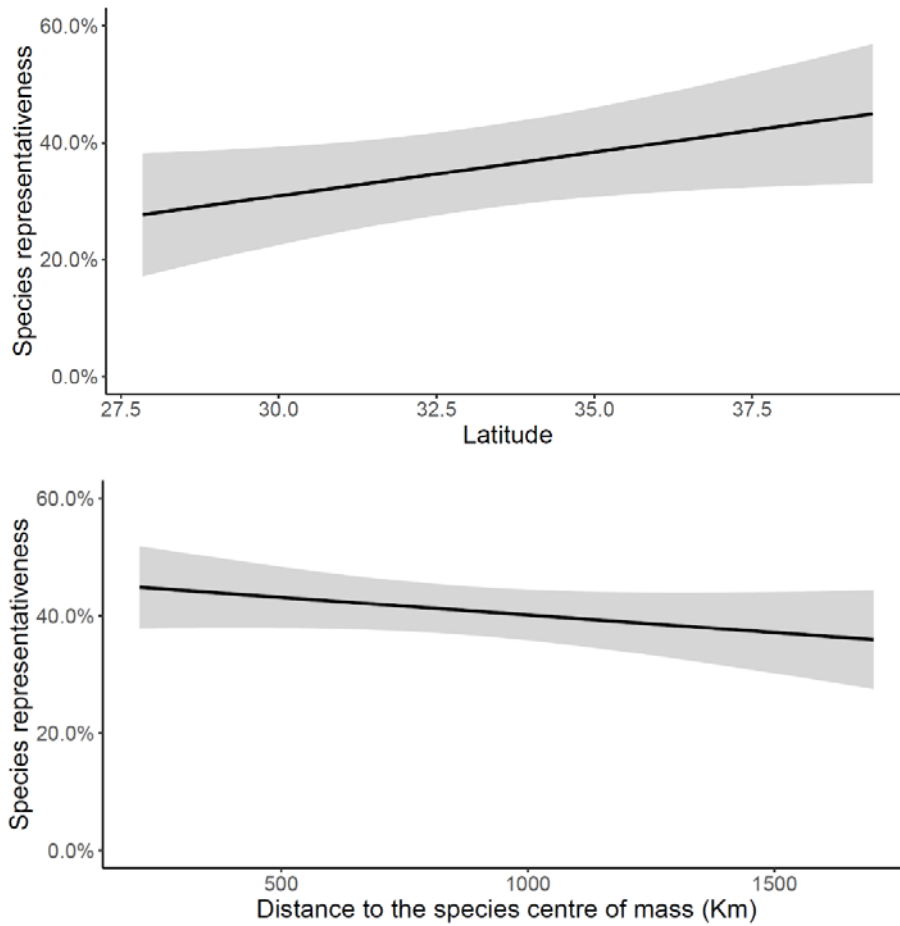


Figure A5

Significant relationship between species representativeness values and colony latitude for Cory's shearwater (top), and representativeness value and distance to the species centre of mass for Scopoli's shearwater (bottom). The black line represents a linear fit and the grey shaded area represents the 95% CI.

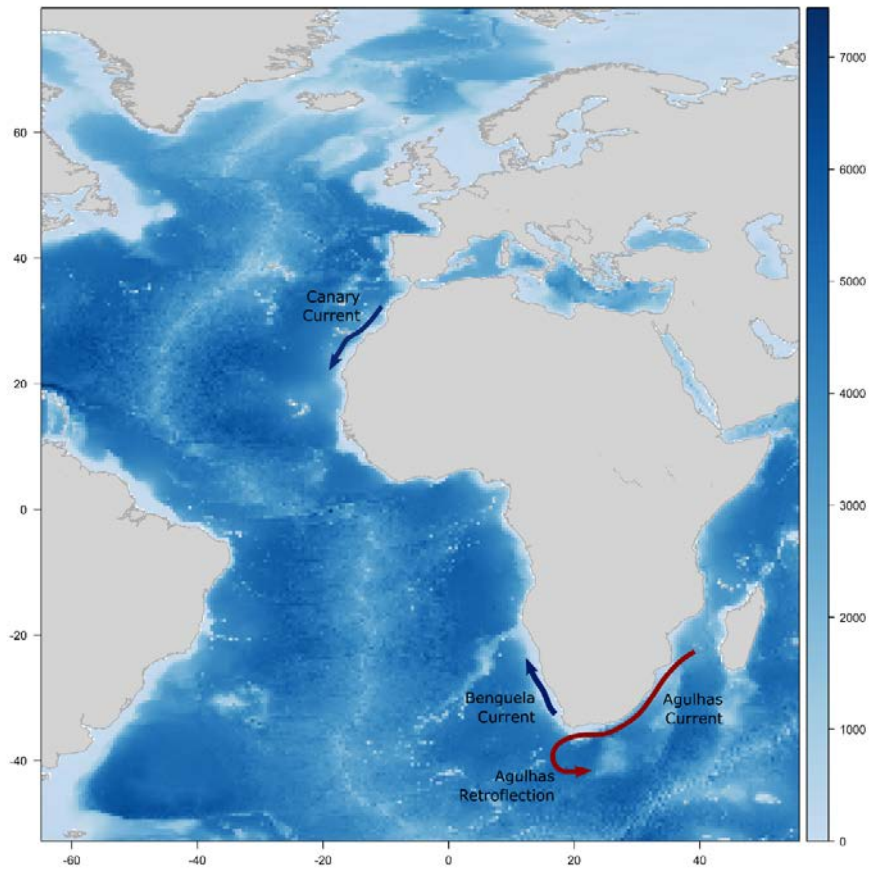


Figure A6

Schematic image of the main currents exploited by Cory's and Scopoli's shearwaters around the African continent. Dark red arrows indicate warm currents and dark blue arrows indicate cold currents. The arrow tip indicates the direction of the flow.

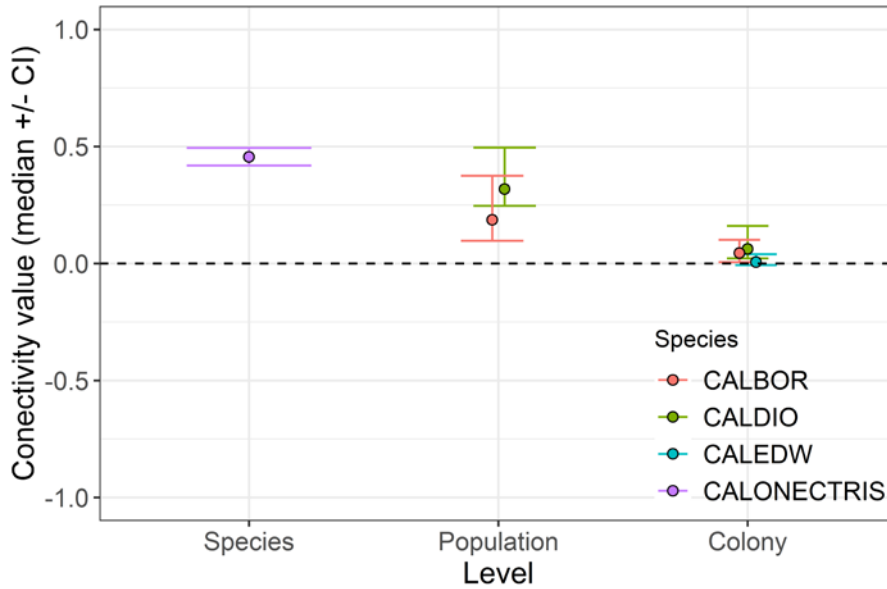


Figure A7

Migratory connectivity (MC) values at a species level, at a population level for Cory's and Scopoli's shearwater and at a colony level for Cory's, Scopoli's and Cape Verde shearwater. Closed circles represent median value, and whiskers represent 95% confidence interval).

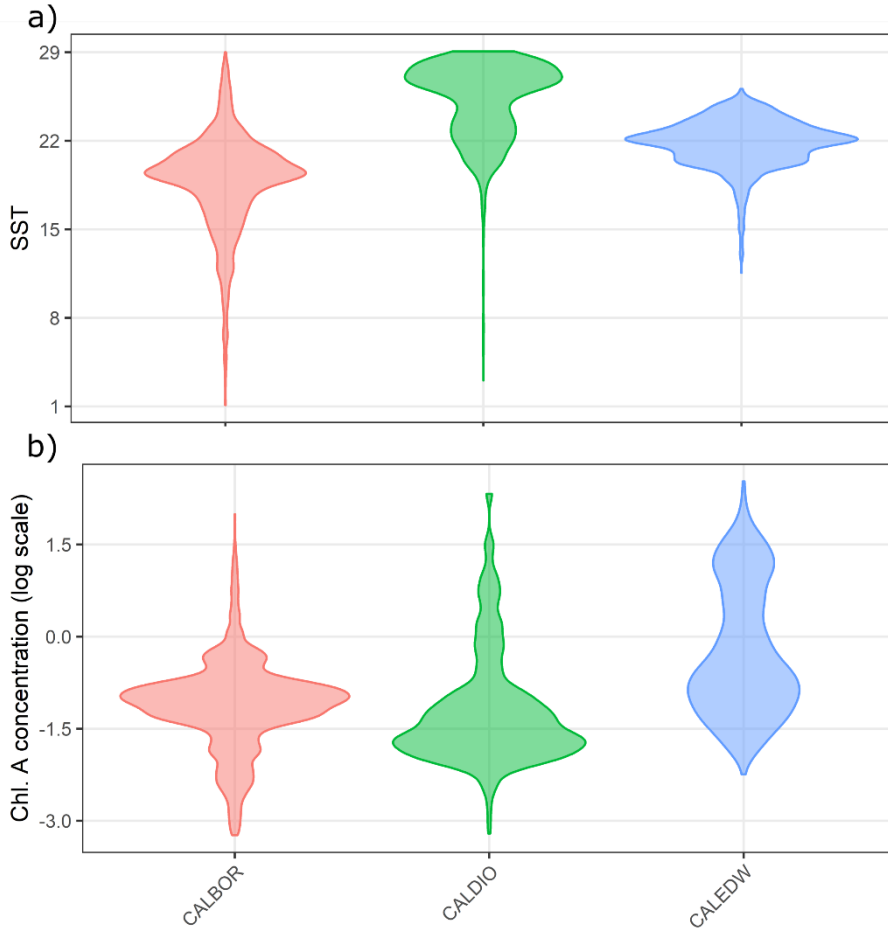
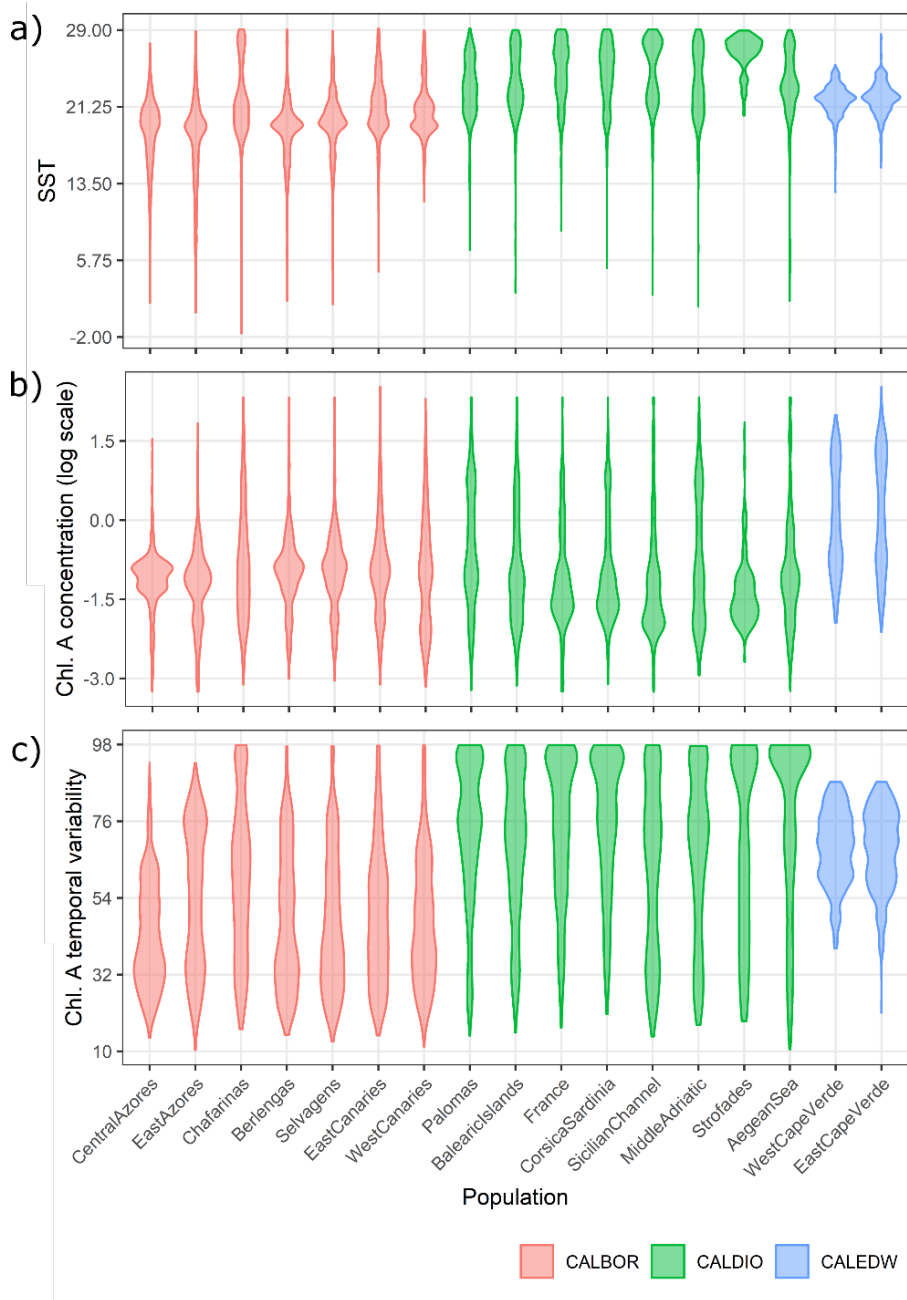


Figure A8

Violin plots representing the values selected in the three ENM at a taxon level for the three most important variables: (a) SST, (b) CHLA. The points represent the value of the environmental variable at each of the presences, and the width of the violin represents the density of points.

Figure A9 (next page)

Violin plots representing the values selected for all the variables reported as most important in at least one of the population-level ENM: (a) SST, (b) CHLA and (c) CHLA_var. The points represent the value of the environmental variable at each of the presences, and the width of the violin represents the density of points.



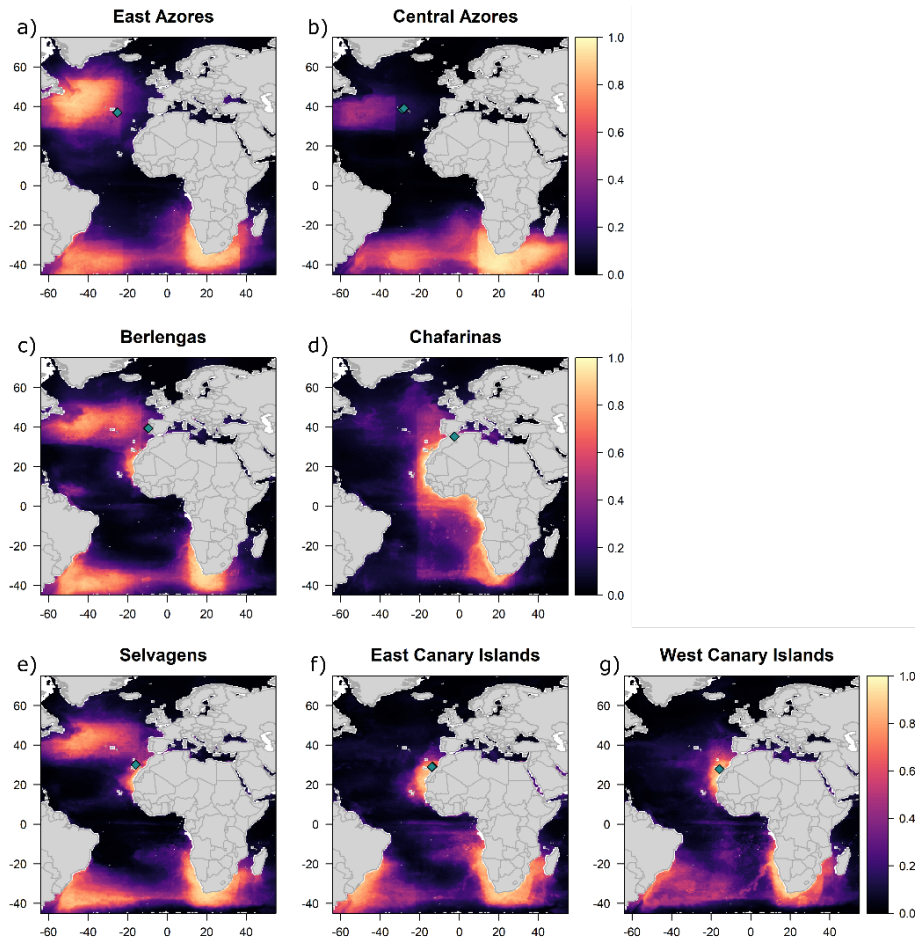
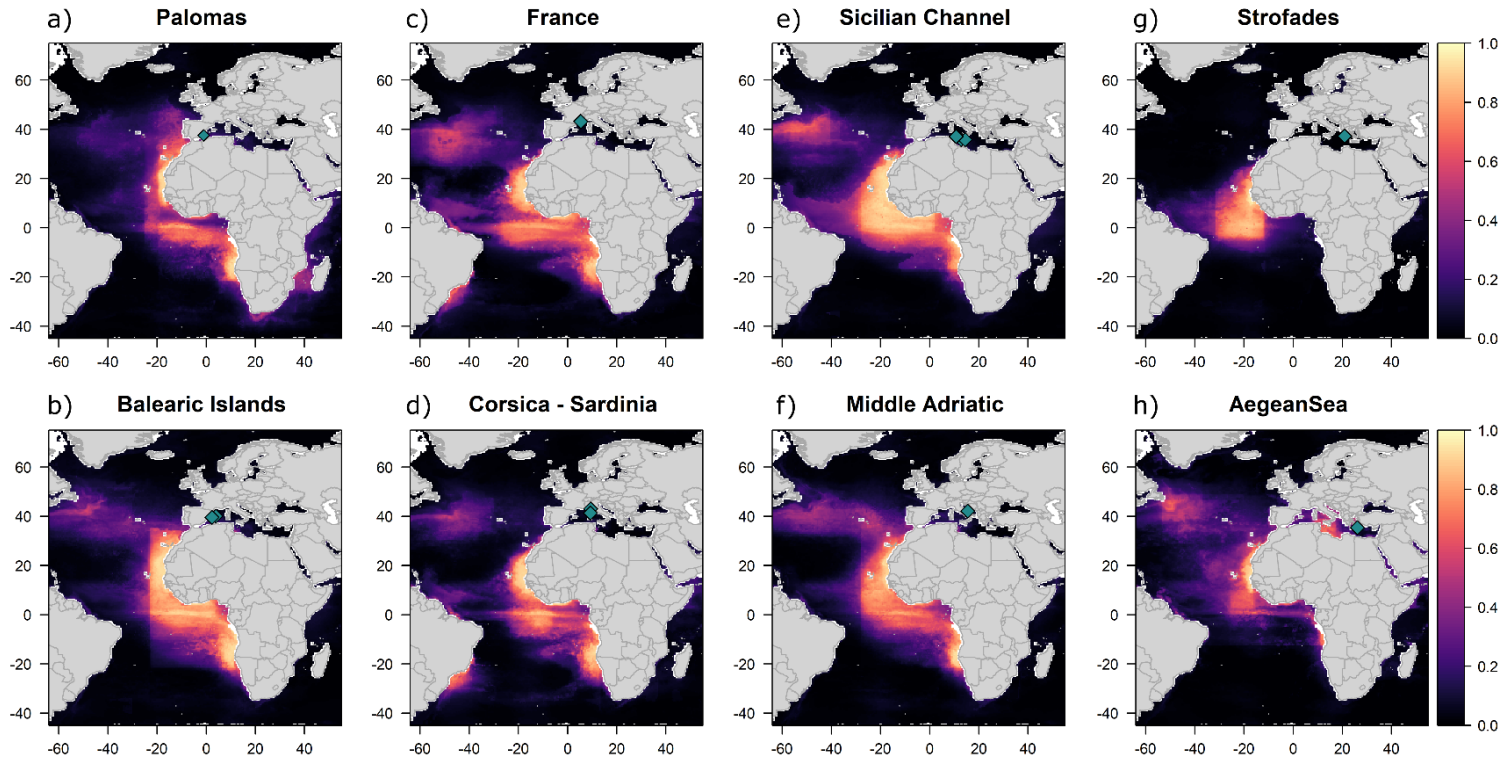


Figure A10

Spatial projection of the habitat suitability obtained from the ENM for the CALBOR populations: East Azores (a), Central Azores (b), Berlengas (c), Chafarinas (d), Selvagens (e), East Canary Islands (f) and West Canary Islands (g). Suitability values range from 0 (not suitable habitat) to 1 (most suitable habitat).

