

Matthiopoulos, J. , Wakefield, E. , Jeglinski, J. W.E., Furness, R. W., Trinder, M., Tyler, G., McCluskie, A., Allen, S., Braithwaite, J. and Evans, T. (2022) Integrated modelling of seabird-habitat associations from multi-platform data: a review. *Journal of Applied Ecology*, 59(4), pp. 909-920.

The material cannot be used for any other purpose without further permission of the publisher and is for private use only.

There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

This is the peer reviewed version of the following article:

Matthiopoulos, J. , Wakefield, E. , Jeglinski, J. W.E., Furness, R. W., Trinder, M., Tyler, G., McCluskie, A., Allen, S., Braithwaite, J. and Evans, T. (2022) Integrated modelling of seabird-habitat associations from multi-platform data: a review. *Journal of Applied Ecology*, 59(4), pp. 909-920, which has been published in final form at: [10.1111/1365-2664.14114](https://doi.org/10.1111/1365-2664.14114)

This article may be used for non-commercial purposes in accordance with [Wiley Terms and Conditions for Self-Archiving](#).

<https://eprints.gla.ac.uk/263103/>

Deposited on 13 January 2022

Matthiopoulos Jason (Orcid ID: 0000-0003-3639-8172)  
 Wakefield Ewan (Orcid ID: 0000-0002-2586-1232)  
 Jeglinski Jana Wibke Elisabeth (Orcid ID: 0000-0002-7997-6756)

Handling Editor: Vitor Paiva

# INTEGRATED MODELLING OF SEABIRD-HABITAT ASSOCIATIONS FROM MULTI-PLATFORM DATA: A REVIEW

JASON MATTHIOPOULOS<sup>\*1,2</sup>, EWAN WAKEFIELD<sup>1</sup>, JANA W.E. JEGLINSKI<sup>1</sup>,  
 ROBERT W. FURNESS<sup>2</sup>, MARK TRINDER<sup>2</sup>, GLEN TYLER<sup>3</sup>, ALY MCCLUSKIE<sup>4</sup>,  
 SOPHY ALLEN<sup>5</sup>, JANELLE BRAITHWAITE<sup>6</sup> & TOM EVANS<sup>6</sup>

1. Institute of Biodiversity, Animal Health & Comparative Medicine, College of Medical, Veterinary & Life Sciences, Graham Kerr Building, University of Glasgow, Glasgow, Scotland, G12 8QQ.

2. MacArthur Green, 93 South Woodside Road, Glasgow, Scotland, G20 6NT

3. NatureScot, Great Glen House, Leachkin Road, Inverness. IV3 8NW,

4. RSPB Centre for Conservation Science, RSPB, Etive House, Beechwood Park, Inverness, Scotland, IV2 3BW

5. Natural England, Sterling House, Exeter, Devon, England

6. Marine Scotland, 1B South Victoria Quay, Edinburgh, Scotland, EH6 6QQ

\*Corresponding author: Tel.: 0141 330 5969, E-mail: [Jason.Matthiopoulos@glasgow.ac.uk](mailto:Jason.Matthiopoulos@glasgow.ac.uk)

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the [Version of Record](#). Please cite this article as doi: [10.1111/1365-2664.14114](https://doi.org/10.1111/1365-2664.14114)

This article is protected by copyright. All rights reserved.

Abstract:

1. Quantifying current and future overlap between human activities and wildlife is a core and growing aim of ecological study, spurring ever more spatial data collection and diversification of observation techniques (surveys, telemetry, citizen science etc.).
2. To meet this aim, data collected via multiple platforms, across different geographical and temporal regions, may need to be integrated, yet many ecologists remain unclear about the relationships between data types and therefore how they can be combined.
3. In seabird research, these applied questions can be particularly pressing because many human activities (e.g. tidal and wind renewables, fishing, shipping, etc.) are concentrated in coastal waters, where many seabirds also aggregate, especially while breeding. In addition, seabird coloniality and density dependence present unique analytical challenges.
4. We review the relevant literature on data integration and illustrate it with example models and data (in an accompanying R-library and vignette (J Matthiopoulos et al., 2022)), to derive methodological and quantitative guidelines for best practice in conducting joint inference for multi-platform data. We use systematic survey data to motivate the key arguments, but also overview developments in integration with other data (e.g., telemetry tracking, citizen science, mark-recapture).
5. We make recommendations on (1) the use of response and explanatory data, (2) the treatment of survey design and observation errors, (3) exploiting dependencies across space and time, (4) accounting for biological phenomena, such as commuting costs from the colony (i.e., accessibility) and density dependence, and (5) the choice of statistical framework.
6. *Synthesis and application:* Integrated analysis of multi-platform data turns many of the seabird-specific challenges into opportunities for inferring habitat associations and predicting future distributions. Our review proposes practical recommendations for data

collection and analysis that will allow seabird conservation to derive maximal benefits from these opportunities.

Keywords: Accessibility constraints, Hierarchical models, Point process models, Marine renewables, Model transferability, Observation models, Spatial modelling, Species distribution modelling

## INTRODUCTION

The fundamental questions in spatial ecology are reassuringly stable (Aarts, MacKenzie, McConnell, Fedak, & Matthiopoulos, 2008; Jason Matthiopoulos, Fieberg, & Aarts, 2020). How many individuals are within a survey area (abundance estimation), where are they (population distribution), why are they there (habitat, intraspecific and interspecific associations), where else might they be, and where might they go if the environment changes (spatial extrapolation and forecasting)? Within this setting, seabirds present us with particular applied challenges. Their seasonal connection to the coast, often in colonial networks, their capabilities for far-ranging travel (Egevang et al., 2010) contrasted with their often strong philopatry and breeding site fidelity (Steiner & Gaston, 2005), their sensitivity to changing wind fields and currents (Weimerskirch, Louzao, De Grissac, & Delord, 2012), and their vulnerability to energetic deficits and commuting costs (Elliott et al., 2013) make the investigation of accessibility constraints and density dependence particularly complicated, compared to other taxa.