Figure A11 (next page)

Spatial projections for the CALDIO populations: Palomas (a), Balearic Islands (b), France (c), Corsica - Sardinia (d), Sicilian Channel (e), Middle Adriatic Sea (f) Strofades (g), and the Aegean Sea (h). Suitability values range from 0 (not suitable habitat) to 1 (most suitable habitat). Pink symbols represent the locations of the breeding colonies represented in the



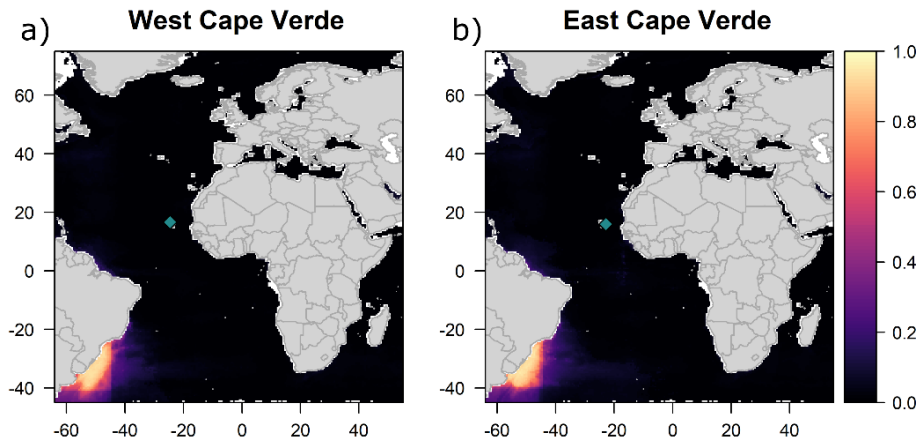


Figure A12

Spatial projection of the habitat suitability obtained from the ENM for the CALEDW populations: West Cape Verde (a), East Cape Verde (b).

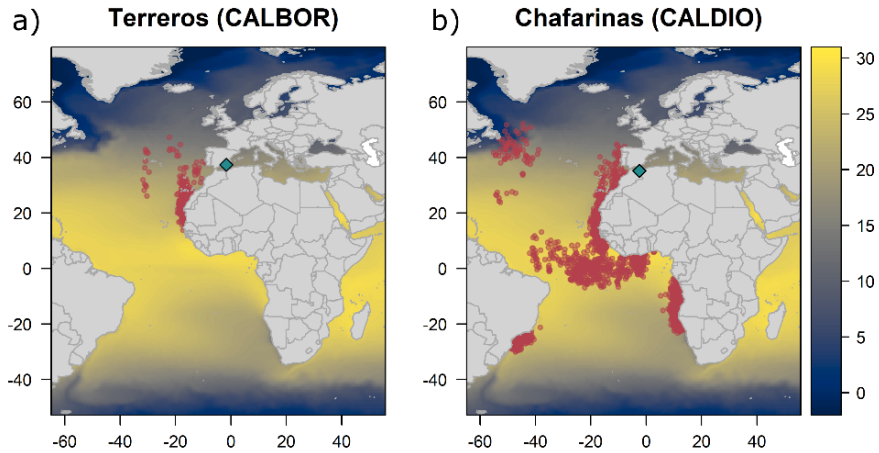


Figure A13

Non-breeding locations (green circles) of birds tracked from the Cory's shearwater colony of Terreros (a) and the Scopoli's shearwater colony of Chafarinas (b). Pink diamonds show the locations of the colonies. The background colour shows SST ($^{\circ}\text{C}$).

Figure A14 (next page)

Average values of the July – September monthly means for the covariates that had significant effect in at least one of the three models: Chlorophyll A concentration, sea surface temperature, slope and sea surface temperature gradient. The open triangle indicates the location of the Concepcion bank, and the open circles the location of the colonies: Montaña Clara (purple), Timanfaya (green), and Veneguera (yellow). Also indicated the positions of Cape Juby (a) and the Dakhla peninsula (b)

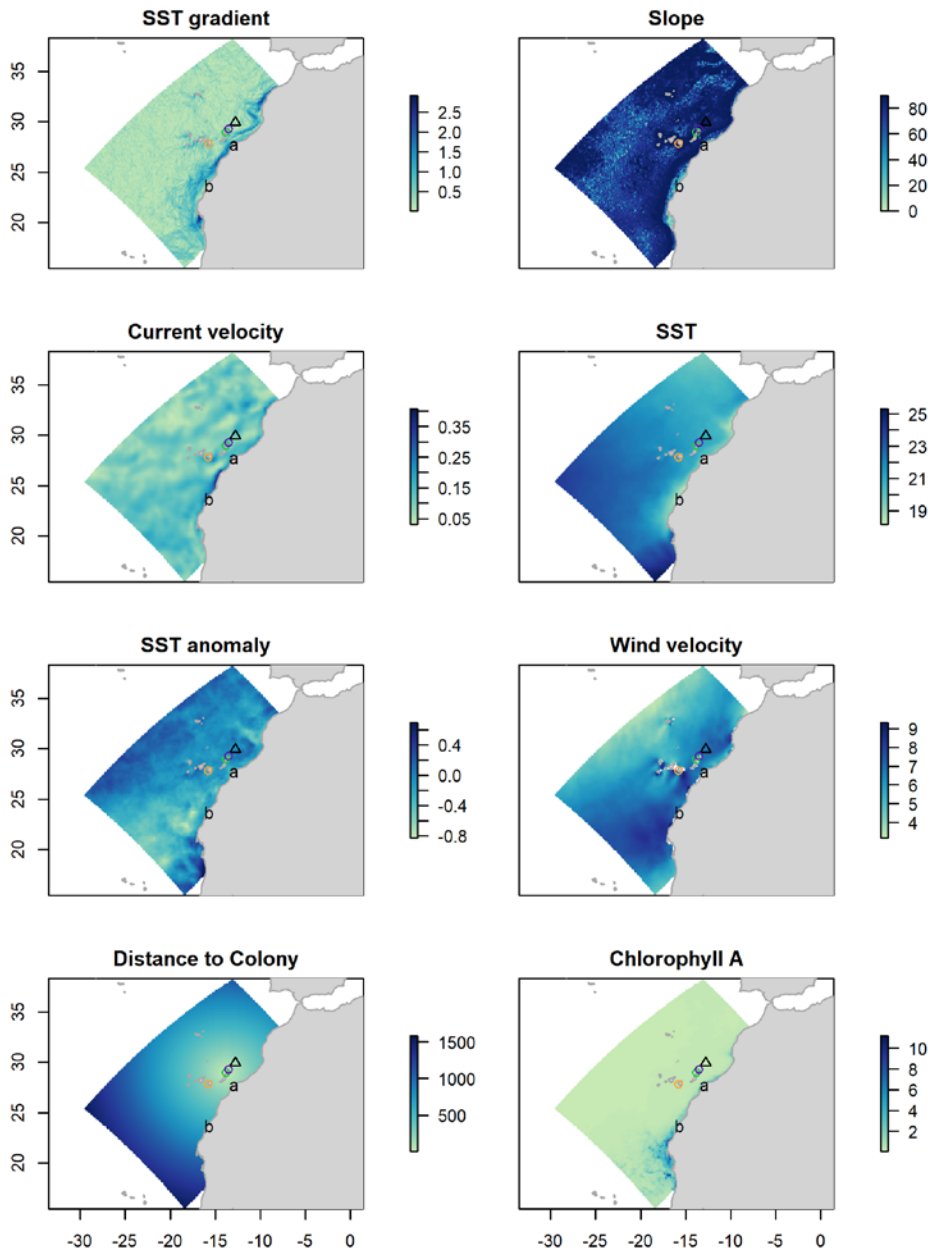
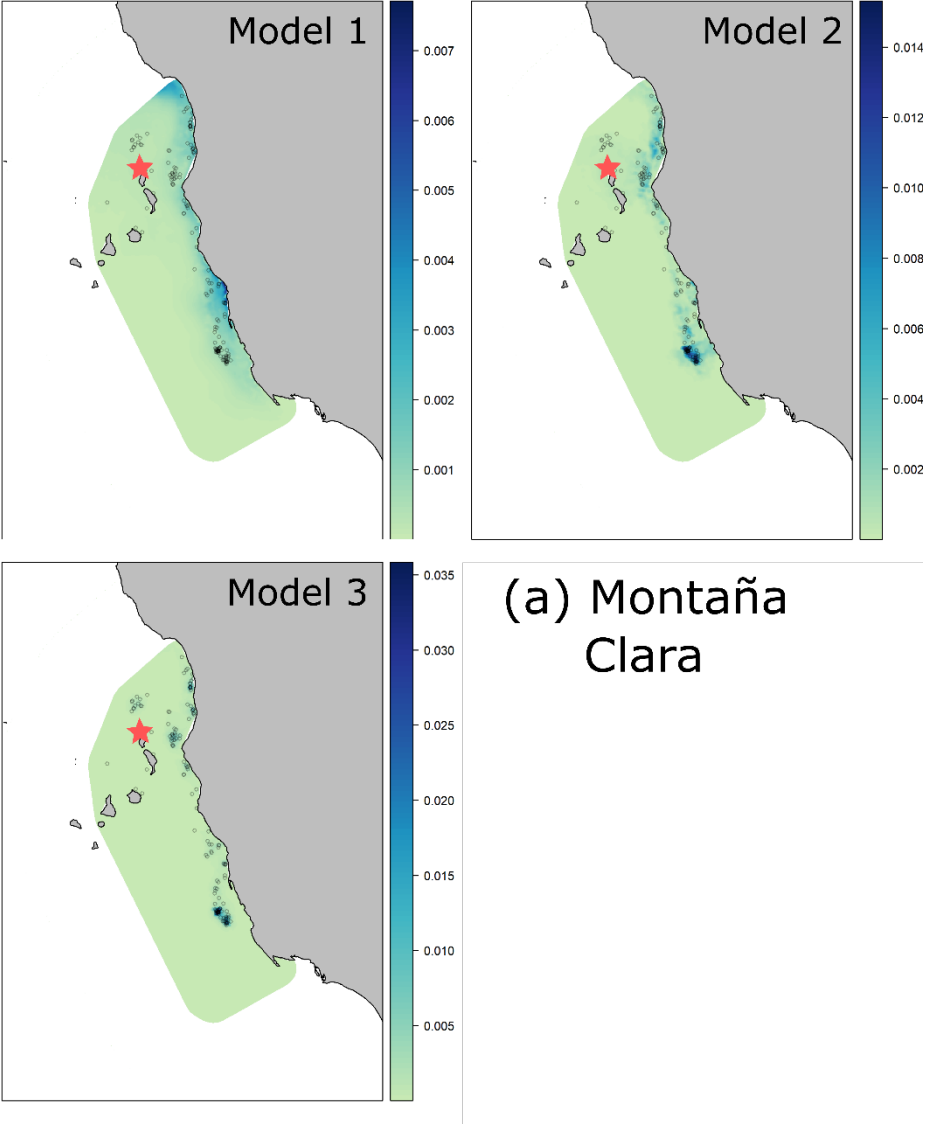
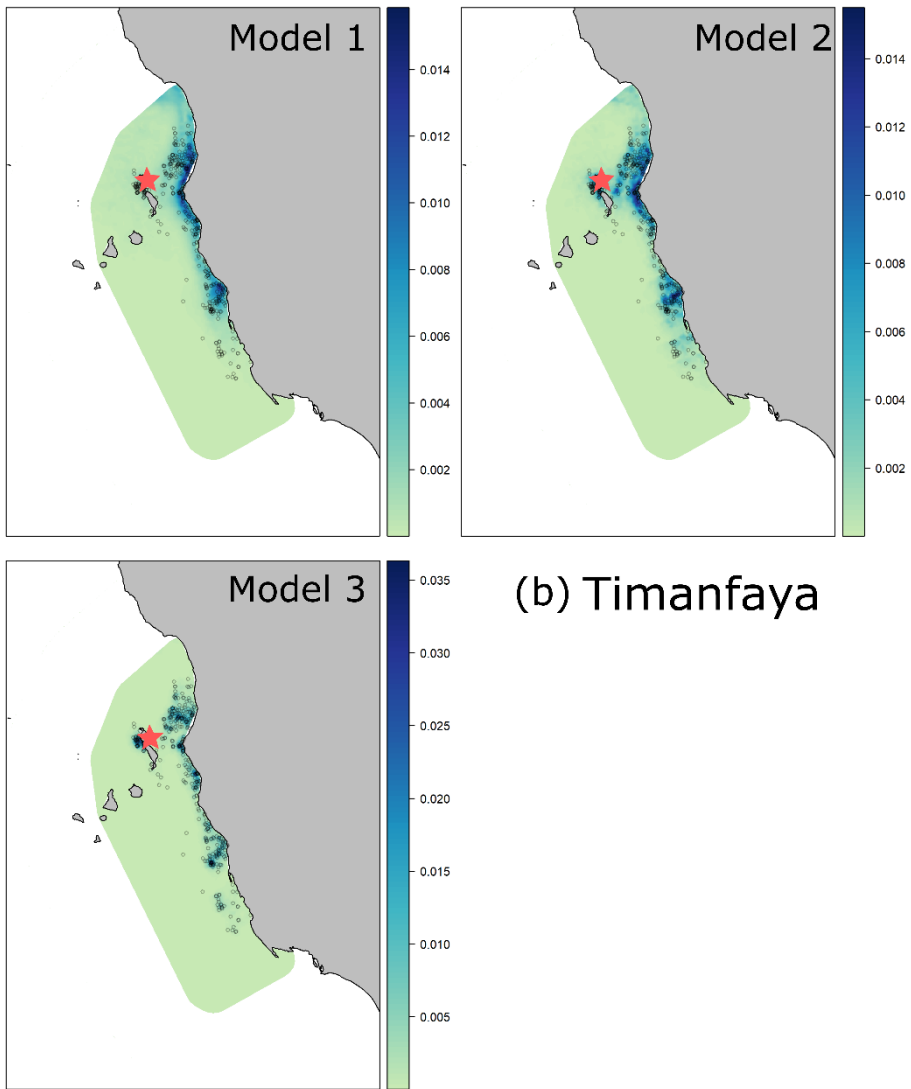


Figure A15

Intensity predicted by the linear (Model 1), non-linear (Model 2), and spatial (Model 3), models for the Montaña Clara (a), Timanfaya (b) and Veneguera (c) colonies.





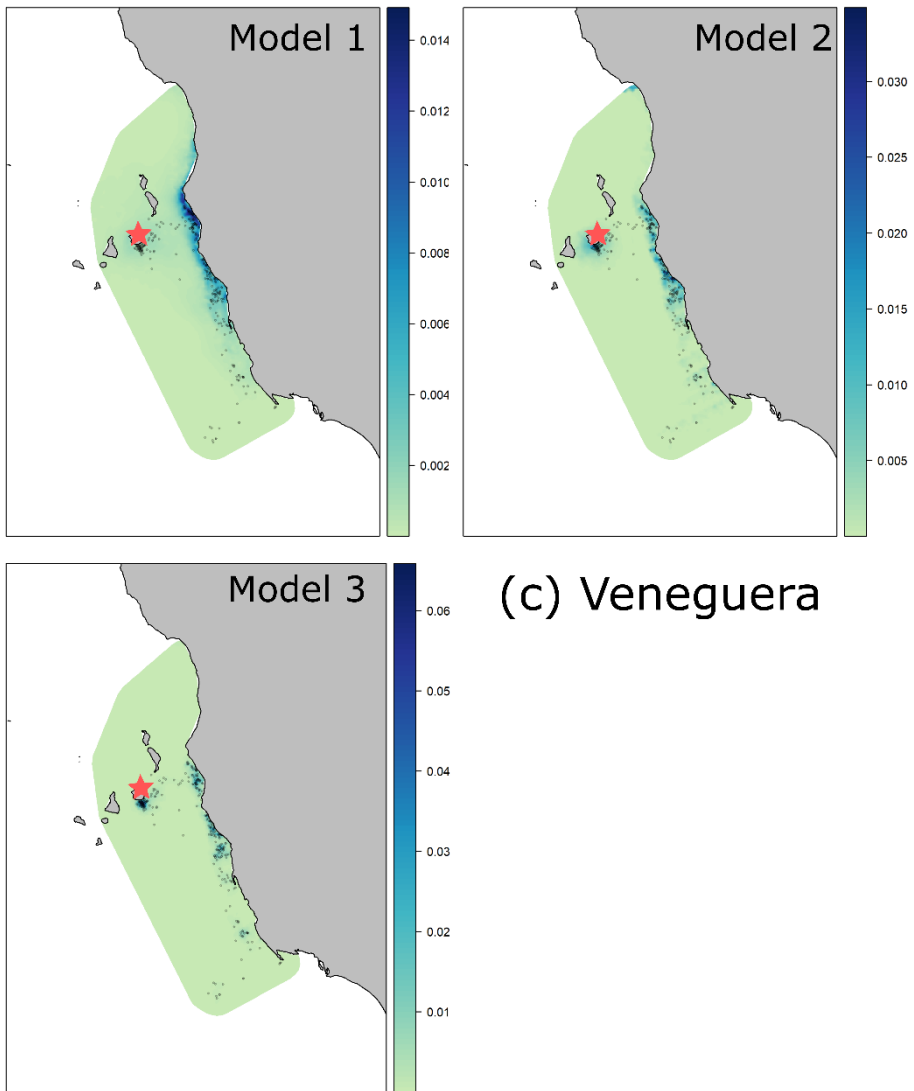
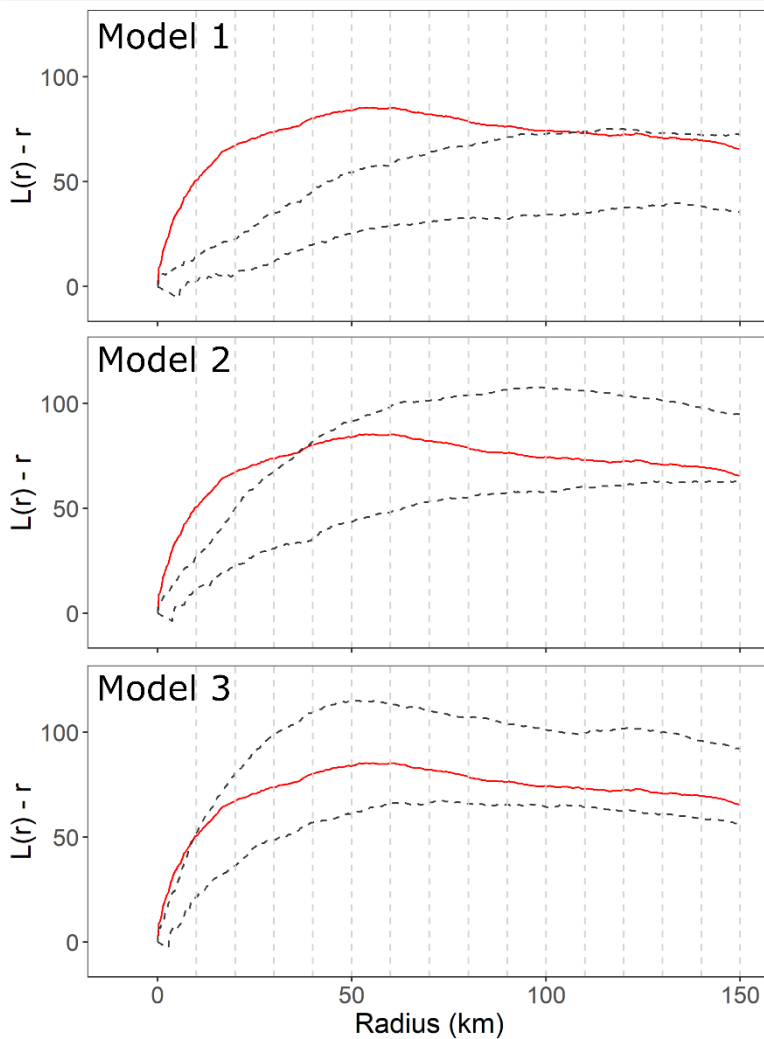


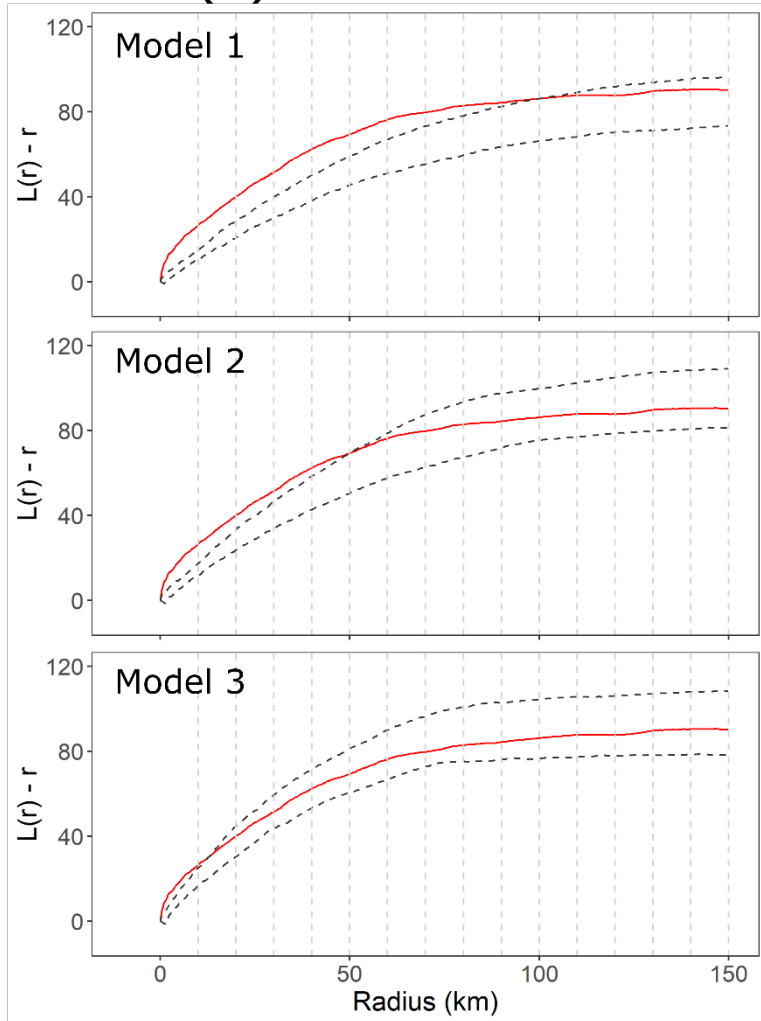
Figure A16

L-functions and envelopes of linear (Model 1) non-linear (Model 2) and spatial (Model 3) models for Montaña Clara (a), Timanfaya (b) and Veneguera (c). The observed point pattern is represented in a continuous red line, and the envelopes (obtained from 1000 simulations from the fitted model) are represented by the dashed black lines. When the L-function for the observed point pattern is above the envelope, there is clustering in the observed point pattern not captured by the fitted model. When the L-function of the observed point pattern is below the envelope there is segregation in the observed point pattern that is not captured by the fitted model.

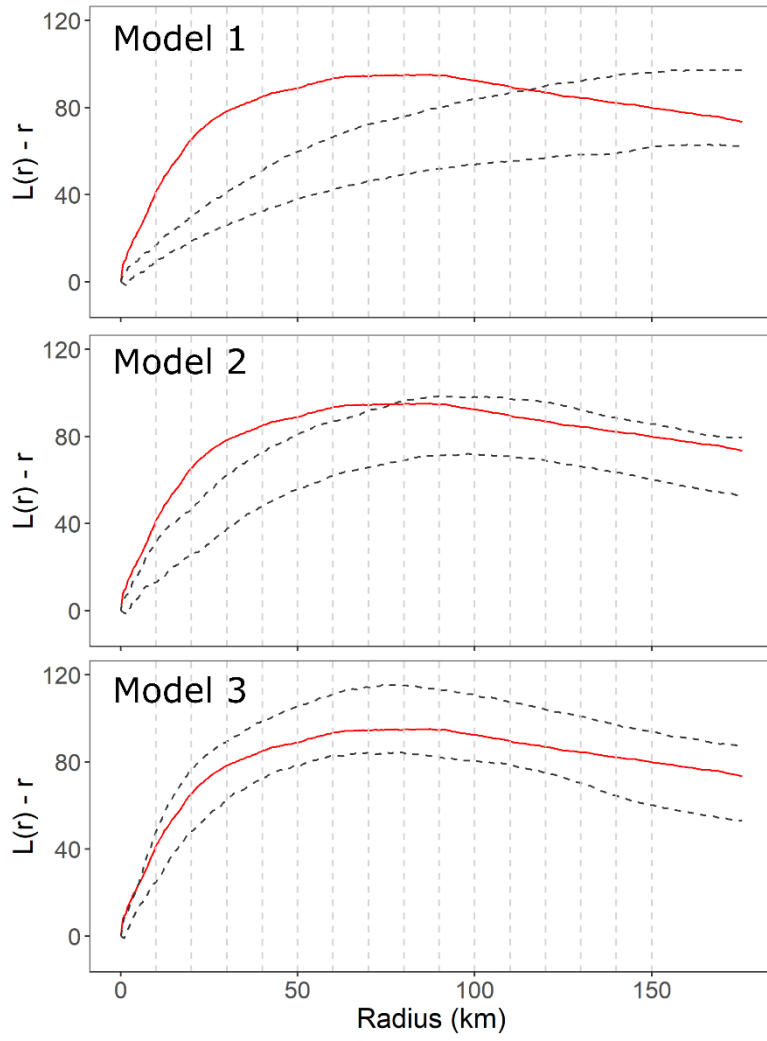
(a) MONTAÑA CLARA



(b) TIMANFAYA



(c) VENEGUERA



References

- Afán I, Navarro J, Cardador L, Ramírez F, Kato A, Rodríguez B, Ropert-Coudert Y, Forero MG. 2014. Foraging movements and habitat niche of two closely related seabirds breeding in sympatry. *Mar Biol.* 161(3):657–668. doi:10.1007/s00227-013-2368-4.
- Anselme L, Durand J. 2012. The Cory's Shearwater *Calonectris diomedea diomedea*: Updated state of knowledge and conservation of the nesting populations of the small Mediterranean islands.
- Arcos JM, Bécares J, Rodríguez B, Ruiz A. 2009. Áreas importantes para la conservación de las aves marinas en España. Madrid: LIFE04NAT/ES/000049 - Sociedad Española de Ornitología (SEO/BirdLife).
- Baccetti N, Capizzi D, Corbi F, Massa B, Nissardi S, Spano G, Sposimo P. 2009. Breeding shearwaters on Italian Islands: population size, selection and co-existence with their main alien predator, the black rat. *Riv Ital di Ornitol.* 78(2):83–100.
- Catry P, Dias MP, Phillips RA, Granadeiro JP. 2011. Different means to the same end: long-distance migrant seabirds from two colonies differ in behaviour, despite common wintering grounds. *PLoS One.* 6(10):4–9. doi:10.1371/journal.pone.0026079.
- Defos du Rau P, Bourgeois K, Thévenet M, Ruffino L, Dromzée S, Ouni R, Abiadh A, Estève R, Durand J-P, Anselme L, et al. 2015. Reassessment of the size of the Scopoli's Shearwater population at its main breeding site resulted in a tenfold increase: implications for the species conservation. *J Ornithol.* 156(4):877–892. doi:10.1007/s10336-015-1187-4. <http://link.springer.com/10.1007/s10336-015-1187-4>.
- Dias MP, Granadeiro JP, Catry P. 2012. Do seabirds differ from other

migrants in their travel arrangements? On route strategies of Cory's shearwater during its trans-equatorial journey. PLoS One. 7(11):e49376. doi:10.1371/journal.pone.0049376.

Dias MP, Granadeiro JP, Catry P. 2013. Individual variability in the migratory path and stopovers of a long-distance pelagic migrant. Anim Behav. 86(2):359–364. doi:10.1016/j.anbehav.2013.05.026. <http://dx.doi.org/10.1016/j.anbehav.2013.05.026>.

Fric J, Portolou D, Manolopoulos A, Kastritis T. 2012. Important Areas for Seabirds in Greece. NAT/GR/000285 L, editor. Athens: Hellenic Ornithological Society (HOS / BirdLife Greece).

Geraldes P. 2000. Censos de Procelariformes na Ilha de Madeira. Época de Nidificação 2000. Relatório final no âmbito do projecto “Novos Atlas das Aves Nidificantes em Portugal”.

González-Solís J, Croxall JP, Oro D, Ruiz X. 2007. Trans-equatorial migration and mixing in the wintering areas of a pelagic seabird. Front Ecol Environ. 5(6):297–301. doi:10.1890/1540-9295(2007)5[297:TMAMIT]2.0.CO;2. [accessed 2017 Jul 8]. [http://doi.wiley.com/10.1890/1540-9295\(2007\)5\[297:TMAMIT\]2.0.CO;2](http://doi.wiley.com/10.1890/1540-9295(2007)5[297:TMAMIT]2.0.CO;2).

González-Solís J, Felicísimo ÁM, Fox JW, Afanasyev V, Kolbeinsson Y, Muñoz J. 2009. Influence of sea surface winds on shearwater migration detours. Mar Ecol Prog Ser. 391:221–230. isi:000271218100019.

Granadeiro JP, Dias MP, Rebelo R, Santos CD, Catry P. 2006. Numbers and Population Trends of Cory's Shearwater *Calonectris diomedea* at Selvagem Grande, Northeast Atlantic. Waterbirds. 29(1):56–60. doi:10.1675/1524-4695(2006)29[56:NAPTOC]2.0.CO;2.

Hazevoet CJ. 1994. Status and conservation of seabirds in the Cape Verde

- Islands. BirdLife Conserv Ser. 1:279–293.
- Hazevoet CJ. 1995. The birds of the Cape Verde Islands. An annotated check list. London, UK: British-Ornitologist Union.
- Karris G. 2014. The breeding ecology of Scopoli's Shearwater (*Calonectris diomedea*) on Strofades Islands. University of Patras.
- Karris G, Xirouchakis S, Grivas K, Sfenthourakis S, Giokas S. 2017. Estimating the population size of Scopoli's Shearwaters (*Calonectris diomedea*) frequenting the Strofades islands (Ionian Sea, western Greece) by raft counts and surveys of breeding pairs. North West J Zool. 13(1):101–108.
- Lecoq M, Geraldès P, Andrade J. 2011. First complete census of Cory's Shearwaters *Calonectris diomedea borealis* breeding at Berlengas Islands (Portugal), including the small islets of the archipelago. Airo. 21(June):31–34.
- Missaglia R V., Ramos JA, Louzao M, Delord K, Weimerskirch H, Paiva VH. 2015. Year-round distribution suggests spatial segregation of Cory's shearwaters, based on individual experience. Araújo M, editor. Mar Biol. 162(11):2279–2289. doi:10.1007/s00227-015-2762-1. <http://link.springer.com/10.1007/s00227-015-2762-1>.
- Munilla I, Genovart M, Paiva VH, Velando A. 2016. Colony Foundation in an Oceanic Seabird. PLoS One. 11(2):e0147222. doi:10.13140/RG.2.1.2966.0248.
- Oliveira P, Menezes D. 2004. Aves do Arquipélago da Madeira.
- Paiva VH, Geraldès P, Rodrigues I, Melo T, Melo J, Ramos JA. 2015. The foraging ecology of the endangered Cape Verde shearwater, a sentinel species for marine conservation off West Africa. PLoS One.

10(10):e0139390. doi:10.1371/journal.pone.0139390.

<http://dx.plos.org/10.1371/journal.pone.0139390>.

Péron C, Grémillet D. 2013. Tracking through life stages: adult, immature and juvenile autumn migration in a long-lived seabird. *PLoS One*.

8(8):e72713. doi:10.1371/journal.pone.0072713.

Péron C, Grémillet D, Culioli J-M, Faggio G, Gillet P, Mante A, Vidal P.

2012. Exploring marine habitats of two shearwater species breeding on French Mediterranean islands. In: Yésou P, Baccetti N, Sultana J, editors. *Ecology and Conservation of Mediterranean Seabirds and other bird species under the Barcelona Convention - Proceedings of the 13th Medmaravis Pan- Mediterranean Symposium*. Proceedings of the 13th Medmaravis Pan-Mediterranean Symposium.

Ramirez I, Geraldés P, Meirinho A, Amorim P, Paiva VH. 2008. Áreas

Importantes para as Aves Marinhas em Portugal [Important Areas for Seabirds in Portugal]. Lisboa: Projeto LIFE04NAT/PT/000213 - Sociedade Portuguesa Para o Estudo das Aves.

Ramos R. 2019. Crossing the Pillars of Hercules: Understanding transoceanic migrations of seabirds throughout their breeding range. *Ecol Evol*.

9(8):4760–4771. doi:10.1002/ece3.5079.

<https://onlinelibrary.wiley.com/doi/abs/10.1002/ece3.5079>.

Ramos R, Granadeiro JP, Rodríguez B, Navarro J, Paiva VH, Bécares J, Reyes-González JM, Fagundes I, Ruiz A, Arcos JM, et al. 2013. Meta-population feeding grounds of Cory's shearwater in the subtropical Atlantic Ocean: implications for the definition of Marine Protected Areas based on tracking studies. *Divers Distrib*. 19(10):1284–1298.

Rodríguez B, De León L, Martín A, Alonso J, Nogales M. 2003. Status and distribution of breeding seabirds in the northern islets of Lanzarote

(Canary Islands). *Atl Seabirds*. 5(2):41–56.

Roscales JL, Gómez-Díaz E, Neves VC, González-Solís J. 2011. Trophic versus geographic structure in stable isotope signatures of pelagic seabirds breeding in the Northeast Atlantic. *Mar Ecol Prog Ser*. 434(2007):1–13. doi:10.3354/meps09211.

Samraoui B, Samraoui F. 2008. An ornithological survey of Algerian wetlands: Important Bird Areas, Ramsar sites and threatened species. *Wildfowl*. 58(2008):71–96.

Telailia S, Boutabia L, Bensouilah M-A, Houhamdi M. 2014. Breeding biology of Cory ' s Shearwater (*Calonectris diomedea*) in North East Algeria. *Int J Curr Sci*. 11:120–127.

https://www.researchgate.net/profile/S_Telailia/publication/282356583_Breeding_biology_of_Cory's_Shearwater_Calonectris_diomedea_in_North_East_Algeria/links/560e292408ae9674201114a8.pdf.

Annex 2

Multi-colony approaches to study migratory and foraging strategies in pelagic seabirds

Virginia Morera Pujol

September 2019

IndEffectTest

Written by Virginia Morera (2018), adapted from `varianceTest()` written by Phil Taylor and Mark Miller (2012)

Changes by Martin Beal (2019): - set default UDLev value to 50 - Remove default Scale parameter of 186km (user must input) - made it so the Matching Package doesn't print a message when the function is run (i.e. 'quietly') - Allow for input of SPDF (projected or un-projected) instead of just dataframe - added a snippet dealing with projecting data that crosses the dataline - Changed Grouping_var name to GroupVar - Added inGroupVar argument (i.e. within Grouping Variable, variable) this is the variable over which overlaps will be calculated - Small changes to allow for subsetting by inGroupVar, rather than "ID" column.

ver3 - add argument conditional from function `kerneloverlap` with default = T - change argument name from inGroupVar to tripID

ARGUMENTS TO THE FUNCTION

Tracks: must be a dataframe or SpatialPointsDataFrame with at least following fields: - Latitude - *Longitude*
- tripID: it can have any other name, the user will specify it in the tripID argument, but it must have a unique identifier for each trip - GroupVar: it can have any other name, and there can be as many as necessary, the user will specify it in the GroupVar argument. Variable to make the within-group vs. between group comparison (Year, Bird, etc.) * If it is a dataframe it must be unprojected (lonlat). If it's a SPDF it can be in any projection but it must be specified in the proj4string slot

method: character, one of the options from the `adehabitatHR::kerneloverlap` function to calculate overlap
conditional: logical. if TRUE the function sets to 0 the pixels of the grid over which the UD is estimated, outside the home range of the animal estimated at a level of probability equal to UDLev. Practically, if TRUE the maximum overlap will be equal to ~UDLev, if FALSE the maximum overlap will be equal to 1.
UDLev: numeric, value containing the % of the UD at which to calculate home-range (irrelevant if conditional is set to FALSE)
Scale is the smoothing factor to be used in the Kernel Density Estimation (in Km) grid is a number giving the size of the grid on which the UD should be estimated.

```
IndEffectTest <- function(Tracks, tripID, GroupVar, plotIt = F, #own arguments
                          method = c("HR", "PHR", "VI", "BA", "UDOI", "HD"),
                          conditional = TRUE, UDLev=50, Scale,
                          Grid = 500, #from adehabitat::kerneloverlap
                          nboots = 1000) #from Matching::ks.boot
{
  # Tracks <- ind_data_list[[1]]
  # packages
  require(sp)
  require(adehabitatHR)
  require(Matching, quietly = T)
  require(tidyverse)

  # initial chceks
  if (!"Latitude" %in% names(Tracks)) stop("Latitude field does not exist")
  if (!"Longitude" %in% names(Tracks)) stop("Longitude field does not exist")
}
```

```

if (!(tripID) %in% names(Tracks)) stop("Within-group field does not exist")
if (!GroupVar %in% names(Tracks)) stop("Group field does not exist")

# MB # Added this section which converts Tracks to spatial dataframe and projects it
# (and if already is SPDF it accepts this)
if (class(Tracks) != "SpatialPointsDataFrame") ## convert to SpatialPointsDF & project
{
  ## filter DF to the minimum fields that are needed
  CleanTracks <- Tracks %>%
    dplyr::select(GroupVar, tripID, Latitude, Longitude)
  mid_point <- data.frame(centroid(cbind(CleanTracks$Longitude, CleanTracks$Latitude)))

  ### PREVENT PROJECTION PROBLEMS FOR DATA SPANNING DATELINE
  if (min(CleanTracks$Longitude) < -170 & max(CleanTracks$Longitude) > 170) {
    longs = ifelse(CleanTracks$Longitude < 0, CleanTracks$Longitude + 360,
                  CleanTracks$Longitude)
    mid_point$lon <- ifelse(median(longs) > 180, median(longs) - 360, median(longs))}

  Tracks.Wgs <- SpatialPoints(data.frame(CleanTracks$Longitude, CleanTracks$Latitude),
                             proj4string = CRS("+proj=longlat + datum=wgs84"))
  proj.UTM <- CRS(paste("+proj=laea +lon_0=", mid_point$lon, " +lat_0=",
                       mid_point$lat, sep = ""))
  Tracks.Projected <- spTransform(Tracks.Wgs, CRS = proj.UTM )
  TracksSpatial <- SpatialPointsDataFrame(Tracks.Projected, data = CleanTracks)
  TracksSpatial@data <- TracksSpatial@data %>% dplyr::select(GroupVar, tripID,
                                                           Latitude, Longitude)

  Tracks.Wgs <- NULL
  Tracks.Projected <- NULL

}else {## if data are already in a SpatialPointsDataFrame then check for projection
  if (is.projected(Tracks)) {
    if ("trip_id" %in% names(Tracks@data)) {
      TracksSpatial <- Tracks }
    TracksSpatial@data <- TracksSpatial@data %>% dplyr::select(GroupVar, tripID,
                                                           Latitude, Longitude)
  }else {## project data to UTM if not projected
    mid_point <- data.frame(centroid(cbind(Tracks@data$Longitude, Tracks@data$Latitude)))

    ### MB This part prevents projection problems around the DATELINE
    if (min(Tracks@data$Longitude) < -170 & max(Tracks@data$Longitude) > 170) {
      longs = ifelse(Tracks@data$Longitude < 0, Tracks@data$Longitude + 360,
                    Tracks@data$Longitude)
      mid_point$lon <- ifelse(median(longs) > 180, median(longs) - 360, median(longs))}

    proj.UTM <- CRS(paste("+proj=laea +lon_0=", mid_point$lon, " +lat_0=", mid_point$lat,
                          sep = ""))
    TracksSpatial <- spTransform(Tracks, CRS = proj.UTM)
    TracksSpatial@data <- TracksSpatial@data %>% dplyr::select(GroupVar, tripID,
                                                           Latitude, Longitude)
  }
}
}