Seabird conservation imperatives increasingly influence marine spatial planning and environmental policy for renewable energy, oil decommissioning, shipping and fishing (Pittman et al., 2021), posing a surfeit of difficult spatial questions on issues such as critical habitat, collisions, spatial displacement, barrier effects and secondary metapopulation effects. This has prompted a rapid increase in Species Distribution Models (SDMs) for

Accepted Article

seabirds, drawing independent inferences from multi-platform data such as ship-board and aerial (Waggitt et al., 2019), ship-board and tracking (Carroll et al., 2019; Perrow, Harwood, Skeate, Praca, & Eglington, 2015) and ship-board, aerial and tracking (Sansom, Wilson, Caldow, & Bolton, 2018) . At the same time, there is a strong trend in ecology towards integration of such multi-platform data (Fletcher et al., 2019; Isaac et al., 2020; Miller, Pacifici, Sanderlin, & Reich, 2019; Pacifici et al., 2017).

Integration across data types lends power to analyses because each observation platform can provide a different view of a species' underlying distribution. Embedding different data into a joint likelihood framework greatly improves the explanatory and predictive ability of species distribution models (Fletcher, McCleery, Greene, & Tye, 2016; Nelli, Ferguson, & Matthiopoulos, 2019; Pacifici et al., 2017; Peel et al., 2019). Integrated analyses also improve model precision (by increasing the effective sample size), model accuracy (by ameliorating the biases of one data type via corroboration with another), and the aggregated spatiotemporal extent and transferability of model estimates and predictions (a natural consequence of pooling study areas and times). Hence, data integration is particularly helpful for ameliorating two types of risk inherent in marine planning. First, by increasing precision we may find that the contracted confidence intervals around estimated impacts no longer include negative effects on wildlife. We may thus avoid unnecessary refusals of environmentally benign developments under the precautionary principle. Second, by reducing the bias in estimates of impact, we may find that the expected anthropogenic effects are, in fact, detrimental. We may therefore be able to act to lower the risk of extirpation of protected populations. Furthermore, seabird survey methodologies have developed over time from ship-based, to aerial and digital-aerial (Stephen T. Buckland et al., 2012), meaning that long-term trends simply cannot be investigated without some form of multi-platform integration.

Accepted Article

Here, we argue that the challenges posed by seabird biology and multi-platform datasets can be turned into opportunities for inference, leading to distinct recommendations for best analytical practice. Our proposed data-integration workflow navigates the difficulties posed by seabird natural history, differences in observation method, overlaps and separation in effort amount and distribution between methods, the imperfections of explanatory data, and the need for statistical robustness. The current state of the literature reviewed here primarily contains examples of integration of different survey types, systematic (i.e., designed scientific surveys) or opportunistic (e.g., citizen science). To introduce several of the key ideas and challenges, we have opted to structure the main body of our paper around integration of different types of systematic surveys. However, we conclude with an overview of emerging possibilities for integration between more diverse data types such as tracking, mark-recapture and multispecies data. To ensure our final recommendations point to a feasible workflow, we experimented with simulations bringing together different suggestions from the literature. We collected our example datasets, simulation functions and Bayesian models into the R-package JointSurvey, together with a supplementary vignette and manual (J Matthiopoulos et al., 2022). The package is instructive, rather than utilitarian (as we discuss below, spatial model fitting in optimised packages is less transparent, but much faster than our illustrative models). We refer to relevant sections of this vignette throughout our review and ensure that we designate sections and recommendations as either seabird-specific, or more general.

## PROCESS AND OBSERVATION MODELS

Species distribution data are the combined result of an underlying biological process and an observation procedure. Process models describe the occurrence of animals in space, while observation models describe the way that distribution data are collected. Two assumptions

allow data sets from multiple surveys to be combined. The first, is that the set of all seabird occurrences (whether observed or not) can be treated as a point process (see section 2.1 in supplementary vignette (J Matthiopoulos et al., 2022)). Underlying any spatial data set, we may envisage a heterogeneous surface that represents the rate of occurrence of seabirds. The actual locations of population members at any given time can then be thought of as a probabilistic realisation from this underlying *intensity surface*. This statistical formalism, known as the Inhomogeneous Point Process (IPP) has emerged (Aarts, Fieberg, & Matthiopoulos, 2012; Chakraborty et al., 2011; Fithian & Hastie, 2013; Fletcher et al., 2019; Isaac et al., 2020; Miller et al., 2019; Warton, Shepherd, & others, 2010) as a unifying framework that allows flexible modelling of covariates, spatiotemporal autocorrelation, observation effort, mechanistic details of ecology and behaviour, hierarchical grouping and transferability (Jason Matthiopoulos, Fieberg, & Aarts, 2020). In addition, it has been argued conclusively that the IPP encompasses all classical approaches to the analysis of animal distribution data (Aarts et al., 2012; Fithian & Hastie, 2013; Warton et al., 2010) and that widely-used spatial modelling packages such as MAXENT, are IPP models (Fithian & Hastie, 2013; Renner & Warton, 2013).