```

```

# remove tripID tracks with < 6 points as they can't be used to calculate kernel
# MB edit # Changed this step to happen after SPDF set-up. Also added tripID column.
UIDs <- names(which(table(TracksSpatial@data[, tripID]) > 5))
TracksSpatial <- TracksSpatial[TracksSpatial@data[, tripID] %in% UIDs, ]
TracksSpatial@data[, tripID] <- droplevels(as.factor(TracksSpatial@data[, tripID]))

# create vector with value of GroupVar for each trip
gid <- TracksSpatial@data[!duplicated(TracksSpatial@data[, tripID]), ][[GroupVar]]

# calculate overlap between tracks
X <- kerneloverlap(xy = TracksSpatial[, tripID], method = method, percent = UDLev,
  conditional = conditional, h = Scale*1000, grid = Grid)
X[lower.tri(X, diag = T)] <- NA

# assign value of GroupVar to rows and columns
rownames(X) <- colnames(X) <- gid

# separate within (WI) and between (BW) group overlaps
WI <- NULL
BW <- NULL
for (i in seq_along(rownames(X))) {
  # i = 1
  x1 <- X[i,]
  x2 <- x1[which(names(x1) == rownames(X)[i])]
  x3 <- x1[which(names(x1) != rownames(X)[i])]
  WI <- c(WI, x2)
  BW <- c(BW, x3)
}
BW <- BW[!is.na(BW)]
WI <- WI[!is.na(WI)]

## VMP commented this out since the ks.boot function is robust to ties.
## Was leftover from when using stats::ks.test function
# BW <- BW[BW != 0]
# WI <- WI[WI != 0]

# organize values in a dataframe for plotting
Overlaps <- data.frame(Overlap = c(WI, BW), Type = c(rep("Within", 1
  length(WI)), rep("Between",
  length(BW))))

# plot boxplot
if (plotIt) {
print(ggplot(data = Overlaps, aes(x = Type, y = Overlap, fill = Type)) +
  geom_boxplot(notch = F) + theme_classic())

# run KS test
ks <- Matching::ks.boot(WI, BW,
  alternative = "two.sided",
  nboots = nboots) # more indicated when data don't come
  # from continuous distr (ours have many 0s)

# Organise output

```



```

Result <- list()
Result[1] <- list(X) # overlaps matrix
Result[2] <- list(Overlaps) # df with overlap values (long format)
Result[3] <- list(ks) # output from the ks.boot function
names(Result) <- c("Overlap Matrix", "Overlaps", "Kolmogorov-Smirnov")
return(Result)
}
}

```

simulate distribution

Tracks must be a data frame or SPDF with at least Longitude and Latitude (not projected), Sp and Colony fields PopulationInfo must be a data frame with at least Colony, Sp and Pairs Scale is the smoothing factor to be used in the Kernel Density Estimation (in Km) Grid is a number giving the size of the grid on which the UD should be estimated. multi_factor is the number by which we want to multiply each population (i.e. simulate that number of positions from each animal). The higher the number, the larger the resulting point pattern

```

simulateDistribution <- function(Tracks, PopulationInfo, Scale, Grid = 500,
                               MultiFactor = 5){

  # Tracks <- colony_boot_list[[2]] #this can be used for testing if something goes wrong
  require(adehabitatHR)
  require(sp)
  require(spatstat, quietly = T)
  require(geosphere)

  if (!"Latitude" %in% names(Tracks)) stop("Latitude field does not exist")
  if (!"Longitude" %in% names(Tracks)) stop("Longitude field does not exist")
  if (!"Species" %in% names(Tracks)) stop("Species field does not exist")
  if (!"Population" %in% names(Tracks)) stop("Population field does not exist")

  # Convert Tracks to spatial dataframe and project them, or check projection if already
  # SPDF (and if already is SPDF it accepts this)
  if (class(Tracks) != "SpatialPointsDataFrame") ## convert to SpatialPointsDF & project
  {
    ## filter DF to the minimum fields that are needed
    CleanTracks <- Tracks %>%
      dplyr::select(Species, Population, Latitude, Longitude)
    mid_point <- data.frame(centroid(cbind(CleanTracks$Longitude, CleanTracks$Latitude)))

    ### PREVENT PROJECTION PROBLEMS FOR DATA SPANNING DATELINE
    if (min(CleanTracks$Longitude) < -170 & max(CleanTracks$Longitude) > 170) {
      longs = ifelse(CleanTracks$Longitude < 0, CleanTracks$Longitude + 360, CleanTracks$Longitude)
      mid_point$lon <- ifelse(median(longs) > 180, median(longs) - 360, median(longs))}

    Tracks.Wgs <- SpatialPoints(data.frame(CleanTracks$Longitude, CleanTracks$Latitude),
                                proj4string = CRS("+proj=longlat + datum=wgs84"))
    proj.UTM <- CRS(paste("+proj=laea +lon_0=", mid_point$lon, " +lat_0=",
                          mid_point$lat, sep = ""))
    Tracks.Projected <- spTransform(Tracks.Wgs, CRS = proj.UTM )
    TracksSpatial <- SpatialPointsDataFrame(Tracks.Projected, data = CleanTracks)
    TracksSpatial@data <- TracksSpatial@data %>% dplyr::select(Species, Population,
                                                                Latitude, Longitude)
  }
}

```

```

Tracks.Wgs <- NULL
Tracks.Projected <- NULL

}else {## if data are already in a SpatialPointsDataFrame then check for projection

if (is.na(Tracks@proj4string))
  stop("proj4string slot can't be NA. Assign the correct CRS object")

if (is.projected(Tracks)) {
  TracksSpatial@data <- TracksSpatial@data %>% dplyr::select(Species, Population,
                                                             Latitude, Longitude)
}else {## project data to UTM if not projected
  mid_point <- data.frame(centroid(cbind(Tracks@data$Longitude, Tracks@data$Latitude)))

  ### MB This part prevents projection problems around the DATELINE
  if (min(Tracks@data$Longitude) < -170 & max(Tracks@data$Longitude) > 170) {
    longs = ifelse(Tracks@data$Longitude < 0, Tracks@data$Longitude + 360,
                  Tracks@data$Longitude)
    mid_point$lon <- ifelse(median(longs) > 180, median(longs) - 360, median(longs))}

  proj.UTM <- CRS(paste("+proj=laea +lon_0=", mid_point$lon, " +lat_0=",
                       mid_point$lat, sep = ""))
  TracksSpatial <- spTransform(Tracks, CRS = proj.UTM)
  TracksSpatial@data <- TracksSpatial@data %>% dplyr::select(GroupVar, tripID,
                                                             Latitude, Longitude)
}
}

map <- rworldmap::getMap(resolution = "coarse")
map <- spTransform(map, TracksSpatial@proj4string)
# generate kernel
Kernel.est <- kernelUD(TracksSpatial, h = Scale*1000, grid = Grid)

# convert to pixel image
r <- raster(as(Kernel.est, "SpatialPixelsDataFrame"))
# raster.as.im function from Jeffrey Evans answer here: https://bit.ly/2TIOFXB
raster.as.im <- function(im) {
  r <- raster::res(im)
  orig <- sp::bbox(im)[, 1] + 0.5 * r
  dm <- dim(im)[2:1]
  xx <- unname(orig[1] + cumsum(c(0, rep(r[1], dm[1] - 1))))
  yy <- unname(orig[2] + cumsum(c(0, rep(r[2], dm[2] - 1))))
  return(spatstat::im(matrix(raster::values(im), ncol = dm[1],
                             nrow = dm[2], byrow = TRUE)[dm[2]:1, ],
                      xcol = xx, yrow = yy))
}
kernel.im <- raster.as.im(r)

# select colony size info
SPopulation <- as.character(unique(Tracks$Population))
SSpecies <- as.character(unique(Tracks$Species))
Pop.size <- PopulationInfo[PopulationInfo$Species == SSpecies &
                           PopulationInfo$Population == SPopulation,]$Pairs*2

```

```

# we're going to simulate a n° of points equal to the pop size * multi_factor
SimulateN <- Pop.size*MultiFactor

# this simulates the points as ppp
SimPoints <- rpoint(SimulateN, kernel.im)

# convert to dataframe, and from there to Spatial points
SimPoints.df <- as.data.frame(SimPoints)
SimPoints.sp <- SimPoints.df
coordinates(SimPoints.sp) <- ~ x+y

# plot to see everything has worked
par(mfrow = c(1,2))
plot(kernel.im, main = SPopulation, xlim = SimPoints.sp@bbox[1,],
      ylim = SimPoints.sp@bbox[2,])
plot(map, add = T, border = "white")
# plot(kernel.im, main = SPopulation)
plot(SimPoints.sp, pch = 20, col = "#ff000030", cex = 0.3)
plot(map, add = T, border = "black")

par(mfrow = c(1,1))

# prepare output
SimPoints.sp <- SpatialPointsDataFrame(coords = SimPoints.sp@coords,
                                     data = data.frame(Population = rep(SPopulation,
                                                                       length(SimPoints.sp)),
                                                         Species = rep(SSpecies,
                                                                       length(SimPoints.sp))),
                                     proj4string = proj.UTM)
SimPoints.sp <- spTransform(SimPoints.sp, CRS(projections$WGS84))
return(SimPoints.sp)
}

```

bootstrap colony effect

Tracks: Tracking data. DF or SPDF, containing Latitude, Longitude, Species and Population
 SimulatedDistributions: DF or SPDF containing the simulated distribution of the studied species, with at least the columns Longitude, Latitude, Population and Species. Can be directly the output of the `simulateDistribution()` function
 tripID: unique identifier for each trip
 UDLev: Numeric, Utilisation Distribution Level. Level of the KDE to be selected. For example, if 50 is used, the overlap is calculated between the smallest areas for which the probability to find the animal is equal to 0.50 (but see explanation for argument “conditional”). From the `adehabitatHR::kerneloverlap()` function. Scale is the smoothing factor to be used in the Kernel Density Estimation (in Km)
 Grid is a number giving the size of the grid on which the UD should be estimated.
 Iterations: Numeric, number of iterations for the bootstrap

```

bootstrapColony <- function(Tracks, SimulatedDistributions, tripID, UDLev = 50, Scale,
                           Grid = 500, Iterations = 50){
  require(sp)
  require(geosphere)
  require(rgdal)
  require(adehabitatHR)
  require(foreach)
  require(doParallel)

```

```

require(parallel)

Tracks$tripID <- Tracks[,tripID]
if (!"Latitude" %in% names(Tracks)) stop("Latitude field does not exist")
if (!"Longitude" %in% names(Tracks)) stop("Longitude field does not exist")
if (!"tripID" %in% names(Tracks)) stop("tripID field does not exist")
if (!"Population" %in% names(Tracks)) stop("Population field does not exist")
if (!"Species" %in% names(Tracks)) stop("Sp field does not exist")

# select species corresponding to the Tracking data
SimulatedDistributions <- SimulatedDistributions[SimulatedDistributions$Species == unique(Tracks$Species)]

# Converts SimulatedDistributions to spatial dataframe and projects it
# (and if already is SPDF check projection)
if (class(SimulatedDistributions) != "SpatialPointsDataFrame") ## convert to SpatialPointsDataFrame
{
  ## filter DF to the minimum fields that are needed
  CleanSimulatedDistributions <- SimulatedDistributions %>%
    dplyr::select(Species, Population, Longitude, Latitude)
  mid_point <- data.frame(centroid(cbind(CleanSimulatedDistributions$Longitude, CleanSimulatedDistributions$Latitude)))

  ### PREVENT PROJECTION PROBLEMS FOR DATA SPANNING DATELINE
  if (min(CleanSimulatedDistributions$Longitude) < -170 &
      max(CleanSimulatedDistributions$Longitude) > 170) {
    longs = ifelse(CleanSimulatedDistributions$Longitude < 0,
                  CleanSimulatedDistributions$Longitude + 360,
                  CleanSimulatedDistributions$Longitude)
    mid_point$lon <- ifelse(median(longs) > 180, median(longs) - 360, median(longs))}

  SimulatedDistributions.Wgs <- SpatialPoints(data.frame(
    CleanSimulatedDistributions$Longitude,
    CleanSimulatedDistributions$Latitude),
    proj4string = CRS("+proj=longlat + datum=wgs84"))
  proj.Sim <- CRS(paste("+proj=laea +lon_0=", mid_point$lon, " +lat_0=",
                      mid_point$lat, sep = ""))
  SimulatedDistributions.Projected <- spTransform(SimulatedDistributions.Wgs,
                                                  CRS = proj.Sim)
  SimulatedDistributionsSpatial <- SpatialPointsDataFrame(SimulatedDistributions.Projected,
                                                         data = CleanSimulatedDistributions)
  SimulatedDistributionsSpatial@data <- SimulatedDistributionsSpatial@data %>%
    dplyr::select(Species, Population)
  SimulatedDistributions.Wgs <- NULL
  SimulatedDistributions.Projected <- NULL
} else {## if data are already in a SpatialPointsDataFrame then check for projection
  if (is.na(SimulatedDistributions@proj4string))
    stop("The proj4string slot of SimulatedDistributions can't be empty.
         Check and assign projection and re-run")
  if (!is.projected(SimulatedDistributions)) { # if it's not projected
    # (lonlat) project to laea around midpoint
    mid_point <- data.frame(centroid(cbind(SimulatedDistributions@coords[,1], SimulatedDistributions@coords[,2])))
    proj.Sim <- CRS(paste("+proj=laea +lon_0=", mid_point$lon, " +lat_0=",
                        mid_point$lat, sep = ""))
  }
}

```

```

### MB This part prevents projection problems around the DATELINE
if (min(SimulatedDistributions@coords[,1]) < -170 &
    max(SimulatedDistributions@coords[,1]) > 170) {
  longs = ifelse(SimulatedDistributions@coords[,1] < 0,
                 SimulatedDistributions@coords[,1] + 360,
                 SimulatedDistributions@coords[,1])
  mid_point$lon <- ifelse(median(longs) > 180, median(longs) - 360,
                         median(longs))}
SimulatedDistributionsSpatial <- spTransform(SimulatedDistributions, proj.Sim)
SimulatedDistributionsSpatial@data <- SimulatedDistributionsSpatial@data %>%
  dplyr::select(Species, Population)
} else { ## if projected, "unproject" to WGS to project laea around midpoint
  SimulatedDistributions.WGS <- spTransform(SimulatedDistributions,
                                           CRS("+proj=longlat + datum=wgs84"))
  mid_point <- data.frame(centroid(
    cbind(SimulatedDistributions.WGS@coords[,1],
          SimulatedDistributions.WGS@coords[,2])))

  ### MB This part prevents projection problems around the DATELINE
  if (min(SimulatedDistributions.WGS@coords[,1]) < -170 &
      max(SimulatedDistributions.WGS@coords[,1]) > 170) {
    longs = ifelse(SimulatedDistributions@coords[,1] < 0,
                   SimulatedDistributions@coords[,1] + 360,
                   SimulatedDistributions@coords[,1])
    mid_point$lon <- ifelse(median(longs) > 180, median(longs) - 360, median(longs))}

  proj.Sim <- CRS(paste("+proj=laea +lon_0=", mid_point$lon, " +lat_0=",
                       mid_point$lat, sep = ""))
  SimulatedDistributionsSpatial <- spTransform(SimulatedDistributions.WGS,
                                              CRS = proj.Sim)
  SimulatedDistributionsSpatial@data <- SimulatedDistributionsSpatial@data %>%
    dplyr::select(Species, Population)
}
}

# Converts Tracks to spatial dataframe and projects it
# (and if already is SPDF check projection)
if (class(Tracks) != "SpatialPointsDataFrame") # convert to SpatialPointsDF & project
{
  ## filter DF to the minimum fields that are needed
  CleanTracks <- Tracks %>%
    dplyr::select(Species, Population, tripID, Latitude, Longitude)

  Tracks.Wgs <- SpatialPoints(data.frame(CleanTracks$Longitude, CleanTracks$Latitude),
                             proj4string = CRS("+proj=longlat + datum=wgs84"))
  Tracks.Projected <- spTransform(Tracks.Wgs,
                                 CRS = SimulatedDistributionsSpatial@proj4string)
  TracksSpatial <- SpatialPointsDataFrame(Tracks.Projected, data = CleanTracks)
  TracksSpatial@data <- TracksSpatial@data %>% dplyr::select(Species, Population, tripID)
  Tracks.Wgs <- NULL
  Tracks.Projected <- NULL
} else {## if data are already in a SpatialPointsDataFrame then check for projection

```

```

if (is.na(Tracks@proj4string)) stop("The proj4string slot of Tracks can't be empty.
                                   Check and assign projection and re-run")
TracksSpatial <- spTransform(Tracks, CRS = SimulatedDistributionsSpatial@proj4string)
TracksSpatial@data <- TracksSpatial@data %>% dplyr::select(Species, Population, tripID)
}

# remove tripID tracks with < 6 points as they can't be used to calculate kernel
UIDs <- names(which(table(TracksSpatial@data[, tripID]) > 6))
TracksSpatial <- TracksSpatial[TracksSpatial@data[, tripID] %in% UIDs, ]
TracksSpatial@data[, tripID] <- droplevels(as.factor(TracksSpatial@data[, tripID]))

TracksSpatial$X <- TracksSpatial@coords[,1]
TracksSpatial$Y <- TracksSpatial@coords[,2]

UIDs <- as.character(unique(TracksSpatial$tripID))
Ntrips <- length(UIDs)
Nloop <- seq(1, (Ntrips - 1), ifelse(Ntrips > 100, 10, 1))
DoubleLoop <- data.frame(SampleSize = rep(Nloop, each = Iterations), Iteration = rep(seq(1:Iterations)
LoopNr <- seq(1:dim(DoubleLoop)[1])
UDLev <- UDLev

#setup parallel backend to use 4 processors
cl <- makeCluster(detectCores(), outfile = "")
registerDoParallel(cl)
Result <- data.frame()

Result <- foreach(LoopN = LoopNr, .combine = rbind, .packages = c("sp", "adehabitatHR", "geosphere", "rgl"),
  # output_list <- list()
  # for (j in seq_along(LoopNr)) {
  # j = 20
  # LoopN <- LoopNr[j]
  N <- DoubleLoop$SampleSize[LoopN]
  i <- DoubleLoop$Iteration[LoopN]
  # Coverage <- NULL
  # Inclusion <- NULL
  # History <- NULL

  Output <- data.frame(SampleSize = N, InclusionMean = 0, Iteration = i)
  # set.seed(123)
  RanNum <- sample(UIDs, N, replace = F)
  sink("D:/selected_ids_colony_bootstrap.txt", append = TRUE)
  cat(RanNum, "\n", "\n")
  sink()
  SelectedCoords <- TracksSpatial[TracksSpatial@data[, tripID] %in% RanNum, ]
  # Ext <- (min(SelectedCoords@coords[,1]) + 3 * diff(range(SelectedCoords@coords[,1])))
  # if(Ext < (Scale * 1000 * 2)) {
  # BExt <- ceiling((Scale * 1000 * 3)/(diff(range(SelectedCoords@coords[,1])))
  # #} else {BExt <- 3}
  KDE.Surface <- kernelUD(SelectedCoords, h = Scale*1000, grid = 500, same4all = FALSE)
  try(KDE.UD <- getverticeshr(KDE.Surface, percent = UDLev))
  if (isTRUE(class(KDE.UD) == "try-error")) {
    sink("errors.txt", append = T)

```

```

cat(paste("Failed in iteration", i, "with sample", RanNum, sep = " "))
sink()} else {
  KDE.UD@proj4string <- SimulatedDistributionsSpatial@proj4string
  Overlain <- over(SimulatedDistributionsSpatial, KDE.UD)$area
  Output$InclusionMean <-
    length(Overlain[!is.na(Overlain)])/nrow(SimulatedDistributionsSpatial@data)
  return(Output)
  # output_list[j] <- list(Output)
}
}

## stop the cluster
stopCluster(cl)
closeAllConnections()

par(mfrow = c(1,1), mai = c(1,1,1,1))
#Result <- Output[1:nrow(Output) - 1,]
Result$Population <- unique(TracksSpatial$Population)
M1 <- try(nls(Result$InclusionMean ~ (a*Result$SampleSize)/(1+b*Result$SampleSize),
  data = Result, start = list(a = 1,b = 0.1)), silent = TRUE)
if (class(M1) != "try-error") { ### run this only if nls was successful
  Result$pred <- predict(M1)
  P2 <- aggregate(pred ~ SampleSize, Result, FUN = mean)
  P2$sd <- aggregate(InclusionMean ~ SampleSize, Result, FUN = sd)[,2]
  plot(InclusionMean ~ SampleSize, data = Result,
    pch = 16, cex = 0.2, col = "darkgray", ylim = c(0, 1),
    ylab = "Inclusion", xlab = "Sample Size",
    main = paste(unique(TracksSpatial$Population), "UDLev", UDLev, sep = "_"))
  yTemp <- c((P2[,2] + P2[,3]), rev(P2[,2] - P2[,3]))
  xTemp <- c(P2[,1], rev(P2[,1]))
  polygon(x = xTemp, y = yTemp, col = "gray93", border = F)
  points(InclusionMean ~ SampleSize, data = Result, pch = 16, cex = 0.2, col = "darkgray")
  lines(P2, lty = 1,lwd = 2)
  Asymptote <- (summary(M1)$coefficients[1]/summary(M1)$coefficients[2])
  RepresentativeValue <- max(P2$pred)/Asymptote*100
  Result$RepresentativeValue <- RepresentativeValue
  print(RepresentativeValue)
  text(x = 2.5, y = 0.9,paste(round(RepresentativeValue,2),
    "%", sep = ""), cex = 2, col = "gray45", adj = 0)
} else{RepresentativeValue <- mean(Result$InclusionMean[
  Result$SampleSize == max(Result$SampleSize)])
### if nls is unsuccessful then use mean output for largest sample size
  Result$RepresentativeValue <- (RepresentativeValue/(UDLev/100))*100
  Result$pred <- NA}# added by Jono Handley to convert to same scale as nls output

Result$Asymptote <- Asymptote
write.table(Result, paste(unique(TracksSpatial$Population), "UDLev", UDLev, "bootout_temp.csv", sep =
return(Result)
}

```

Annex 3

Multi-colony approaches to study migratory and foraging strategies in pelagic seabirds

Virginia Morera Pujol

September 2019

With this script we generate the boundaries and mesh that we'll use for the INLA models. From our data (ds2) as SpatialPointsDataframe we obtain a `spatstats` `ppp` object to define a convex hull around them that will be our study area boundary

```
ds.ppp <- ppp(ds2@coords[,1], ds2@coords[,2], ds2@bbox[1,], ds2@bbox[2,],
             marks = ds2@data$marks)
w <- convexhull(ds.ppp)
Window(ds.ppp) <- w
boundary <- spoly(data.frame(w)) # turn window object into spatial polygon
```

We buffer 50 km around this polygon to keep points away from the edge

```
boundary@proj4string <- CRS(projections$LambertKM)
boundary <- rgeos::gBuffer(boundary, width = 50)
# plot(boundary)
# points(ds2, pch = 1, col = factor(ds2$Colony))
```

We now load a map of the world to intersect the boundary with the coastline and islands and exclude “land” from our boundary. We also simplify the lines, as the boundaries in `inla` can't be very complicated (can't have more resolution than the edge length of the triangles) or it will cause problems. Further, we save it as shapefile because we need to fix some problems (loops, inland “lakes”), in ArcGis as I don't know how to fix that in R, and then re-load it to continue to work with it

```
wrld <- rworldmap::getMap(resolution = "high")
wrld <- wrld[wrld$NAME %in% c("Spain", "Morocco", "W. Sahara", "Algeria", "Mauritania"),]
#just to reduce extent to the area we need
wrld2 <- spTransform(wrld, CRS(projections$LambertKM))
wrld2 <- gBuffer(wrld2, width = 0) # open the borders!
wrld3 <- gSimplify(wrld2, tol = 10)
# plot(boundary)
# plot(wrld3, add = T)
b_in <- gDifference(boundary, wrld3, byid = F, drop_lower_td = F)
# plot(b_in)
b_in2 <- rgeos::gBuffer(b_in, width = 5)
# plot(b_in2)
# points(ds2, pch = 20, col = factor(ds2$Colony))
bound <- SpatialPolygonsDataFrame(b_in2, data.frame(ID = "buffer"), match.ID = F)
# save(bound, file = "data/new_bound_chick.Rdata")
# load("data/new_bound_chick.Rdata")
```

We now need to generate an outer boundary, that will be used to avoid boundary effects. For that, we generate a 200 km buffer from the boundary we already have. Finally we plot everything to check that nothing has gone terribly wrong, and assign projection to everything (just in case, although we'll remove them for the modelling)

```
boundary2 <- rgeos::gBuffer(bound, width = 300) # 300 km between inner and outer boundaries
# plot(boundary2)
```



```

# plot(bound, add = T)
# plot(ds2, add = T, pch = ".", col = factor(ds2$Colony))

bound@proj4string <- CRS(projections$LambertKM)
boundary2@proj4string <- CRS(projections$LambertKM)
ds2@proj4string <- CRS(projections$LambertKM)

```

Now, finally, with these two boundaries we can generate the mesh. First we need to convert the two boundaries into “inla segments”. Then, we use the `inla.mesh.2d` function, pass to it the two boundaries, and values for the edges of the triangles (smaller for the inner boundary, larger for the outer boundary), and the cutoff (segments can’t be smaller than this value, this is why we’ve had to simplify the boundary earlier), and assign the same crs as the boundary.

We are going to generate a mesh for each colony, as the location of the points plays a part in defining the mesh triangles and the points are different for every colony

By colony

```

boundary_in <- inla.sp2segment(bound)
boundary_out <- inla.sp2segment(boundary2)
V_data <- ds2[ds2$Colony == "Veneguera",]
meshV <- inla.mesh.2d(loc = V_data, boundary = list(boundary_in, boundary_out),
                    max.edge = c(20, 50), cutoff = 10, crs = bound@proj4string)
plot(meshV)
# save(meshV, file = "data/definitive_mesh_Veneguera.Rdata")

M_data <- ds2[ds2$Colony == "MClara",]
meshM <- inla.mesh.2d(loc = M_data, boundary = list(boundary_in, boundary_out),
                    max.edge = c(20, 50), cutoff = 10, crs = bound@proj4string)
# plot(meshM)
# save(meshM, file = "data/definitive_mesh_MClara.Rdata")

T_data <- ds2[ds2$Colony == "Timanfaya",]
meshT <- inla.mesh.2d(loc = T_data, boundary = list(boundary_in, boundary_out),
                    max.edge = c(20, 50), cutoff = 10, crs = bound@proj4string)
# plot(meshT)
# save(meshT, file = "data/definitive_mesh_Timanfaya.Rdata")

```

We will now generate a “mask” that will hide the outer boundary for the plots. (The model will make predictions and fit covariates in the outer boundary as well, but we’re not interested in that, so we want to hide it for the plots)

This code is taken from Haakon Bakka’s barrier model from the website https://haakonbakka.bitbucket.io/btopic107.html#4_barrier_models

```

mesh <- meshT
t1 = length(mesh$graph$tv[,1]) # Number of triangles of the mesh
posTri = matrix(0, t1, 2) # matrix containing the coordinates of each triangle's

# we fill it with the vertex "coordinates here
for (t in 1:t1) {
  # Take the vertex of triangles
  temp = mesh$loc[mesh$graph$tv[t, ], ]
  # Compute barycenter which sintetize the triangle
  posTri[t,] = colMeans(temp)[c(1,2)]
}

```

```

#transform it in Spatial Points
posTri = SpatialPoints(posTri)
posTri@proj4string <- bound@proj4string

normal = over(bound, posTri, returnList = T) # this are the polygons contained
                                           # in the inner boundary, the "good" ones

normal = unlist(normal)
barrier.triangles = setdiff(1:tl, normal) # this are the polygons of the outer boundary,
                                           # the ones we want to hide

# build a polygon that contains all polygons that are outer boundary and not inner.
poly.barrier = inla.barrier.polygon(mesh, barrier.triangles)
# plot(poly.barrier)
# save(poly.barrier, file = "data/barrier.Rdata")

```

Annex 4

Multi-colony approaches to study migratory and foraging strategies in pelagic seabirds

Virginia Morera Pujol

September 2019

Sensitivity test for SPDE prior range and sigma

Load all necessary data

```
# load data
load("definitive_datasets/ALLth_chick.Rdata")
load("data/projections.Rdata")
load("data/new_bound_chick.Rdata")
load("data/definitive_mesh_Veneguera.Rdata")
load("data/definitive_mesh_Timanfaya.Rdata")
load("data/barrier.Rdata")

# make sure everything has the same CRS (i.e. none)
proj4string(bound) = inla.CRS()
proj4string(ds2) = inla.CRS()
proj4string(poly.barrier) = inla.CRS()
meshV$crs <- NULL
V_data <- ds2[ds2$Colony == "Veneguera",]
ds.ppp <- ppp(V_data@coords[,1], V_data@coords[,2],
             bound@bbox[1,], bound@bbox[2,]) # first generate the ppp from our data

wrld <- rworldmap::getMap(resolution = "high")
wrld <- spTransform(wrld, CRS(projections$LambertKM))
wrld <- rgeos::gBuffer(wrld, width = 0)
```

Prepare all elements of the model (better explained in supplementary material 4)

```
# prepare response
(nv <- meshV$n) # number of mesh nodes
(n <- nrow(V_data)) # number of data points

# calculate weights for each meshnode depending on whether they are inside or
# outside the boundary (outside weight is 0)
dmesh <- inla.mesh.dual(meshV) ##? Extracted from the SPDE manual, not sure what it does

sum(w <- sapply(1:length(dmesh), function(i) {
  if (gIntersects(dmesh[i,], bound))
    return(gArea(gIntersection(dmesh[i,], bound)))
  else return(0)
}))

table(w > 0) # false = nodes with weight 0 (outside the boundary)
           # true = nodes with some weight (inside the boundary)
```

```

length(y.pp <- rep(0:1, c(nv, n))) # response (0 for all meshnodes,
                                   # 1 for all data points)
summary(y.pp) # all 0 and 1

length(e.pp <- c(w, rep(0, n)))
summary(e.pp) # different values

# a matrix
lmat <- inla.spde.make.A(meshV, V_data@coords)
imat <- Diagonal(nv, rep(1, nv))
A.pp <- rbind(imat, lmat)

# prepare covariates
load("data/cov_list_Ve_normalised.Rdata")

covars <- do.call(cbind.data.frame, cov_list)

new_names <- list() # generate new names for the grouped covariates,
                   # with the original name and "LEV" (for "levels")
grouped_covars <- list()
for (i in seq_along(names(covars))) {
  # i = 1
  new_names[i] <- list(paste0(names(covars)[i], "LEV"))
  grouped_covars[i] <- list(inla.group(covars[,i],
                                       n = 20)) # 20 groups to begin with. If
                                                # it's too wiggly you can reduce the
                                                # number of groups (cheating,
                                                # but it works)
}

# add the grouped covars to the dataframe of original covars
new_covars <- do.call(cbind.data.frame, grouped_covars)
names(new_covars) <- unlist(new_names)
covars <- cbind(covars, new_covars)
covars$b0 <- 1 # add intercept

# formula
frml3V <- y ~ -1 + b0 +
  log_Ch1A +
  # log_current_strength +
  log_SST_Gradient +
  slope +
  SST +
  SST_Anomaly +
  Veneguera_dist +
  f(log_Ch1ALEV, model = 'rw2', scale.model = T,
    hyper = list(theta = list(prior = "pc.prec",
                              param = c(sd(covars$log_Ch1ALEV)*2.5, 0.01)))) +
  # f(log_current_strengthLEV, model = 'rw2', scale.model = T,
  # hyper = list(theta = list(prior = "pc.prec",
  #                             param = c(sd(covars$log_current_strengthLEV)*10, 0.01)))) +
  #

```

```

f(log_SST_GradientLEV, model = 'rw2', scale.model = T,
  hyper = list(theta = list(prior = "pc.prec",
    param = c(sd(covars$log_SST_GradientLEV)*50, 0.01)))) +
# f(slopeLEV, model = 'rw2', scale.model = T,
#   hyper = list(theta = list(prior = "pc.prec",
#     param = c(sd(covars$slopeLEV)*50, 0.01)))) +
# f(SSTLEV, model = 'rw2', scale.model = T,
#   hyper = list(theta = list(prior = "pc.prec",
#     param = c(sd(covars$SSTLEV)*10, 0.01)))) +
# f(SST_AnomalyLEV, model = 'rw2', scale.model = T,
#   hyper = list(theta = list(prior = "pc.prec",
#     param = c(sd(covars$SST_AnomalyLEV)*10, 0.01)))) +
f(Veneguera_distLEV, model = 'rw2', scale.model = T,
  hyper = list(theta = list(prior = "pc.prec",
    param = c(sd(covars$Veneguera_distLEV)*50, 0.01)))) +
f(sp_rdm, model = spde) +
# f(iid, model="iid", hyper=hyper.iid) +
NULL

```

Loop along sigma and range values

```

range_vector <- c(20, 50, 100, 150, 200, 250, 300)
sigma_vector <- c(0.05, 0.10, 0.15, 0.20, 0.50, 1, 2)

different_range_V_models <- list()
for (i in 1:7) {
  # i = 1
  range <- range_vector[i]
  different_sigma_V_models <- list()
  for (j in 1:7) {
    # j = 1
    sigma_hyp <- sigma_vector[j]
    print(paste("model", i*j, "of 49, range of", range, "and sigma of", sigma_hyp))
    # spatial field (start loop here)
    spde <- inla.spde2.pcmatern(mesh = meshV, alpha = 3/2, # mesh and smoothness parameter
      prior.range = c(range,
        0.01), # P(practic.range<50) = 0.01 = 1%
      prior.sigma = c(sigma_hyp,
        0.01)) # P(sigma > 0.2) = 0.01 = 1%

    field.indices <- inla.spde.make.index("sp_rdm", n.spde = spde$n.spde)

    # stack
    stack.spat <- inla.stack(data = list(y = y.pp, e = e.pp),
      A = list(A.pp, # for the spatial effect
        # 1, # for the rw and linear
        1), # for the iid model
      tag = 'dat',
      effects = list(field.indices,
        covars))

    mdl3V <- inla(frml3V,
      family = "poisson",
      data = inla.stack.data(stack.spat),

```

```

control.predictor = list(A = inla.stack.A(stack.spat),
                        compute = TRUE, link = 1),
E = inla.stack.data(stack.spat)$e,
control.compute = list(dic = T, waic = T, config = T),
# control.fixed = list(prec = 0.5, prec.intercept = 0.5,
#                       mean.intercept = 0),
control.inla = list(int.strategy = "eb"),
verbose = F)
different_sigma_V_models[j] <- list mdl3V)
names(different_sigma_V_models)[j] <- paste("range", range_vector[i],
                                           "sigma", sigma_vector[j], sep = "_")
}
different_range_V_models[i] <- list(different_sigma_V_models)
}
beepr::beep(sound = 8)

save(different_range_V_models, file = "sensitivity_tests_V.Rdata")

```

Plot posterior distributions of range and sigma

```

# plot ranges
# pdf("outputs/V_ranges.pdf", width = 3*10)
par(mfrow = c(7,7), mar = c(2, 2, 0, 0))
for (i in 1:7) {
  for (j in 1:7) {
    plot(different_range_V_models[[i]][[j]]$marginals.hyperpar$`Stdev for sp_rdm`,
         type = 'l', xlab = '', ylab = '') #, xlim = c(0, 5))
  }
}

# pdf("outputs/V_sigma.pdf", width = 3*10)
par(mfrow = c(7,7), mar = c(2, 2, 0, 0))
for (i in 1:7) {
  for (j in 1:7) {
    plot(different_range_V_models[[i]][[j]]$marginals.hyperpar$`Stdev for sp_rdm`,
         type = 'l', xlab = '', ylab = '', xlim = c(0, 5))
  }
}
}

```

Obtain posterior samples from all models

```

posterior_samples_1 <- list()
posterior_samples_2 <- list()
for (i in 1:7) {
  for (j in 1:7) {
    print(paste("i =", i, "and j =", j))
    mdl3M.posterior.sample <- inla.posterior.sample(n = 1000,
                                                  result = different_range_V_models[[i]][[j]],
                                                  use.improved.mean = T)
    posterior_samples_1[j] <- list(mdl3M.posterior.sample)
    names(posterior_samples_1)[j] <- paste("range", range_vector[i], "sigma",

```

```

        sigma_vector[j], sep = "_")
    }
    posterior_samples_2[i] <- list(posterior_samples_1)
}

```

Plotting specifications

```

(contents <- mdl3V$misc$configs$contents) # check the contents
effect <- "sp_rdm" # select the effect we want to sample
id.effect <- which(contents$tag == effect) # the numerical id of the effect
ind.effect <- contents$start[id.effect] - 1 +
             (1:contents$length[id.effect]) # all the indices for the effect

pltt <- RColorBrewer::brewer.pal(11, "RdYlBu")
sp_fld_pallette <- colorRampPalette(rev(pltt))

projV = inla.mesh.projector(meshV,
                            xlim = poly.barrier@bbox[1,], ylim = poly.barrier@bbox[2,],
                            dims = c(280*2, 245*2)) #this is the dimensions of the covars
                                                    #rasters/10. For definitive plots make
                                                    #with more resolution (*2)