The second assumption is that all transect data can ultimately be analysed as if they came from strip transects (see section 3 in supplementary vignette). Strip transect methods assume that all individuals within a particular distance band are detected with equal probability, and no detections are recorded outside that band. Many historical seabird datasets come in this form (Jespersen, 1924; Tasker, Jones, Dixon, & Blake, 1984) and strip transects are still used for shipboard observations of birds in flight and aerial digital image capture (Oedekoven, Mackenzie, Scott-Hayward, & Rexstad, 2012). The alternative, continuous or punctuated records from line transect surveys (K. C. J. Camphuysen, Fox, Leopold, & Petersen, 2004; Oedekoven et al., 2012), are analysed by distance sampling methods (S.T.

Buckland et al., 2001; S T Buckland, Anderson, Burnham, & Laake, 2008), to effectively bring them into line with the assumptions of strip transect surveys.

By unifying the data-generating process across surveys and quantifying the effort implicit in different survey methods, it is possible to simply pool data from different surveys on a shared point process platform, allowing the model's offset term to account for effort.

## CHALLENGES AND OPPORTUNITIES IN MULTI-SURVEY MODELLING

The sections below impinge on data integration approaches for all species, but the section on accessibility and density dependence is particularly relevant to central place aggregations, such as seabird colonies.

### *Imperfect observations*

Biases and imprecisions inherent in different survey methodologies may be amplified by seabird behaviour. For example, the localised strong attraction of scavenging seabirds to boats (Bodey et al., 2014) can cause heterogeneities in detectability from on-board observers. Additionally, variation in observer ability to detect, identify and accurately count seabirds can be considerable (C. J. Camphuysen et al., 1995). False negatives and positives in detection will bias the intercept of the eventual SDM. For instance, if two species are easily mistaken for each other (e.g., red-throated and black-throated divers), then the bias will be positive (a large intercept) or negative (a small intercept) depending on whether the true prevalence of the focal species is respectively smaller or larger than the prevalence of the non-focal species. It is also possible for imperfect detection to affect the inferred relationships (i.e. slopes) with covariates (Lahoz-Monfort, Guillera-Arroita, & Wintle, 2014).

Errors in the detection and identification of individuals may vary temporally (Furnas, Newton, Capehart, & Barrows, 2019) or according to ambient conditions (Frair et al., 2010) making it impossible to use counts as a relative index of abundance (e.g. Oppel et al. 2012)



without a correction to the effective strip width and the baseline probability of detection, or much better, incorporation of these covariates in the distance-sampling analysis (S T Buckland et al., 2008). When both the probability of detection and the abundance of a species depend on the same environmental covariate, teasing apart its effects depends on whether repeat visits to the same location, by the same data collection method, occurred within a short time window (Guillera-Arroita & Lahoz-Monfort, 2012). For most situations, multiple concurrent visits to any location by the same observation platform are unlikely, and quantifying the magnitude and covariates of detection errors must rely on the species responding similarly to similar habitats surveyed at different places or times (Lele, Moreno, & Bayne, 2012) .

Heterogeneity in observation effort can generate many of the patterns seen in survey data, potentially confounding true biological processes. Hence, observation covariates may need to be included in the intensity function of an IPP (Chakraborty et al., 2011).

#### *Accessibility and density dependence: A feature of colonial species*

For at least some parts of the year, breeding adults and some non-breeding/immature seabirds are central-place foragers from the breeding colony. Colony accessibility shapes the distributions of colonial species (Aarts et al., 2008; Lewis, Sherratt, Hamer, & Wanless, 2001; Jason Matthiopoulos, 2003; Jason Matthiopoulos, Fieberg, Aarts, Barraquand, & Kendall, 2020; Thaxter et al., 2012; Waggitt et al., 2019; E. D. E. D. Wakefield et al., 2011), leading to potential resource depletion in the regions surrounding the colonies (Ashmole, 1963; Lewis et al., 2001). Ultimately, the use of particular locations at sea is a trade-off between commuting costs (including the risks of provisioning chicks) and foraging benefits (including depletion by conspecifics).

Both accessibility and depletion/interference may be thought of as functions of travel distance from the colony, but they are complex, non-linear processes for distinct reasons (see

Accepted Article

section 2.2. of supplementary vignette and abridged explanations in Box 1). Many seabird species avoid flying over land, may take circuitous routes dictated by wind conditions and avoid anthropogenic structures. It is important that these effects are accounted for in measures of distance. To capture declines in accessibility with distance, it is possible, as a first approximation, to introduce a distance-decay function, parameterised identically for different colonies (Jason Matthiopoulos, McConnell, Duck, & Fedak, 2004). However, the fact that the available area of water around each colony will depend on coastal morphology, means that the resulting marine distribution from such a function would not allocate equal numbers of birds at units of area that are the same distance from different colonies (E. D. Wakefield et al., 2017). These behaviours will also depend on age and breeding stage. Seabird populations include a high proportion of immatures that are less competitive or constrained compared to provisioning adults, so may tend to go further away from colonies (Fayet et al., 2015). So, we can expect some ‘infilling’ of marine areas away from large colonies by immatures, especially younger age classes (Ashmole, 1963; Votier et al., 2017).

Density dependence is primarily driven by resource competition between colony members (Lewis et al., 2001). As the size of the colony grows, individuals need to travel further to escape the density-dependent effects of depletion. Depletion from neighbouring colonies can also lead to the appearance of home ranging behaviour at the colony level (Aarts et al., 2021; E. D. Wakefield et al., 2013). In addition, seabirds may experience interspecific competition (Petalas, Lazarus, Lavoie, Elliott, & Guigueno, 2021) leading to asymmetries in colony domains, driven by relative colony sizes, trophic niche overlap and competitive dominance between species.

The sophistication that is used for modelling accessibility and density dependence will determine computational feasibility. A parsimonious approach (see Box 1 and further details and numerical examples in section 2.2. of supplementary vignette) must include the key

features of accessibility and competition (intra-colony, inter-colony and inter-specific). This approach allows the strength of accessibility and density dependence for a particular species to be quantified. Hence, fitting it simultaneously with environmental covariates should allow intrinsic and environmental regulation of spatial usage to be teased apart.