```

Plot spatial fields of all models

```

spatial_rdm_1 <- list()
spatial_rdm_2 <- list()
for (i in 1:7) {
  for (j in 1:7) {
    # i = 1
    # j = 1
    sp_sample3V <- lapply(posterior_samples_2[[i]][[j]],
                        function(x) x$latent[ind.effect]) # extract this part of the
                                                            # posterior from all the
                                                            # 1000 samples
    spLog3V <- matrix(unlist(sp_sample3V), byrow = T, nrow = 1000) #prediction in
                                                                # the log scale
    spResponse3V <- exp(matrix(unlist(sp_sample3V), byrow = T,
                                nrow = 1000)) #prediction in the response scale

    # calculate mean and sd from previous matrix
    mean_spResponse3V <- (colMeans(spResponse3V))[1:nv]
    mean_spLog3V <- (colMeans(spLog3V))[1:nv]
    lim <- max(abs(range(mean_spLog3V)))
    lo_lim <- -lim
    hi_lim <- lim
    sp.mean3V <- my.levelplot(projV, mean_spLog3V, col.regions = sp_fld_pallette(100),
                             at = seq(lo_lim, hi_lim, length.out = 100),
                             xlab = "", ylab = "",
                             scales = list(draw = F), contour = F) +
    latticeExtra::layer(sp.polygons(wrld, lwd = 1, col = "black", fill = "gray")) +
    latticeExtra::layer(sp.points(V_data, pch = 1, cex = 0.5, col = "#000000"))
  }
}

```

```

    spatial_rdm_1[j] <- list(sp.mean3V)
    names(spatial_rdm_1)[j] <- names(posterior_samples_2[[i]])[j]
  }
  spatial_rdm_2[i] <- list(spatial_rdm_1)
}

plots <- do.call(c, spatial_rdm_2)

lay = rbind(c(1, 2, 3, 4, 5, 6, 7),
            c(8, 9, 10, 11, 12, 13, 14),
            c(15, 16, 17, 18, 19, 20, 21),
            c(22, 23, 24, 25, 26, 27, 28),
            c(29, 30, 31, 32, 33, 34, 35),
            c(36, 37, 38, 39, 40, 41, 42),
            c(43, 44, 45, 46, 47, 48, 49))

grStr <- arrangeGrob(grobs = plots, layout_matrix = lay)

# pdf("outputs/V_rdm_effects.pdf", width = 3*10)
grid.draw(grStr)
# dev.off()

```


Annex 5

Multi-colony approaches to study migratory and foraging strategies in pelagic seabirds

Virginia Morera Pujol

September 2019

#####SPDE - INLA modelling for tracking data with fixed, random and spatial effect

We first load the dataset, the object with the projections, the mesh, the boundary polygon and the polygon to cover the outer boundary.

```
load("definitive_datasets/ALLth_chick.Rdata") # point data
load("data/new_bound_chick.Rdata")
load("data/definitive_mesh_MClara.Rdata") # mesh
load("data/barrier.Rdata") # mask for outer boundary
load("data/new_bound_chick.Rdata") #boundary of the study area

# make sure everything has the same CRS (i.e. none)
proj4string(bound) = inla.CRS()
proj4string(ds2) = inla.CRS()
proj4string(poly.barrier) = inla.CRS()
meshM$crs <- NULL
M_data <- ds2[ds2$Colony == "MClara",]
ds.ppp <- ppp(M_data@coords[,1], M_data@coords[,2],
             bound@bbox[1,], bound@bbox[2,]) # generate the ppp from our data

wrld <- rworldmap::getMap(resolution = "high")
wrld <- spTransform(wrld,
                  CRS("+proj=aea +lat_1=20 +lat_2=60 +lat_0=40 +lon_0=-96\n
                    +x_0=0 +y_0=0 +ellps=GRS80 +datum=NAD83 +units=km +no_defs"))
```

1 Prepare the response vector

First of all, we need to prepare the response for the model, the y . For that, we will calculate a weight for each mesh node, that is related to the number of points present in each triangle (so, related to the intensity). However, we need to specify that the weight is 0 in the nodes of the outer boundary, because we don't want to model intensity in the outer boundary. The code has been extracted from the chapter 4 of the spde manual here: <http://inla.r-inla-download.org/r-inla.org/tutorials/spde/spde-tutorial.pdf>

```
(nv <- meshM$n) # number of mesh nodes
(n <- nrow(M_data)) # number of data points

# calculate weights for each meshnode depending on whether they are inside
# or outside the boundary (outside weight is 0)

dmesh <- inla.mesh.dual(meshM) # ? More info https://doi.org/10.1093/biomet/asv064

sum(w <- sapply(1:length(dmesh), function(i) {
  if (gIntersects(dmesh[i,], bound))
    return(gArea(gIntersection(dmesh[i,], bound)))
  else return(0)
}))
```

```
table(w > 0) # false = nodes with weight 0 (outside the boundary)
            # true = nodes with some weight (inside the boundary)
```

There are two response vectors, the y that has 1 for every data point and 0 for every mesh node (i.e. presence/absence), and the e or expected values, that has 0 for every data point and the weight at the node for every node. It is important that in both the order is the same: values corresponding to the meshnodes first, values corresponding to the data later.

```
length(y.pp <- rep(0:1, c(nv, n))) # response (0 for all meshnodes, 1 for all data points)
summary(y.pp) # all 0 and 1
```

```
length(e.pp <- c(w, rep(0, n)))
summary(e.pp) # different values
```

2 Prepare the covariates

From the original set of ~20 covariates, the ones used here have been selected taking into account colinearity between them (Pearson correlation coefficients) and their effect on the data assessed elsewhere.

We'll model the covariates with a random walk model of order 2 (rw2), which needs the covariate values to be grouped in equidistant classes. This model is useful for capturing potential non-linear effects of the covariate (for more info, type `inla.doc("rw2")`). For that, we'll load the geographical and environmental covariates as rasters, extract the value at the mesh nodes and data points, scale them to mean = 0 and sd = 1 with the function `scale()` and then group each of the covariate vectors using the function `inla.group()`. All these will be done inside loops for all the covariates at the same time, and needs to be done for each model (each colony) as the data points will be different so the value of the covariates at each point must be extracted for each.

```
# load rasters
env_layers <- load_var(
  path = "env_vars",
  files = c("log_Ch1A.tif", "log_current_strength.tif", "log_SST_Gradient.tif",
            "slope.tif", "SST.tif", "SST_anomaly.tif", "MClara_dist.tif",
            "wind_strength.tif"),
  format = ".tif", Norm = T)

raster_env <- stack()
for (i in 1:nlayers(env_layers)) {
  raster_env <- stack(raster_env, env_layers[[i]])}

# extract covariate values at presences for rug of random effect plots
for_rug <- list()
for (i in 1:nlayers(raster_env)) {
  # i = 1
  cov <- raster::extract(raster_env[[i]], M_data)
  for_rug[i] <- list(cov)
}
names(for_rug) <- names(raster_env)

# convert mesh nodes into spatial points to extract covariates
mshpts2D <- SpatialPoints(coords = cbind(meshM$loc[,1], meshM$loc[,2]))

# extract covariate values at mesh points and datapoints
cov_list <- list()
for (i in 1:(nlayers(env_layers))) {
```

```

# i = 2
cov <- env_layers[[i]]
covariate <- c(raster::extract(cov, mshpts2D),
              raster::extract(cov, M_data)) # order is important!
covariate[is.na(covariate)] <- mean(covariate, na.rm = T)
cov_list[i] <- list(covariate)
}
names(cov_list) <- names(env_layers)

## do the same with original values for plotting covar effects
env_layers_orig <- load_var(
  path = "env_vars",
  files = c("log_Ch1A.tif", "log_current_strength.tif", "log_SST_Gradient.tif",
            "slope.tif", "SST.tif", "SST_anomaly.tif", "MClara_dist.tif",
            "wind_strength.tif"),
  format = ".tif", Norm = F)

raster_env_orig <- stack()
for (i in 1:nlayers(env_layers_orig)) {
  raster_env_orig <- stack(raster_env_orig, env_layers_orig[[i]])}

# extract covariate values at mesh points and datapoints
cov_list_orig <- list()
for (i in 1:nlayers(raster_env_orig)) {
  # i = 2
  cov <- raster_env_orig[[i]]
  covariate <- c(raster::extract(cov, mshpts2D), raster::extract(cov, M_data))
  covariate[is.na(covariate)] <- mean(covariate, na.rm = T)
  cov_list_orig[i] <- list(covariate)
}

names(cov_list_orig) <- names(raster_env_orig)

covars <- do.call(cbind.data.frame, cov_list)
# covars is now a df with covars in the columns and as many rows
# as meshpoints + datapoints

# generate new names for the grouped covariates, with the original name
# and "LEV" (for "levels")
new_names <- list()
grouped_covars <- list()
for (i in seq_along(names(covars))) {
  # i = 1
  new_names[i] <- list(paste0(names(covars)[i], "LEV"))
  grouped_covars[i] <- list(inla.group(covars[,i],
                                     n = 20)) # 20 groups to begin with. If it's too
                                             # wiggly you can reduce the
                                             # number of groups
                                             # ("cheating", but it works)
}

# add the grouped covars to the dataframe of original covars
new_covars <- do.call(cbind.data.frame, grouped_covars)

```

```
names(new_covars) <- unlist(new_names)
covars <- cbind(covars, new_covars)
covars$b0 <- 1 # add intercept
names(covars)
```

3 Spatial models

3.a Construct A matrix

We need to construct a projector matrix (A) to connect the mesh nodes to our data points. The following lines are taken from the spde tutorial <http://inla.r-inla-download.org/r-inla.org/tutorials/spde/spde-tutorial.pdf>

```
lmat <- inla.spde.make.A(meshM, M_data@coords)
imat <- Diagonal(nv, rep(1, nv))
A.pp <- rbind(imat, lmat)
```

3.b Create projector matrix

Don't ask. Just... do it, it works.

```
projM = inla.mesh.projector(meshM,
                             xlim = poly.barrier@bbox[1,], ylim = poly.barrier@bbox[2,],
                             dims = c(280*2, 245*2)) # this is the dimensions of the covars
                                                         # rasters/10. For definitive plots
                                                         # make with more resolution (*2)
```

3.c Define spatial random effect model

For more info in the matern model try `inla.doc("matern")` and `?inla.spde2.pcmatern` for an explanation on the use of pc priors for the matern model. The alpha for the spde model is set to $3/2$ following Finn's suggestion here <https://groups.google.com/forum/#!searchin/r-inla-discussion-group/overfitting%20spde%7Csort:date/r-inla-discussion-group/ZhZVu8YPI8I/jUIOjFDBQAJ>

We begin with a small range, as small as the mesh's triangle side length, so it is the smaller resolution the spatial effect can detect, as per Hakoona's suggestion here <https://groups.google.com/forum/embed/?parenturl=http%3A%2F%2Fwww.r-inla.org%2Fcomments-1&service=jotspot&ul=1&theme=default&place=forum%2Fr-inla-discussion-group&showpopout=true&showsearch=true#!searchin/r-inla-discussion-group/long%20tail%20range%20spde%20random%20effect/r-inla-discussion-group/ZhZVu8YPI8I/xB9xWHkHCQAJ>.

After trying different values for the sigma, we settle for 0.2

However, we are going to perform a sensitivity test for both these parameters (in another script)

```
spde <- inla.spde2.pcmatern(mesh = meshM, alpha = 3/2, ### mesh and smoothness parameter
                             prior.range = c(80, 0.01), ### P(practic.range < 50)
                             prior.sigma = c(0.2, 0.01)) ### P(sigma > 0.2)
field.indices <- inla.spde.make.index("sp_rdm", n.spde = spde$n.spde)
```

3.d Define priors for unstructured random effect (for now, we're not using this, because it keeps causing trouble)

Now we define the priors for the unstructured (iid) random effect. `inla.doc("iid")` for more info

We are not doing this in the end, as it was confounding with the covariates and spatial random effect and capturing all the variability of the data. Probably would be ideal to find a spde version of this paper: <https://doi.org/10.1111/rssc.12321>

```
hyper.iid <- list(prec = list(prior = 'pc.prec',
                             param = c(0.1, 0.01))) # half the sigma of the spde
```

3.e Stack

Now we can stack everything (the parts of the stack corresponding to the iid random effect are commented out)

```
stack.spat <- inla.stack(data = list(y = y.pp, e = e.pp),
                        A = list(A.pp, #, # for the spatial effect
                                # 1, # for the iid model
                                1), # for the rw and linear
                        tag = 'dat',
                        effects = list(field.indices, #,
                                       covars))
                                # idx = 1:length(y.pp))) #
```

3.f Null model

We first run the model only with spatial effect, to compare the spde posterior to that of the model with covariates and see that the covariates “make a difference”

```
frml0M <- y ~ -1 + b0 +
  f(sp_rdm, model = spde) +
  # f(iidx, model="iid", hyper=hyper.iid) +
  NULL

mdl0M <- inla(frml0M,
              family = "poisson",
              data = inla.stack.data(stack.spat),
              control.predictor = list(A = inla.stack.A(stack.spat),
                                       compute = TRUE, link = 1),
              E = inla.stack.data(stack.spat)$e,
              control.compute = list(dic = T, waic = T, config = T),
              # control.fixed = list(prec = 0.5, prec.intercept = 0.5, mean.intercept = 0),
              # control.inla = list(int.strategy = "eb", strategy = "gaussian"),
              verbose = T)
```

3.g Formula

We now specify the formula, where we remove the default intercept (-1) because we have added it to our covars as a new column of 1s (b0). We specify all linear and non linear effects and will comment them out in successive steps as we discard them. We also specify the random effect, and the iid effect (even if we comment it out for now). The last NULL row is there so we can comment in and out the lines of the formula without worrying about adding or removing the “+” sign at the end of the formula.

We add all the covariates as linear AND non-linear following the example in Lombardo *et al.* 2018, <https://link.springer.com/article/10.1007/s00477-018-1518-0>.

For the non-linear effects, `scale.model = T` makes the model scaled to have an average variance of 1. This makes prior specification much easier, and the effects comparable. Inside hyper, we specify the hyperparameters for the model. We use PC priors (penalised complexity, Simpson *et al.* 2015, <https://arxiv.org/abs/1403.4630>), for which we only have to specify the precision (inverse of the σ) which controls the wigglyness as it controls the strenght of the dependency between a value and its neighbours. The problem is that it also controls the size of the effect, so shrinking it too much renders the effects negligible. For that reason, we should play with

τ and the number of groups in the covariate to find a compromise between wigglyness and size of the effect (not statistically ideal, but practical). The specification of the prior is detailed in `inla.doc("pc.prec")`

The prior for the `rw2` models is for the hyperparameter “precision”, but it is actually set to the sigma (inverse). We initially set the prior as the SD of the covariate, and multiply it by increasing values (2, 5, 10, 20, ...) until we achieve a “wigglyness” that makes sense and a posterior precision that is not an absurd value. (very “heuristic”, I know, but we’ve got nothing better)

```
names(covars)

frml3M <- y ~ -1 + b0 +
  log_Ch1A +
  # log_current_strength +
  # log_SST_Gradient +
  slope +
  SST +
  # SST_Anomaly +
  # wind_strength +
  MClara_dist +
  f(log_Ch1ALEV, model = 'rw2', scale.model = T,
    hyper = list(theta = list(prior = "pc.prec",
      param = c(sd(covars$log_Ch1ALEV)*25, 0.01)))) +
  # f(log_current_strengthLEV, model = 'rw2', scale.model = T,
  # hyper = list(theta = list(prior = "pc.prec",
  #   param = c(sd(covars$log_current_strengthLEV)*20, 0.01)))) +
  # f(log_SST_GradientLEV, model = 'rw2', scale.model = T,
  # hyper = list(theta = list(prior = "pc.prec",
  #   param = c(sd(covars$log_SST_GradientLEV)*200, 0.01)))) +
  # f(slopeLEV, model = 'rw2', scale.model = T,
  # hyper = list(theta = list(prior = "pc.prec",
  #   param = c(sd(covars$slopeLEV)*100, 0.01)))) +
  f(SSTLEV, model = 'rw2', scale.model = T,
    hyper = list(theta = list(prior = "pc.prec",
      param = c(sd(covars$SSTLEV)*70, 0.01)))) +
  # f(SST_AnomalyLEV, model = 'rw2', scale.model = T,
  # hyper = list(theta = list(prior = "pc.prec",
  #   param = c(sd(covars$SST_AnomalyLEV)*20, 0.01)))) +
  # f(wind_strengthLEV, model = 'rw2', scale.model = T,
  # hyper = list(theta = list(prior = "pc.prec",
  #   param = c(sd(covars$wind_strengthLEV)*50, 0.01)))) +
  f(MClara_distLEV, model = 'rw2', scale.model = T,
    hyper = list(theta = list(prior = "pc.prec",
      param = c(sd(covars$MClara_distLEV)*5, 0.01)))) +
  f(sp_rdm, model = spde) +
  # f(iidx, model="iid", hyper=hyper.iid) +
  NULL
```

3.h Run the model

And finally we run the spatial model. This might take several minutes

```
mdl3M <- inla(frml3M,
  family = "poisson",
  data = inla.stack.data(stack.spat),
  control.predictor = list(A = inla.stack.A(stack.spat), compute = TRUE,
    link = 1),
```

```

E = inla.stack.data(stack.spat)$e,
control.compute = list(dic = T, waic = T, config = T),
# control.fixed = list(prec = 0.5, prec.intercept = 0.5, mean.intercept = 0),
control.inla = list(int.strategy = "eb",
                    strategy = "gaussian"), # this speeds up computation
                                           # comment it out for final model
verbose = T) # to check process and output. Only for the brave
beepr::beep(sound = 10)

```

3.g Output

* Fixed effects

Once more, we extract the coefficients for the linear effects

```

fxd.effects3M <- cbind(rownames(mdl3M$summary.fixed), mdl3M$summary.fixed)
names(fxd.effects3M) <- c("effect", "mean", "sd", "infCI",
                        "median", "supCI", "mode", "kld")
fxd.effects3M <- fxd.effects3M[-1,]

median3M <- ggplot(fxd.effects3M, aes(x = effect, y = median)) + #ylim(-1.7, 1.7) +
  geom_errorbar(aes(ymin = infCI, ymax = supCI), width = 0.1) +
  geom_point() +
  geom_hline(yintercept = 0) + ggtitle("spatial model") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 30, hjust = 1))

median3M

```

* Random effects

This shows only the random effect (i.e. the deviations from the mean “fixed effect”). If the deviations are important we’ll keep both random and fixed, if they are not we’ll keep only the fixed. For final plotting we’ll add up the fixed and random effect to have a general view of the effect of the covariate

```

par(mfrow = c(3, 1), mar = c(3,3,1,0.3), mgp = c(2,1,0))
for (i in 1:3) {
  # i = 1
  cov_name <- substr(names(mdl3M$summary.random)[i], 1,
                    nchar(names(mdl3M$summary.random)[i]) - 3)
  scaled_vals <- cov_list[[cov_name]]
  unscaled_vals <- cov_list_orig[[cov_name]]
  post <- mdl3M$summary.random[[i]]
  ylims <- c(min(post$`0.5quant` - 1),
             max(post$`0.5quant` + 1))
  lims <- max(abs(ylims))
  plot(x = post$ID,
       y = post$`0.5quant`,
       type = 'l', xlab = cov_name, ylab = '', xaxt = "n", #ylim = c(-10, 20))
       ylim = c(-lims, lims))
  axis(side = 1, at = seq(from = (min(scaled_vals)),
                          to = (max(scaled_vals)), length = 6),
       labels = round(seq(from = min(unscaled_vals),
                          to = max(unscaled_vals), length = 6)))
  abline(h = 0, lty = 3)
}

```

```

lines(x = post$ID,
      y = (post$`0.025quant`),
      lty = 2)
lines(x = post$ID,
      y = (post$`0.975quant`),
      lty = 2)
rug(x = covars[,cov_name], ticksize = 0.04, col = "gray")
rug(x = for_rug[[cov_name]], ticksize = 0.02, side = 1)
}
par(mfrow = c(1,1))

```

* Hyperparameters for covariates

```

print(round(md13M$summary.hyperpar[1:3, ], 2))
par(mfrow = c(3, 1), mar = c(3,3,1,0.3), mgp = c(2,1,0))
for (i in 1:3) {
  # i = 5
  plot(md13M$marginals.hyperpar[[i]], type = 'l',
       xlab = names(md13M$marginals.hyperpar)[i],
       ylab = 'Density', xlim = c(0,100))
}
par(mfrow = c(1,1))

```

* Sample from posteriors

To predict, we are going to sample from the posterior of the fitted models. We could try to fit the model again with NA in the response in the places where we want to predict. However, that is problematic, and produces a lot of integration problems for some reason (Finn dixit), so we sample from the posterior which is just as good.

```

md10M.posterior.sample <- inla.posterior.sample(n = 1000, result = md10M,
                                              use.improved.mean = T)
md13M.posterior.sample <- inla.posterior.sample(n = 1000, result = md13M,
                                              use.improved.mean = T)

```

* Plot spatial random effects

We can now plot the spatial random effect. This has captured all the structure in our data that has not been captured by our covariates. Positive (red) values show areas where there is a clustering of points that is not captured by our covariates, and negative (blue) values denote areas where there are less points than predicted by our model. If this effect was flat we could assume that our covariates captured all the structure of our data. The range of this effect indicates the distance at which two points in this field are no longer correlated, and it can reflect the scale at which the “unobserved effect” is acting.