### *Autocorrelation*

A common assumption of SDMs is that their residuals (*conditional* on the covariates included) are independent (Dormann et al., 2007). Residual spatial autocorrelation can result from missing or mis-specified covariates, or from social clustering. Unmodelled heterogeneity generated by spatial autocorrelation can lead to overdispersion in the residuals. Seabird analyses, commonly capture this with overdispersed likelihoods (e.g. Lieske et al. 2014) or zero-inflated models (e.g. Oppel et al. 2012, Waggitt et al. 2019), making the unlikely assumption that nearby residuals are uncorrelated (see example model in section 5.1 of supplementary vignette). It is generally more informative to model spatial autocorrelation explicitly by using flexible functions of latitude and longitude (Mendel et al., 2019), using density as a local autocovariate (Augustin, Muggleston, & Buckland, 1996), or including an autocorrelated random effect within the error structure of a hierarchical model (Beale, Lennon, Yearsley, Brewer, & Elston, 2010). We present an illustration of the structure and application of such a model in section 5.2 of the supplementary vignette. It is important to note that none of these approaches can distinguish between extrinsically driven spatial autocorrelation (e.g., unmodelled environmental covariates) and intrinsically driven clustering (e.g., social aggregations at sea), both of which are likely to be features of seabird distributions.

Temporal dependence offers us valuable opportunities for exploiting multi-survey data that have been collected at different times. Currently, investigation of multiannual trends and relative changes in usage usually falls beyond the scope of seabird SDMs (Perrow et al.,

2015), but such features are essential to correctly integrate multiannual survey data. Counts from two surveys conducted over the same region and season in different years should be expected to be more similar, the closer the two years were. Within the range of spatiotemporal autocorrelation, a model should be able to acquire additional support from the fact that even when two surveys do not exactly coincide in time and space, they can share similar information depending on their spatiotemporal proximity (Hothorn, Müller, Schröder, Kneib, & Brandl, 2011). Therefore, for multi-survey SDMs, modelling spatial and temporal autocorrelation explicitly can be an asset (Fig. 2), within the spatiotemporal frame of the pooled survey data.

### *Model transferability*

Due to rapid anthropogenic change, ecologists are increasingly tasked with predicting outside the spatiotemporal frame of their data (Yates et al., 2018). An ideal dataset for empirical modelling in this context includes sampling effort spanning a wide range of covariate values, and combinations thereof (Oedekoven et al., 2012), criteria that can potentially be met cost-effectively by combining multiple surveys as distinct sampling instances in a functional response framework. Functional responses broadly describe how organisms change their use of a habitat as the availabilities of that, and all other habitats change and have been shown to bring considerable gains in predictive power for environmental scenarios that are within the range of environmental values observed in the pooled data (Holbrook et al., 2019; Jason Matthiopoulos, Hebblewhite, Aarts, & Fieberg, 2011; Paton & Matthiopoulos, 2018).

Extrapolations outside the observed spatiotemporal *and* environmental window (i.e. the environmental profiles used for model-training) are more problematic (Sinclair, White, & Newell, 2010) and yet form the main objective of anticipatory ecological modelling.

Arguably, increasing the mechanistic content of SDMs increases predictive ability. Hence,

there is now a clear tendency in the literature to consider species' distributions in the contexts of their population dynamics (Ehrlén & Morris, 2015; Jason Matthiopoulos et al., 2015; Mcloughlin, Morris, Fortin, Vander Wal, & Contasti, 2010) and wider ecological communities (Fleming et al., 2014; Ovaskainen, Abrego, Halme, & Dunson, 2016).

### *Computational efficiency*

Spatially autoregressive models, particularly combined with non-linear predictors, are computationally very expensive so approximations are necessary to make them tractable (e.g. the "covariate" model in Pacifici et al. (2017), or the list of four methods cited Chakraborty et al. (2011)). For seabirds, the key challenge lies in simultaneously estimating biologically important parameters (e.g. pertaining to density dependence) and accounting for spatial and temporal autocorrelation. Computation has been revolutionised by approximate Bayesian methods that either deal with fully non-linear models (as in Approximate Bayesian Computation - (Beaumont, 2010) or deal with linearised versions of these models (as in Integrated Nested Laplace Approximation - INLA - (Bachl, Lindgren, Borchers, & Illian, 2019; Rue, Martino, & Chopin, 2009). ABC methods have yet to meet with broad application in SDMs, but INLA methodology, developed specifically for IPPs, is ideal for the purposes of SDMs, particularly since it may soon be possible to fit mildly non-linear models (such as those of Box 1 and the model "spatial" in the supplementary vignette) within the inlabru package (Bachl et al., 2019).

## BEST PRACTICE FOR MULTI-SURVEY ANALYSES OF SEABIRD DISTRIBUTIONS

Building upon the above literature, and practical experimentation with simulated data (Supplements), we make the following practical recommendations for multi-survey analyses. Recommendations under category 4, below, are particularly relevant to seabirds.

### *1) Response and explanatory variables*

**Keep data in their highest-information form:** Thresholding abundance data into occupancy represents considerable information loss and precludes predictions of spatial distribution (yielding instead, surfaces for the probability of presence). If individual detections are available, we should use these in preference to spatially aggregated counts.

**Analyse even low-information data as if originating from abundance:** Irrespective of how detailed our species observations are, the underlying biological variable is the abundance of a species. Hence, although some data sets solely record occupancy, we should still model the underlying data-generating process as an intensity surface. Such approaches will work particularly well if some abundance records (from different surveys) are also integrated into the analysis (Fletcher et al., 2019).

**Avoid inflated error structures until the end of modelling:** Zero-inflated and over-dispersed data are the norm in spatial ecology. However, covariates will generally explain some of that variability, and use of spatially and temporally autocorrelated fixed or random effects will better describe unexplained high or low density regions in the underlying seabird distribution.

## *2) Treatment of survey design attributes and observation errors*

**Prioritise cross-calibration between surveys:** Surveys for which the detectability errors have been quantified (e.g., multiple observer platforms), are highly desirable because, within a joint analysis, they can cross-calibrate other, less detailed surveys. The overlap required for such cross-calibration may not need to be exact or complete, in explicitly spatiotemporal models (i.e. models that can recognise temporal and spatial proximity between observations via autocorrelated structures, see below).

## *3) Treatment of space time*

**Use point process models:** IPP approaches subsume all valid approaches to species distribution modelling and are fast becoming the benchmark for spatiotemporal analyses

Their implementation in speed-optimised frameworks such as INLA favours their use in wildlife management, which often requires time-limited decision making.

**Use autocorrelated structures:** Spatially and temporally autocorrelated error terms can account for wholly or partly missing covariates (hence explaining residual overdispersion). In multi-survey data, they can be used to exploit spatiotemporal proximity between observations so that, even if exact replication is not part of the survey design, an indirect form of replication can be achieved.

**Take complex dynamics into account:** If we need to account for multi-survey data that include before-and-after control impact, it is important to account for temporal trends. In some cases, non-linearity in the responses of a species can be captured by simple extensions such as statistical interaction terms in the linear predictors of models. In other cases, a more explicitly biological model or autoregressive error structures may be required.

#### *4) Accessibility and density dependence*

**Use biologically relevant travel distance measures:** For colonial species, accessibility and density dependence in spatial usage are most often represented as non-linear transformations of distance between points at sea and colonies. Therefore, using distance measures that avoid obstacles is essential if birds do not transit between locations in straight lines.

**For the present, use simplified models of density dependence:** Currently, the computational demands of a fully spatially explicit model of intra-colony, inter-colony and interspecific competition are prohibitive for the purposes of applied SDMs. Models such as those of Box 1 may be crude approximations of the truth, but even such relatively simple formulations are currently missing from most seabird SDM approaches.

**In the future, consider spatially explicit models for density dependence:** As computational approaches (particularly ABC and INLA) become more widespread in the field of SDMs, it may become possible to model competition in an explicitly spatial way. Modelling multiple, coupled response variables would allow the spatial interactions of different colonies to be captured as part of simultaneous regression where the animals from any given colony are allowed to affect (and be affected by) the distributions of members of other colonies and species.

### *5) Statistical Frameworks*

**Use hierarchical models:** Three important features of multi-survey models described above rely on hierarchical models: cross-calibration of observation models, covariate imputation and latency and use of spatiotemporal proximity to allow the predictions to borrow strength from multiple surveys.

**Use Bayesian approaches:** Computer-intensive Bayesian model-fitting is implemented in flexible software frameworks (such as JAGS or Stan), that allow state-space and hierarchical structures. Bayesian inference permits the elicitation of expert opinion in the form of parameter priors. INLA is Bayesian by design, so attention could be given to expert specification of priors in its application.

**Fully propagate uncertainty to the final predictions:** Currently, the majority of published marine SDM studies do not report the amount of uncertainty derived from data deficiencies and model parameters (94% of papers reviewed by Robinson et al. 2017). Above, we have reviewed several areas where this could lead to unnecessarily optimistic or conservative conclusions. For example, unmodelled spatial and temporal autocorrelation in the data may artificially inflate apparent sample sizes. Such concerns about pseudoreplication apply particularly for multi-survey analyses because different surveys may have overlapped in



space or in time. Alternatively, uncertainty contained in the pre-analysis of transects, if not propagated to the final results, may under-represent uncertainty in distribution.

## FUTURE EXTENSIONS

All the challenges and opportunities discussed above apply beyond the strict assumptions of transect surveys, into other types of seabird distribution data. The key difficulty to extending integration between other types of data is in formulating appropriate observation models (e.g., to take account of observation effort in citizen science protocols) and joining together very different likelihood models (e.g., step selection functions for telemetry data with resource selection functions for tracking data).

The different methodologies of data collection might, at first sight, imply an overwhelming number of data integration frameworks (Fig. 3a). However, from a statistical point of view, the simplifying assumptions discussed above on process and observation models lead to a considerably simpler picture (Fig. 3b), essentially requiring the integration between different types of surveys (e.g., line, point transect, citizen science data) and qualitatively different data types (e.g., multi-species data, telemetry, and mark-recapture). We review developments in these areas below.

### *Multi-species surveys*

Surveys at sea usually record multiple species. This is an alternative interpretation of the multi-survey idea, where the same platform provides multiple datasets. Interest in hotspots of biodiversity has led to the idea of stacking single-species SDMs (Calabrese, Certain, Kraan, & Dormann, 2014). Although stacking is not an integrated analysis in the sense outlined here, it has been useful in demonstrating the magnitude and duration of seabird aggregations or partitioning in the open sea from both survey (Nur et al., 2011; E. D. Wakefield et al., 2021) and tracking (Davies et al., 2021; Grecian et al., 2016; Jones et al., 2015)

Accepted Article

data. An interesting research direction lies in allowing data sets from multiple species to gain strength from each other. We outlined earlier how spatiotemporal proximity can be used to borrow strength by jointly analysing a collection of surveys. The same idea could be extended to develop hierarchical models using taxonomic or functional proximity (Kindsvater et al., 2018). Multispecies SDMs could be developed to quantify the (apparent) associations between species (Ovaskainen et al., 2016; Tikhonov et al., 2020), and these could be used to reconstruct the distribution for any-and-all of the species participating in the model. By exploiting niche differentiation between species, this approach also has potential in modelling unknown observation effort (Peel et al., 2019) because an apparent absence of a species in a region can be better diagnosed as a true absence if it is known that we have recorded other species there.