The posterior samples contain a lot of information, and are difficult to navigate. We need to extract the *Predictor*, and we do it following Haakon’s code in [https://haakonbakka.bitbucket.io/btopic112.html#52_how_to_access_the_right_part_of_\\$latent](https://haakonbakka.bitbucket.io/btopic112.html#52_how_to_access_the_right_part_of_$latent)

```

effect <- "sp_rdm" # select the effect we want to sample
(contents3M <- md13M$misc$configs$contents) # check the contents
id.effect3M <- which(contents3M$tag == effect) # the numerical id of the effect
ind.effect3M <- contents3M$start[id.effect3M] - 1 +
  (1:contents3M$length[id.effect3M]) # all the indices for the effect

(contents0M <- md10M$misc$configs$contents) # check the contents

```



```

id.effect0M <- which(contents0M$tag == effect) # the numerical id of the effect
ind.effect0M <- contents0M$start[ind.effect0M] - 1 +
  (1:contents0M$length[ind.effect0M]) # all the indices for the effect

# extract this part of the posterior from all the 1000 samples
spLog3M <- matrix(unlist(lapply(mdl3M.posterior.sample,
                              function(x) x$latent[ind.effect3M])),
                 byrow = T, nrow = 1000)
spLog0M <- matrix(unlist(lapply(mdl0M.posterior.sample,
                              function(x) x$latent[ind.effect0M])),
                 byrow = T, nrow = 1000)

# palette for plotting
pltt <- RColorBrewer::brewer.pal(11, "RdYlBu")
sp_cols <- colorRampPalette(rev(pltt))

# calculate median from previous matrix.02.3
median_spLog3M <- apply(spLog3M, 2, quantile, probs = c(0.5), na.rm = TRUE)[1:nv]
median_spLog0M <- apply(spLog0M, 2, quantile, probs = c(0.5), na.rm = TRUE)

rr <- range(c(median_spLog0M, median_spLog3M), na.rm = T)
lim <- max(abs(rr))
lo_lim <- -lim
hi_lim <- lim

sp.median0M <- my.levelplot(projM, median_spLog0M, col.regions = sp_cols(100),
                           at = seq(-6, 7, by = 1),
                           xlab = "", ylab = "",
                           scales = list(draw = F), contour = T)

sp.median3M <- my.levelplot(projM, median_spLog3M, col.regions = sp_cols(100),
                           at = seq(-6, 7, by = 1),
                           xlab = "", ylab = "",
                           scales = list(draw = F), contour = T)

sp.median3M <- sp.median3M +
  latticeExtra::layer(sp.polygons(poly.barrier, lwd = 1, fill = "white", col = "white")) +
  latticeExtra::layer(sp.polygons(wrld, lwd = 1, col = "black", fill = "gray"))

sp.median0M <- sp.median0M +
  latticeExtra::layer(sp.polygons(poly.barrier, lwd = 1, fill = "white", col = "white")) +
  latticeExtra::layer(sp.polygons(wrld, lwd = 1, col = "black", fill = "gray"))

grid.arrange(sp.median0M, sp.median3M, ncol = 2)

```

We can also plot the spatial random effect hyperparameters (again, beware of different x axes scale)

```

print(round(mdl3M$summary.hyperpar[4:5, ], 2))
par(mfrow = c(2,1), mar = c(3,3,1,0.3), mgp = c(2,1,0))
plot(mdl3M$marginals.hyperpar$`Stdev for sp_rdm`, type = 'l',
     xlab = expression(sigma^2), ylab = 'Density') #, xlim = c(0, 150))
plot(mdl3M$marginals.hyperpar$`Range for sp_rdm`, type = 'l',
     xlab = 'Nominal range', ylab = 'Density') #, xlim = c(0,1000))
par(mfrow = c(1,1))

```

* Predict intensity

```
effect <- "APredictor" # select the effect we want to sample
(contents3M <- mdl3M$misc$configs$contents) # check the contents
id.effect3M <- which(contents3M$tag == effect) # the numerical id of the effect
ind.effect3M <- contents3M$start[id.effect3M] - 1 +
  (1:contents3M$length[id.effect3M]) # all the indices for the effect

# extract this part of the posterior from all the 1000 samples
PredResp3M <- exp(matrix(unlist(lapply(mdl3M.posterior.sample,
                                     function(x) x$latent[ind.effect3M])),
                       byrow = T, nrow = 1000))
PredLog3M <- matrix(unlist(lapply(mdl3M.posterior.sample,
                                 function(x) x$latent[ind.effect3M])),
                   byrow = T, nrow = 1000)

# palette for plotting
pltt <- RColorBrewer::brewer.pal(9, "YlGnBu")
pred_cols <- colorRampPalette(pltt[3:9])

# calculate median from previous matrix.02.3
median_predLog3M <- apply(PredLog3M, 2, quantile, probs = c(0.5), na.rm = TRUE)[1:nv]
median_predResp3M <- apply(PredResp3M, 2, quantile, probs = c(0.5), na.rm = TRUE)[1:nv]

pred.medianResp3M <- my.levelplot(projM, median_predResp3M, col.regions = pred_cols(100),
                                at = seq(min(median_predResp3M),
                                         max(median_predResp3M), length.out = 100),
                                xlab = "", ylab = "",
                                scales = list(draw = F), contour = F,
                                main = "Predicted intensity - Response scale")

pred.medianResp3M <- pred.medianResp3M +
  latticeExtra::layer(sp.polygons(poly.barrier, lwd = 1, fill = "white", col = "white")) +
  latticeExtra::layer(sp.polygons(wrld, lwd = 1, col = "black", fill = "gray")) +
  latticeExtra::layer(sp.points(M_data, pch = 1, col = "#00000025"))

pred.medianResp3M
```

4. Repeat the model without spatial effect to compare outputs

4.a Construct stack

The first step is to construct our linear predictor with the covariates only as linear, and without spatial random field.

For that, we construct a “stack” with all the covariates, the response vectors and the projector matrix A . A is 1 in this case as there is no spatial effect.

```
stack.non_spat <- inla.stack(data = list(y = y.pp, e = e.pp),
                           A = list(1),
                           effects = list(covars))
```

4.b Formula

```
frml2M <- y ~ -1 + b0 +
  log_Ch1A +
  # log_current_strength +
  # log_SST_Gradient +
  slope +
  SST +
  # SST_Anomaly +
  # wind_strength +
  MClara_dist +
  f(log_Ch1ALEV, model = 'rw2', scale.model = T,
    hyper = list(theta = list(prior = "pc.prec",
                              param = c(sd(covars$log_Ch1ALEV)*25, 0.01)))) +
  # f(log_current_strengthLEV, model = 'rw2', scale.model = T,
  # hyper = list(theta = list(prior = "pc.prec",
  #                             param = c(sd(covars$log_current_strengthLEV)*20, 0.01)))) +
  # f(log_SST_GradientLEV, model = 'rw2', scale.model = T,
  # hyper = list(theta = list(prior = "pc.prec",
  #                             param = c(sd(covars$log_SST_GradientLEV)*200, 0.01)))) +
  # f(slopeLEV, model = 'rw2', scale.model = T,
  # hyper = list(theta = list(prior = "pc.prec",
  #                             param = c(sd(covars$slopeLEV)*100, 0.01)))) +
  f(SSTLEV, model = 'rw2', scale.model = T,
    hyper = list(theta = list(prior = "pc.prec",
                              param = c(sd(covars$SSTLEV)*70, 0.01)))) +
  # f(SST_AnomalyLEV, model = 'rw2', scale.model = T,
  # hyper = list(theta = list(prior = "pc.prec",
  #                             param = c(sd(covars$SST_AnomalyLEV)*20, 0.01)))) +
  # f(wind_strengthLEV, model = 'rw2', scale.model = T,
  # hyper = list(theta = list(prior = "pc.prec",
  #                             param = c(sd(covars$wind_strengthLEV)*50, 0.01)))) +
  f(MClara_distLEV, model = 'rw2', scale.model = T,
    hyper = list(theta = list(prior = "pc.prec",
                              param = c(sd(covars$MClara_distLEV)*5, 0.01)))) +
  NULL
```

4.c Model

Now we run the model. The options in the `inla` call are the same as for the previous model, there will be changes when we add the spatial random effect

```
mdl2M <- inla(frml2M,
  family = "poisson",
  data = inla.stack.data(stack.non_spat),
  control.predictor = list(A = inla.stack.A(stack.non_spat), compute = TRUE),
  E = inla.stack.data(stack.non_spat)$e,
  control.compute = list(dic = T, waic = T, config = T),
  control.inla = list(int.strategy = "eb", strategy = "gaussian"),
  verbose = T)
summary(mdl2M)
```

4.d Output

* Fixed effects

We now extract the coefficients of the linear effects, and compare them with those of the only linear model.

```
fxd.effects2M <- cbind(rownames(mdl2M$summary.fixed), mdl2M$summary.fixed)
names(fxd.effects2M) <- c("effect", "mean", "sd",
                        "infCI", "median", "supCI", "mode", "kld")
fxd.effects2M <- fxd.effects2M[-1,]

median2M <- ggplot(fxd.effects2M, aes(x = effect, y = median)) + #ylim(-1.7, 1.7) +
  geom_errorbar(aes(ymin = infCI, ymax = supCI), width = 0.1) +
  geom_point() +
  geom_hline(yintercept = 0) + ggtitle("spatial model") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 30, hjust = 1))

fxd.effects3M$model <- "spatial"
fxd.effects2M$model <- "non-linear"
fxd.effects_MC <- rbind(fxd.effects3M, fxd.effects2M)

pd <- 0.4
ggplot(fxd.effects_MC, aes(x = effect, y = median, col = model)) + #ylim(-1.7, 1.7) +
  geom_errorbar(aes(ymin = infCI, ymax = supCI), width = 0.1,
              position = position_dodge(pd)) +
  geom_point(position = position_dodge(pd)) +
  geom_hline(yintercept = 0) + ggtitle("non-linear") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 30, hjust = 1))
```

* Random effects

And now we plot the shape of the non-linear effects, to which we add the coefficient of the corresponding linear effect

```
par(mfrow = c(3, 1), mar = c(3,3,1,0.3), mgp = c(2,1,0))
for (i in 1:3) {
  # i = 1
  cov_name <- substr(names(mdl2M$summary.random)[i], 1,
                    nchar(names(mdl2M$summary.random)[i]) - 3)
  scaled_vals <- cov_list[[cov_name]]
  unscaled_vals <- cov_list_orig[[cov_name]]
  post <- mdl2M$summary.random[[i]]
  ylims <- c(min(post$`0.5quant` +
                fxd.effects2M[rownames(fxd.effects2M) == cov_name,4] - 1),
            max(post$`0.5quant` +
                fxd.effects2M[rownames(fxd.effects2M) == cov_name,6] + 1))
  lims <- max(abs(ylims))
  plot(x = post$ID,
       y = post$`0.5quant` + fxd.effects2M[rownames(fxd.effects2M) == cov_name,5],
       type = 'l', xlab = cov_name, ylab = '', xaxt = "n", #ylim = c(-10, 20))
       ylim = c(-lims, lims))
  axis(side = 1, at = seq(from = (min(scaled_vals)),
                          to = (max(scaled_vals)), length = 6),
       labels = round(seq(from = min(unscaled_vals), to = max(unscaled_vals),
                          length = 6)))
  abline(h = 0, lty = 3)
  lines(x = post$ID,
```

```

    y = (post$`0.025quant` + fxd.effects2M[rownames(fxd.effects2M) == cov_name,5]),
    lty = 2)
lines(x = post$ID,
      y = (post$`0.975quant` + fxd.effects2M[rownames(fxd.effects2M) == cov_name,5]),
      lty = 2)
rug(x = covars[,cov_name], ticksize = 0.04, col = "gray")
rug(x = for_rug[[cov_name]], ticksize = 0.02, side = 1)
}
par(mfrow = c(1,1))

```

Compare with random effects of spatial model

```

par(mfrow = c(2, 2), mar = c(3,3,1,0.3), mgp = c(2,1,0))
for (i in 1:3) {
  # i = 1
  cov_name <- substr(names mdl3M$summary.random)[i], 1,
                    nchar(names(mdl3M$summary.random)[i]) - 3)
  scaled_vals <- cov_list[[cov_name]]
  unscaled_vals <- cov_list_orig[[cov_name]]
  post3 <- mdl3M$summary.random[[i]]

  post2 <- mdl2M$summary.random[[i]]
  ylims <- c(min(min(post3$`0.5quant` +
                    fxd.effects3M[rownames(fxd.effects3M) == cov_name,4] - 1),
              min(post2$`0.5quant` +
                    fxd.effects2M[rownames(fxd.effects2M) == cov_name,4] - 1)),
            max(max(post3$`0.5quant` +
                    fxd.effects3M[rownames(fxd.effects3M) == cov_name,6] + 1),
              max(post2$`0.5quant` +
                    fxd.effects2M[rownames(fxd.effects2M) == cov_name,6] + 1)))
  lims <- max(abs(ylims))
  plot(x = post3$ID,
       y = post3$`0.5quant` + fxd.effects3M[rownames(fxd.effects3M) == cov_name,5],
       type = 'l', xlab = cov_name, ylab = '', xaxt = "n", #ylim = c(-10, 20))
       ylim = c(-lims, lims))
  abline(h = 0, lty = 3)
  lines(x = post3$ID,
        y = (post3$`0.025quant` + fxd.effects3M[rownames(fxd.effects3M) == cov_name,5]),
        lty = 2)
  lines(x = post3$ID,
        y = (post3$`0.975quant` + fxd.effects3M[rownames(fxd.effects3M) == cov_name,5]),
        lty = 2)
  lines(x = post2$ID,
        y = post2$`0.5quant` + fxd.effects2M[rownames(fxd.effects2M) == cov_name,5],
        xlab = cov_name, ylab = '', xaxt = "n", #ylim = c(-10, 20))
        ylim = c(-lims, lims), col = "gray")
  lines(x = post2$ID,
        y = (post2$`0.025quant` + fxd.effects2M[rownames(fxd.effects2M) == cov_name,5]),
        lty = 2, col = "gray")
  lines(x = post2$ID,
        y = (post2$`0.975quant` + fxd.effects2M[rownames(fxd.effects2M) == cov_name,5]),
        lty = 2, col = "gray")
  rug(x = covars[,cov_name], ticksize = 0.04, col = "gray")
  rug(x = for_rug[[cov_name]], ticksize = 0.02, side = 1)
}

```

```

axis(side = 1, at = seq(from = (min(scaled_vals)),
                        to = (max(scaled_vals)), length = 6),
      labels = round(seq(from = min(unscaled_vals), to = max(unscaled_vals),
                          length = 6)))
}
# add slope (linear)
post_lin3 <- post2
post_lin3$`0.5quant` <- fxd.effects3M[2,5]
post_lin3$`0.025quant` <- fxd.effects3M[2,4]
post_lin3$`0.975quant` <- fxd.effects3M[2,6]
post_lin2 <- post2
post_lin2$`0.5quant` <- fxd.effects2M[2,5]
post_lin2$`0.025quant` <- fxd.effects2M[2,4]
post_lin2$`0.975quant` <- fxd.effects2M[2,6]
ylims <- c(min(post_lin3$`0.025quant`, post_lin2$`0.025quant`) - 0.5,
           max(post_lin3$`0.975quant`, post_lin2$`0.975quant`) + 0.5)
lims <- max(abs(ylims))
plot(x = post_lin3$ID,
     y = post_lin3$`0.5quant`,
     type = 'l', xlab = "Slope", ylab = '', xaxt = "n", #ylim = c(-10, 20))
     ylim = c(-lims, lims))
abline(h = 0, lty = 3)
lines(x = post_lin3$ID,
      y = post_lin3$`0.025quant`,
      lty = 2)
lines(x = post_lin3$ID,
      y = post_lin3$`0.975quant`,
      lty = 2)
lines(x = post_lin2$ID,
      y = post_lin2$`0.5quant`,
      xlab = cov_name, ylab = '', xaxt = "n", #ylim = c(-10, 20))
      ylim = c(-lims, lims), col = "gray")
lines(x = post_lin2$ID,
      y = post_lin2$`0.025quant`,
      lty = 2, col = "gray")
lines(x = post_lin2$ID,
      y = post_lin2$`0.975quant`,
      lty = 2, col = "gray")
rug(x = covars[,"slope"], ticksize = 0.04, col = "gray")
rug(x = for_rug[["slope"]], ticksize = 0.02, side = 1)
axis(side = 1, at = seq(from = min(post_lin3$ID),
                        to = max(post_lin3$ID), length = 6),
      labels = round(seq(from = min(cov_list_orig[["slope"]]),
                          to = max(cov_list_orig[["slope"]]), length = 6), 2))
par(mfrow = c(1, 1))

```

We also plot the hyperparameters of the non-linear effects

```

par(mfrow = c(3, 1), mar = c(3,3,1,0.3), mgp = c(2,1,0))
for (i in 1:3) {
  # i = 1
  plot(md12M$marginals.hyperpar[[i]], type = 'l',
       xlab = names(md12M$marginals.hyperpar)[i],
       ylab = 'Density', xlim = c(0,10))
}

```

```
}
par(mfrow = c(1,1))
```

5 Finally, model with all covariates as linear, to compare.

5.a Formula

Next, we construct the formula, which specifies each of the linear effects and the intercept. In this case, we can add the covariates without grouping, the grouped covariates will be used for the rw2 models.

```
frml1M <- y ~ -1 + b0 +
  log_Ch1A +
  # log_current_strength +
  # log_SST_Gradient +
  slope +
  SST +
  # SST_Anomaly +
  # wind_strength +
  MClara_dist +
  NULL
```