#### *Combination with vantage point data*

Several data collection methods could come under this category, most commonly, observations made from onshore stations (e.g., by theodolite/distancer). These could be important sources of information for near-shore distribution. Their combination with line transect survey data is straightforward since both data types belong to the broader class of transect methods (S.T. Buckland et al., 2001). Nesting habitat preferences for seabirds are a considerably less studied aspect of their biology, but one that is particularly pertinent for determining the placement of potential new colonies. Of particular relevance for studying human-seabird interactions is the terrestrial distribution of scavenging species such as gulls, which frequently switch between marine and terrestrial foraging.

#### *Combination with citizen science data*

Citizen science programmes are flourishing in ecology and new statistical methods are being developed to deal with the resulting data (Bird et al., 2014), which are often of variable quality (Hochachka et al., 2012). The main issue with citizen scientist data is that often we

have limited information on the spatial and temporal distribution of survey effort and the heterogeneity in bias or imprecision in species identification across individual observers (Dickinson, Zuckerberg, & Bonter, 2010). Such gaps in knowledge often need to be supplemented by proxies (such as plausible assumptions about the behaviour and distribution of citizen observers).

### *Combination with telemetry data*

The combination of survey data with tracking data has proved particularly challenging.

Studies that have attempted this marriage in the seabird literature have often inflicted heavy censoring on the data (Louzao et al., 2009) or taken a comparative (rather than integrative) approach (e.g. Carroll et al. 2019). The default analysis frameworks used for each data type are a major obstacle to joint inference. Telemetry data are most conveniently analysed via step selection functions (SSFs), while resource selection functions (RSFs) are most appropriate for survey data. These two analytical approaches do not, by default, lead to the same results (Signer, Fieberg, & Avgar, 2017). A promising development in this area is the convergence between the frameworks of resource selection and step selection analyses (Michélot, Blackwell, & Matthiopoulos, 2019). This work has established the conditions under which SSF and RSF frameworks agree, and has begun to derive methods for joint inference (Michélot, Blackwell, Chamaillé-Jammes, & Matthiopoulos, 2019).

### *Combination with mark-recapture data*

Mark-recapture data have rarely been used to map seabird distributions and fit habitat models (K. C. J. Camphuysen et al., 2004), however they are a potentially valuable repository of spatial data that are also individually referenced (e.g. colour-ringing). In a sense therefore, mark-recapture data carry intermediate information between point transects and telemetry tracking and could, in the longer-term benefit from current developments in the integration between these two.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## AUTHORS' CONTRIBUTIONS

JM, TE, JB formulated the focus of the review. JM wrote the first draft, JointSurvey R library and accompanying vignette. MT collated data for the simulations. EW, JJ, RWF, MT, GT, AMcC, SA, JB & TE provided biological grounding for the modelling synthesis. All authors contributed to the structure of the review, literature selection and drafting of the final text.

## DATA AVAILABILITY STATEMENT

This paper is accompanied by the R-library JointSurvey and an explanatory vignette available via <https://github.com/JasonMat/SeabirdMultiplatform> (Matthiopoulos et al, 2022)

## ACKNOWLEDGEMENTS

This review was supported by grants TG-0247 and MS-P-0020 by Marine Scotland.

## REFERENCES

- Aarts, G., Fieberg, J., & Matthiopoulos, J. (2012). Comparative interpretation of count, presence-absence and point methods for species distribution models. *Methods in Ecology and Evolution*, 3(1), 177–187. doi: 10.1111/j.2041-210X.2011.00141.x
- Aarts, G., MacKenzie, M., McConnell, B., Fedak, M., & Matthiopoulos, J. (2008). Estimating space-use and habitat preference from wildlife telemetry data. *Ecography*, 31(1), 140–160. doi: 10.1111/j.2007.0906-7590.05236.x
- Aarts, G., Mul, E., Fieberg, J., Brasseur, S., van Gils, J. A., Matthiopoulos, J., & Riotte-Lambert, L. (2021). Individual-level memory is sufficient to create spatial segregation among neighboring colonies of central place foragers. *American Naturalist*, 198(2), E37–E52. doi: 10.1086/715014
- Ashmole, N. P. (1963). the Regulation of Numbers of Tropical Oceanic Birds. *Ibis*, 103 b(3), 458–473. doi: 10.1111/j.1474-919X.1963.tb06766.x
- Augustin, N. H., Muggleston, M. A., & Buckland, S. T. (1996). An autologistic model for the spatial distribution of wildlife. *Journal of Applied Ecology*, 33, 339–347.
- Bachl, F. E., Lindgren, F., Borchers, D. L., & Illian, J. B. (2019). inlabru: an R package for Bayesian spatial modelling from ecological survey data. *Methods in Ecology and Evolution*, 10(6), 760–766. doi: 10.1111/2041-210X.13168
- Beale, C. M., Lennon, J. J., Yearsley, J. M., Brewer, M. J., & Elston, D. A. (2010). Regression analysis of spatial data. *Ecology Letters*, 13(2), 246–264. doi: 10.1111/j.1461-0248.2009.01422.x
- Beaumont, M. A. (2010). Approximate Bayesian Computation in Evolution and Ecology. *Annual Review of Ecology, Evolution, and Systematics*, 41(1), 379–406. doi:

- Bird, T. J., Bates, A. E., Lefcheck, J. S., Hill, N. A., Thomson, R. J., Edgar, G. J., ... Frusher, S. (2014). Statistical solutions for error and bias in global citizen science datasets. *Biological Conservation*, 173, 144–154. doi: 10.1016/j.biocon.2013.07.037
- Bodey, T. W., Jessopp, M. J., Votier, S. C., Gerritsen, H. D., Cleasby, I. R., Hamer, K. C., ... Bearhop, S. (2014). Seabird movement reveals the ecological footprint of fishing vessels. *Current Biology*, 24(11), R514–R515. doi: 10.1016/j.cub.2014.04.041
- Buckland, S.T., Anderson, D. R., Burnham, K. ., Laake, J. L., Borchers, D., & Thomas, L. (2001). *Introduction to Distance Sampling: Estimating Abundance of Biological Populations*. Oxford University Press.
- Buckland, S T, Anderson, D. R., Burnham, K. P., & Laake, J. L. (2008). *Advanced distance sampling*.
- Buckland, Stephen T., Burt, M. L., Rexstad, E. A., Mellor, M., Williams, A. E., & Woodward, R. (2012). Aerial surveys of seabirds: The advent of digital methods. *Journal of Applied Ecology*, 49(4), 960–967. doi: 10.1111/j.1365-2664.2012.02150.x
- Calabrese, J. M., Certain, G., Kraan, C., & Dormann, C. F. (2014). Stacking species distribution models and adjusting bias by linking them to macroecological models. *Global Ecology and Biogeography*, 23(1), 99–112. doi: 10.1111/geb.12102
- Camphuysen, C. J., Calvo, B., Durinck, J., Ensor, K., Follestad, A., Furness, R. W., ... Winter, C. J. N. (1995). *Consumption of discards by seabirds in the North Sea*.
- Camphuysen, K. C. J., Fox, T. A. D., Leopold, M. M. F., & Petersen, I. K. (2004). Towards standardised seabirds at sea census techniques in connection with environmental impact assessments for offshore wind farms in the U.K. *Cowrie - Bam - 02-2002*, (June),

1-38. doi: 10.13140/RG.2.1.2230.0244

Carroll, M. J., Wakefield, E. D., Scragg, E. S., Owen, E., Pinder, S., Bolton, M., ... Evans, P. G.

H. (2019). Matches and mismatches between seabird distributions estimated from at-sea surveys and concurrent individual-level tracking. *Frontiers in Ecology and Evolution*, 7(SEP). doi: 10.3389/fevo.2019.00333

Chakraborty, A., Gelfand, A. E., Wilson, A. M., Latimer, A. M., Silander, J. A., Journal, S., ...

Chakraborty, A. (2011). Point pattern modelling for degraded presence-only data over large regions. *Journal of the Royal Statistical Society. Series C: Applied Statistics*, 60(5), 757–776. Retrieved from <https://www.jstor.org/stable/41262305>

Davies, T. E., Carneiro, A. P. B., Tarzia, M., Wakefield, E., Hennicke, J. C., Frederiksen, M.,

... Dias, M. P. (2021). Multispecies tracking reveals a major seabird hotspot in the North Atlantic. *Conservation Letters*, 14(5), 1–14. doi: 10.1111/conl.12824

Dickinson, J. L., Zuckerberg, B., & Bonter, D. N. (2010). Citizen Science as an Ecological

Research Tool: Challenges and Benefits. *Annual Review of Ecology, Evolution, and Systematics*, 41(1), 149–172. doi: 10.1146/annurev-ecolsys-102209-144636

Dormann, C. F., McPherson, J. M., Araújo, Miguel, B., Bivand, R., Bolliger, J., Carl, G., ...

Wilson, R. (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, 30(5), 609–628. doi: 10.1111/j.2007.0906-7590.05171.x

Egevang, C., Stenhouse, I. J., Phillips, R. A., Petersen, A., Fox, J. W., & Silk, J. R. D. (2010).

Tracking of Arctic terns *Sterna paradisaea* reveals longest animal migration. *Proceedings of the National Academy of Sciences of the United States of America*, 107(5), 2078–2081. doi: 10.1073/pnas.0909493107

Ehrlén, J., & Morris, W. F. (2015). Predicting changes in the distribution and abundance of species under environmental change. *Ecology Letters*, 18(3), 303–314. doi: 10.1111/ele.12410

Elliott, K. H., Ricklefs, R. E., Gaston, A. J., Hatch, S. A., Speakman, J. R., & Davoren, G. K. (2013). High flight costs, but low dive costs, in auks support the biomechanical hypothesis for flightlessness in penguins. *Proceedings of the National Academy of Sciences of the United States of America*, 110(23), 9380–9384. doi: 10.1073/pnas.1304838110

Fayet, A. L., Freeman, R., Shoji, A., Padget, O., Perrins, C. M., & Guilford, T. (2015). Lower foraging efficiency in immatures drives spatial segregation with breeding adults in a long-lived pelagic seabird. *Animal Behaviour*, 110, 79–89. doi: 10.1016/j.anbehav.2015.09.008

Fithian, W., & Hastie, T. (2013). Finite-sample equivalence in statistical models for presence-only data. *The Annals of Applied Statistics*, 7(4), 1917. doi: 10.1214/13-AOAS667

Fleming, C. H., Calabrese, J. M., Mueller, T., Olson, K. A., Leimgruber, P., & Fagan, W. F. (2014). Non-Markovian maximum likelihood estimation of autocorrelated movement processes. *Methods in Ecology and Evolution*, 5(5), 462–472.