5.b Run model

We run the model with the following code, which specifies the family (Poisson, tupe `inla.doc("poisson")` for more info), where to take the data from, where to find the A matrix, the expected values, and several control parameters such as whether to compute dic and waic, and the integration strategy.

```
mdl1M <- inla(frml1M,
  family = "poisson",
  data = inla.stack.data(stack.non_spat),
  control.predictor = list(A = inla.stack.A(stack.non_spat), compute = TRUE),
  E = inla.stack.data(stack.non_spat)$e,
  control.compute = list(dic = T, waic = T, config = T),
  control.inla = list(int.strategy = "eb", strategy = "gaussian"),
  verbose = T)
summary(mdl1M)
```

5.c Output

From this model we can extract the coefficients of the linear effects, and plot them to see how much they differ from 0 (i.e. how significant they are)

```
fxd.effects1M <- cbind(rownames(mdl1M$summary.fixed), mdl1M$summary.fixed)
names(fxd.effects1M) <- c("effect", "mean", "sd",
  "infCI", "median", "supCI", "mode", "kld")
fxd.effects1M <- fxd.effects1M[-1,] # remove intercept, we don't want to plot it

median1M <- ggplot(fxd.effects1M, aes(x = effect, y = median)) + #ylim(-1.7, 1.7) +
  geom_errorbar(aes(ymin = infCI, ymax = supCI), width = 0.1) +
  geom_point() +
  geom_hline(yintercept = 0) + ggtitle("spatial model") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 30, hjust = 1))

fxd.effects1M$model <- "linear"
fxd.effects_MC <- rbind(fxd.effects_MC, fxd.effects1M)
```

```

pd <- 0.4
fxd_MC <- ggplot(fxd.effects_MC, aes(x = effect, y = median, col = model)) +
  geom_errorbar(aes(ymin = infCI, ymax = supCI), width = 0.1,
               position = position_dodge(pd)) +
  geom_point(position = position_dodge(pd)) +
  geom_hline(yintercept = 0) + ggtitle("non-linear") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 30, hjust = 1))

fxd_MC

```

6 Predict from the three models

6.a Sample from the posterior

Now we have the three models, we'll predict from the three to compare outputs and see if the model has improved with the addition of non-linear covariates and a spatially structured random effect.

```

mdl1M.posterior.sample <- inla.posterior.sample(n = 1000, result = mdl1M,
                                              use.improved.mean = T)
mdl2M.posterior.sample <- inla.posterior.sample(n = 1000, result = mdl2M,
                                              use.improved.mean = T)
mdl3M.posterior.sample <- inla.posterior.sample(n = 1000, result = mdl3M,
                                              use.improved.mean = T)

```

6.b Extract Predictor

These posterior samples contain a lot of information, and are difficult to navigate. We need to extract the *Predictor*, and we do it following Haakon's code in https://haakonbakka.bitbucket.io/btopic112.html#52_how_to_access_the_right_part_of_latent

```

(contents1M <- mdl1M$misc$configs$content) # check the contents
effect <- "APredictor" # select the effect we want to sample
id.effect <- which(contents1M$tag == effect) # the numerical id of the effect
ind.effect <- contents1M$start[id.effect] - 1 +
  (1:contents1M$length[id.effect]) # all the indices for the effect
Predictor_sample1M <- lapply(mdl1M.posterior.sample,
                            function(x) x$latent[ind.effect]) # extract this part of the
                                                                # posterior from all the
                                                                # 1000 samples

(contents2M <- mdl2M$misc$configs$content)
effect <- "APredictor"
id.effect <- which(contents2M$tag == effect)
ind.effect <- contents2M$start[id.effect] - 1 + (1:contents2M$length[id.effect])
Predictor_sample2M <- lapply(mdl2M.posterior.sample,
                            function(x) x$latent[ind.effect])

(contents3M <- mdl3M$misc$configs$content)
effect <- "APredictor"
id.effect <- which(contents3M$tag == effect)
ind.effect <- contents3M$start[id.effect] - 1 + (1:contents3M$length[id.effect])
Predictor_sample3M <- lapply(mdl3M.posterior.sample,
                            function(x) x$latent[ind.effect])

```


From these predictor samples, we can extract mean and σ of the posterior at each of the integration points (nodes of the mesh) and then use the projector matrix to plot it. We can extract the posterior at the response scale or at the log scale. The response scale is usually easier to interpret, but the log scale is visually clearer.

6.c Project and plot predicted intensity

We now convert the lists that are the output of `inla.posterior.sample()` into matrices, with one column for each integration point (mesh node) and one row for each sample. And then we calculate `colMeans()` and `colSds()` and with the custom function `my.levelplot()`* (I should find out where I found it) and the help of the projector matrix generated above we plot it.

The function `my.levelplot()` contains the function `inla.mesh.project()` that is an important step to this in combination with `inla.mesh.projector()` although I've long given up on understanding exactly what it does. It just works.

```
Pred_Response1M <- exp(matrix(unlist(Predictor_sample1M),
                             byrow = T, nrow = 1000)) #prediction in the response scale
Pred_Log1M <- matrix(unlist(Predictor_sample1M), byrow = T,
                    nrow = 1000) #prediction in the log scale
median_PredResp1M <- apply(Pred_Response1M, 2, quantile, probs = c(0.5),
                           na.rm = TRUE)[1:nv]

# RWIQR: relative width of the interquantile range (https://doi.org/10.1214/17-AOAS1078)
CI_predResp1M <- apply(Pred_Response1M, 2, quantile, probs = c(0.025, 0.975),
                      na.rm = TRUE)[,1:nv]
RWIQR1M <- abs((CI_predResp1M[2,] - CI_predResp1M[1,])/median_PredResp1M)
Pred_Response2M <- exp(matrix(unlist(Predictor_sample2M),
                             byrow = T, nrow = 1000)) #prediction in the response scale

Pred_Log2M <- matrix(unlist(Predictor_sample2M), byrow = T,
                    nrow = 1000) #prediction in the log scale
median_PredResp2M <- apply(Pred_Response2M, 2, quantile, probs = c(0.5),
                           na.rm = TRUE)[1:nv]
CI_predResp2M <- apply(Pred_Response2M, 2, quantile, probs = c(0.025, 0.975),
                      na.rm = TRUE)[,1:nv]
RWIQR2M <- abs((CI_predResp2M[2,] - CI_predResp2M[1,])/median_PredResp2M)

Pred_Response3M <- exp(matrix(unlist(Predictor_sample3M),
                             byrow = T, nrow = 1000)) #prediction in the response scale
Pred_Log3M <- matrix(unlist(Predictor_sample3M), byrow = T,
                    nrow = 1000) #prediction in the log scale
median_PredResp3M <- apply(Pred_Response3M, 2, quantile, probs = c(0.5),
                           na.rm = TRUE)[1:nv]
CI_predResp3M <- apply(Pred_Response3M, 2, quantile, probs = c(0.025, 0.975),
                      na.rm = TRUE)[,1:nv]
RWIQR3M <- abs((CI_predResp3M[2,] - CI_predResp3M[1,])/median_PredResp3M)

pred.median1M <- my.levelplot(projM, median_PredResp1M, col.regions = pred_cols(100),
                             at = seq(min(median_PredResp1M),
                                       max(median_PredResp1M), length.out = 100),
                             # at = seq(0, 14, length.out = 100),
                             xlab = "", ylab = "",
                             scales = list(draw = F), contour = F) +
  latticeExtra::layer(sp.polygons(poly.barrier, lwd = 1, fill = "white", col = "white")) +
  latticeExtra::layer(sp.polygons(wrld, lwd = 1, col = "black", fill = "gray")) +
```

```

latticeExtra::layer(sp.points(M_data, pch = 1, col = "#00000050", cex = 0.5))

pred.median2M <- my.levelplot(projM, median_PredResp2M, col.regions = pred_cols(100),
  at = seq(min(median_PredResp2M),
    max(median_PredResp2M), length.out = 100),
  # at = seq(0, 14, length.out = 100),
  xlab = "", ylab = "",
  scales = list(draw = F), contour = F) +
latticeExtra::layer(sp.polygons(poly.barrier, lwd = 1, fill = "white", col = "white")) +
latticeExtra::layer(sp.polygons(wrld, lwd = 1, col = "black", fill = "gray")) +
latticeExtra::layer(sp.points(M_data, pch = 1, col = "#00000050", cex = 0.5))

pred.median3M <- my.levelplot(projM, median_PredResp3M, col.regions = pred_cols(100),
  at = seq(min(median_PredResp3M),
    max(median_PredResp3M), length.out = 100),
  # at = seq(0, 14, length.out = 100),
  xlab = "", ylab = "",
  scales = list(draw = F), contour = F) +
latticeExtra::layer(sp.polygons(poly.barrier, lwd = 1, fill = "white", col = "white")) +
latticeExtra::layer(sp.polygons(wrld, lwd = 1, col = "black", fill = "gray")) +
latticeExtra::layer(sp.points(M_data, pch = 1, col = "#00000050", cex = 0.5))

pred.RWIQR3M <- my.levelplot(projM, RWIQR3M, col.regions = pred_cols(100),
  at = seq(0, 600, length.out = 100),
  # at = seq(0, 14, length.out = 100),
  xlab = "", ylab = "",
  scales = list(draw = F), contour = F,
  main = "RWIQR - Spatial Model") +
latticeExtra::layer(sp.polygons(poly.barrier, lwd = 1, fill = "white", col = "white")) +
latticeExtra::layer(sp.polygons(wrld, lwd = 1, col = "black", fill = "gray")) +
latticeExtra::layer(sp.points(M_data, pch = ".", col = "#00000075"))

pred.RWIQR2M <- my.levelplot(projM, RWIQR2M, col.regions = pred_cols(100),
  at = seq(0, 1500, length.out = 100),
  # at = seq(0, 14, length.out = 100),
  xlab = "", ylab = "",
  scales = list(draw = F), contour = F,
  main = "RWIQR - Non-linear model") +
latticeExtra::layer(sp.polygons(poly.barrier, lwd = 1, fill = "white", col = "white")) +
latticeExtra::layer(sp.polygons(wrld, lwd = 1, col = "black", fill = "gray")) +
latticeExtra::layer(sp.points(M_data, pch = ".", col = "#00000075"))

pred.RWIQR1M <- my.levelplot(projM, RWIQR1M, col.regions = pred_cols(100),
  at = seq(0, 2, length.out = 100),
  # at = seq(0, 14, length.out = 100),
  xlab = "", ylab = "",
  scales = list(draw = F), contour = F,
  main = "RWIQR - Linear model") +
latticeExtra::layer(sp.polygons(poly.barrier, lwd = 1, fill = "white", col = "white")) +
latticeExtra::layer(sp.polygons(wrld, lwd = 1, col = "black", fill = "gray")) +
latticeExtra::layer(sp.points(M_data, pch = ".", col = "#00000075"))

```

```
plots <- list(pred.median1M, pred.median2M, pred.median3M,
             pred.RWIQR1M, pred.RWIQR2M, pred.RWIQR3M)

lay = rbind(c(1, 2, 3),
           c(4, 5, 6))

grStr <- arrangeGrob(grobs = plots, layout_matrix = lay)

grid::grid.draw(grStr)
```

7 Model validation for the three models

Cross Validation is not possible with point pattern models because the actual data are the points, so if you split the dataset in two you'll actually change the structure of the y . Also, `dic` and `waic` are not accurate for these type of models, which leaves us with no “off the shelf” way to validate our models. However, we can use some summary functions from point patterns that will allow us at least to evaluate how well our models have captured the structure of our data.

The first thing we need to do is to simulate 1000 point patterns from each of the models. The code is from Janine’s “model validation” session at the `inlabru` course (https://www.dropbox.com/s/ok2gimujqfyvq7f/06_b_mod_valid.html?dl=0). We need to make the simulations from the prediction in the log scale (for some reason).

7.a Simulate 1000 point patterns from each model

```
for_sampl1M_median <- apply(Pred_Log1M, 2, quantile, probs = c(0.5),
                           na.rm = TRUE)[1:nv]
mdl1M.simulations <- lapply(1:1000, function(i) {
  spatial.to.ppp(sample.lgcp(meshM, for_sampl1M_median, samplers = bound),
                 samplers = bound)})

for_sampl2M_median <- apply(Pred_Log2M, 2, quantile, probs = c(0.5),
                           na.rm = TRUE)[1:nv]
mdl2M.simulations <- lapply(1:1000, function(i) {
  spatial.to.ppp(sample.lgcp(meshM, for_sampl2M_median, samplers = bound),
                 samplers = bound)})

for_sampl3M_median <- apply(Pred_Log3M, 2, quantile, probs = c(0.5),
                           na.rm = TRUE)[1:nv]
mdl3M.simulations <- lapply(1:1000, function(i) {
  spatial.to.ppp(sample.lgcp(meshM, for_sampl3M_median, samplers = bound),
                 samplers = bound)})
```

7.b Plot examples to check it worked

We can plot a couple of examples of simulated patterns from each of the models to see it has worked

```
# par(mfrow = c(3,3), oma = c(0, 0, 2, 0))
par(mfrow = c(1, 3), oma = c(0, 0, 2, 0))
for (i in sample(x = 1:1000, replace = F, size = 3)) {plot(mdl1M.simulations[[i]],
                                                         main = "")}

mtext("Model 1", outer = TRUE, cex = 1.5)
par(mfrow = c(1, 1))
```

```

par(mfrow = c(1, 3), oma = c(0, 0, 2, 0))
for (i in sample(x = 1:1000, replace = F, size = 3)) {plot(md12M.simulations[[i]],
                                                         main = "")}

mtext("Model 2", outer = TRUE, cex = 1.5)
par(mfrow = c(1, 1))

par(mfrow = c(1, 3), oma = c(0, 0, 2, 0))
for (i in sample(x = 1:100, replace = F, size = 3)) {plot(md13M.simulations[[i]],
                                                         main = "")}

mtext("Model 3", outer = TRUE, cex = 1.5)
par(mfrow = c(1, 1))

# add the observed ppp to compare
par(mfrow = c(1,3))
plot(md11M.simulations[[1]], main = "Linear model")
plot(ds.ppp, add = T, col = "red")

plot(md12M.simulations[[1]], main = "Non-linear model")
plot(ds.ppp, add = T, col = "red")

plot(md13M.simulations[[1]], main = "Spatial model")
plot(ds.ppp, add = T, col = "red")
par(mfrow = c(1,1))

```

Then, we need to convert our data to a point pattern, to calculate the L Function. This L Function is related to the number of points inside an area of radius R from each point, and is calculated for several distances for each point. High values for the L function (a lot of points inside the area of radius r) means the point pattern is clustered. Low values for the L function (very few points inside the area of radius r) means the point pattern is dispersed. We will calculate the L function values from the 1000 simulated patterns that we've generated, and plot it for all the distances (r) at which is calculated. This will generate a "ribbon" or "envelope", over which we'll plot the L-function values for our observed pattern. If the observed pattern falls inside the envelope we can interpret that our model correctly captures the structure of our data. If our observed pattern is above the envelope, we can interpret that there is some clustering in our data that our model is not capturing, while if the observed pattern falls below the envelope we interpret there is some dispersion in our data that the model has not captured.

To make the plot more visually clear we will subtract the r value to the L-function values for plotting. This should generate a L-function plot more centered around 0.

7.c Calculate and plot L functions for the three models

```

md11M.Lest <- spatstat::envelope(ds.ppp, fun = spatstat::Lest,
                               simulate = md11M.simulations)

H1M <- md11M.Lest

H1M$obs <- H1M$obs - H1M$r
H1M$mmean <- H1M$mmean - H1M$r
H1M$lo <- H1M$lo - H1M$r
H1M$hi <- H1M$hi - H1M$r

plot1M <- ggplot(data = H1M) +
  theme_bw() +
  theme(axis.title.x = element_text(size = 16),

```

```

axis.text.x = element_text(size = 14),
axis.title.y = element_text(size = 16),
axis.text.y = element_text(size = 14),
panel.grid.minor.x = element_blank(), panel.grid.major.x = element_blank(),
panel.grid.minor.y = element_blank(), panel.grid.major.y = element_blank())

env1M <- plot1M + geom_line(aes(x = r, y = obs), col = "red") +
  geom_line(aes(x = r, y = lo), col = "Gray20", linetype = 2) +
  geom_line(aes(x = r, y = hi), col = "Gray20", linetype = 2) +
  ylim(-10, 125) +
  xlim(-10, 150) +
  geom_vline(xintercept = seq(10, 150, by = 10), lty = 2, col = "lightgray") +
  xlab("Radius (km)") +
  ylab("L(r) - r")

mdl2M.Lest <- spatstat::envelope(ds.ppp, fun = spatstat::Lest,
  simulate = mdl2M.simulations)

H2M <- mdl2M.Lest

H2M$obs <- H2M$obs - H2M$r
H2M$mmean <- H2M$mmean - H2M$r
H2M$lo <- H2M$lo - H2M$r
H2M$hi <- H2M$hi - H2M$r

plot2M <- ggplot(data = H2M) +
  theme_bw() +
  theme(axis.title.x = element_text(size = 16),
    axis.text.x = element_text(size = 14),
    axis.title.y = element_text(size = 16),
    axis.text.y = element_text(size = 14),
    panel.grid.minor.x = element_blank(), panel.grid.major.x = element_blank(),
    panel.grid.minor.y = element_blank(), panel.grid.major.y = element_blank())

env2M <- plot2M + geom_line(aes(x = r, y = obs), col = "red") +
  geom_line(aes(x = r, y = lo), col = "Gray20", linetype = 2) +
  geom_line(aes(x = r, y = hi), col = "Gray20", linetype = 2) +
  ylim(-10, 125) +
  xlim(-10, 150) +
  geom_vline(xintercept = seq(10, 150, by = 10), lty = 2, col = "lightgray") +
  xlab("Radius (km)") +
  ylab("L(r) - r")

mdl3M.Lest <- spatstat::envelope(ds.ppp, fun = spatstat::Lest,
  simulate = mdl3M.simulations)

H3M <- mdl3M.Lest

H3M$obs <- H3M$obs - H3M$r
H3M$mmean <- H3M$mmean - H3M$r
H3M$lo <- H3M$lo - H3M$r
H3M$hi <- H3M$hi - H3M$r

```

```

plot3M <- ggplot(data = H3M) +
  theme_bw() +
  theme(axis.title.x = element_text(size = 16),
        axis.text.x = element_text(size = 14),
        axis.title.y = element_text(size = 16),
        axis.text.y = element_text(size = 14),
        panel.grid.minor.x = element_blank(), panel.grid.major.x = element_blank(),
        panel.grid.minor.y = element_blank(), panel.grid.major.y = element_blank())

(env3M <- plot3M + geom_line(aes(x = r, y = obs), col = "red") +
  geom_line(aes(x = r, y = lo), col = "Gray20", linetype = 2) +
  geom_line(aes(x = r, y = hi), col = "Gray20", linetype = 2) +
  ylim(-10, 125) +
  xlim(-10, 150) +
  geom_vline(xintercept = seq(10, 150, by = 10), lty = 2, col = "lightgray") +
  xlab("Radius (km)") +
  ylab("L(r) - r"))

grid.arrange(env1M, env2M, env3M, ncol = 1)

```

7.d Estimate abundance and compare to observed abundance

Lastly, another way to see how well our models fit to our data, we can try to calculate the total abundance estimated from the model, and compare it to the actual number of observed points. A way to do that is from the point patterns simulated from each of the models (as we've done above). We count the number of points in each of them, and because we have 1000 simulations we obtain a mean and an sd.

```

nrow(M_data) #real abundance
Lambda1 <- unlist(lapply(md11M.simulations, function(x) x$n))
Lambda2 <- unlist(lapply(md12M.simulations, function(x) x$n))
Lambda3 <- unlist(lapply(md13M.simulations, function(x) x$n))

mean(Lambda3)
sd(Lambda3)

Lambdas <- cbind(Lambda1, Lambda2, Lambda3)
x <- as.data.frame(apply(Lambdas, 2, quantile, probs = c(0.025, 0.5, 0.975),
  na.rm = TRUE))

```