Fletcher, R. J., Hefley, T. J., Robertson, E. P., Zuckerberg, B., McCleery, R. A., & Dorazio, R. M. (2019). A practical guide for combining data to model species distributions. *Ecology*, 0(0), e02710. doi: 10.1002/ecy.2710

Fletcher, R. J., McCleery, R. A., Greene, D. U., & Tye, C. A. (2016). Integrated models that unite local and regional data reveal larger-scale environmental relationships and improve predictions of species distributions. *Landscape Ecology*, 31(6), 1369–1382. doi: 10.1007/s10980-015-0327-9



Frair, J. L., Fieberg, J., Hebblewhite, M., Cagnacci, F., DeCesare, N. J., & Pedrotti, L. (2010).

Resolving issues of imprecise and habitat-biased locations in ecological analyses using {GPS} telemetry data. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 365(1550), 2187–2200.

Furnas, B. J., Newton, D. S., Capehart, G. D., & Barrows, C. W. (2019). Hierarchical distance sampling to estimate population sizes of common lizards across a desert ecoregion.

*Ecology and Evolution*, 9(6), 3046–3058. doi: 10.1002/ece3.4780

Grecian, W. J. J., Witt, M. J. J., Attrill, M. J. J., Bearhop, S., Becker, P. H. H., Egevang, C., ...

Votier, S. C. C. (2016). Seabird diversity hotspot linked to ocean productivity in the Canary Current Large Marine Ecosystem. *Biology Letters*, 12(8). doi:

10.1098/rsbl.2016.0024

Guillera-Arroita, G., & Lahoz-Monfort, J. J. (2012). Designing studies to detect differences in species occupancy: Power analysis under imperfect detection. *Methods in Ecology and*

*Evolution*, 3(5), 860–869. doi: 10.1111/j.2041-210X.2012.00225.x

Hochachka, W. M., Fink, D., Hutchinson, R. A., Sheldon, D., Wong, W. K., & Kelling, S.

(2012). Data-intensive science applied to broad-scale citizen science. *Trends in Ecology and Evolution*, 27(2), 130–137. doi: 10.1016/j.tree.2011.11.006

Holbrook, J. D., Olson, L. E., DeCesare, N. J., Hebblewhite, M., Squires, J. R., & Steenweg, R.

(2019). Functional responses in habitat selection: clarifying hypotheses and interpretations. *Ecological Applications*, 29(3), e01852. doi: 10.1002/eap.1852

Hothorn, T., Müller, J., Schröder, B., Kneib, T., & Brandl, R. (2011). Decomposing

environmental, spatial, and spatiotemporal components of species distributions.

*Ecological Monographs*, 81(2), 329–347. doi: 10.1890/10-0602.1

Isaac, N. J. B., Jarzyna, M. A., Keil, P., Dambly, L. I., Boersch-Supan, P. H., Browning, E., ...

O'Hara, R. B. (2020). Data Integration for Large-Scale Models of Species Distributions.

*Trends in Ecology and Evolution*, 35(1), 56–67. doi: 10.1016/j.tree.2019.08.006

Jespersen, P. (1924). The frequency of birds over the high atlantic ocean. *Nature*, 114(2860),

281–283. doi: 10.1038/114281a0

Jones, E. L. L., McConnell, B. J. J., Smout, S., Hammond, P. S. S., Duck, C. D. D., Morris, C. D.

D., ... Matthiopoulos, J. (2015). Patterns of space use in sympatric marine colonial predators reveal scales of spatial partitioning. *Marine Ecology Progress Series*, 534. doi:

10.3354/meps11370

Kindsvater, H. K. H. K., Dulvy, N. K. N. K., Horswill, C., Juan-Jordá, M. J. M.-J. M. J. M.-J.,

Mangel, M., & Matthiopoulos, J. (2018). Overcoming the Data Crisis in Biodiversity

Conservation. *Trends in Ecology and Evolution*, 33(9), 676–688. doi:

10.1016/j.tree.2018.06.004

Lahoz-Monfort, J. J., Guillera-Arroita, G., & Wintle, B. A. (2014). Imperfect detection impacts

the performance of species distribution models. *Global Ecology and Biogeography*, 23(4),

504–515. doi: 10.1111/geb.12138

Lele, S. R., Moreno, M., & Bayne, E. (2012). Dealing with detection error in site occupancy

surveys: What can we do with a single survey? *Journal of Plant Ecology*, 5(1), 22–31. doi:

10.1093/jpe/rtr042

Lewis, S., Sherratt, T. N., Hamer, K. C., & Wanless, S. (2001). Evidence of intra-specific

competition for food in a pelagic seabird. *Nature*, 412(August), 816–819.

Lieske, D. J., Fifield, D. A., & Gjerdrum, C. (2014). Maps, models, and marine vulnerability:

Assessing the community distribution of seabirds at-sea. *Biological Conservation*, 172,

15–28. doi: 10.1016/j.biocon.2014.02.010

Louzao, M., Bécarea, J., Rodríguez, B., Hyrenbach, K. D., Ruiz, A., & Arcos, J. M. (2009).

Combining vessel-based surveys and tracking data to identify key marine areas for seabirds. *Marine Ecology Progress Series*, 391, 183–197. doi: 10.3354/meps08124

Matthiopoulos, J., Wakefield, E., Jeglinski, J. W. E., Furness, R. W., Trinder, M., Tyler, G., ...

Evans, T. (2022). Data from: Software for Integrated modelling of seabird-habitat associations from multi-platform data: a review. Github  
<https://github.com/JasonMat/SeabirdMultiplatform>

Matthiopoulos, Jason. (2003). The use of space by animals as a function of accessibility and preference. *Ecological Modelling*, 159(2–3), 239–268. doi: 10.1016/S0304-3800(02)00293-4

Matthiopoulos, Jason, Fieberg, J., & Aarts, G. (2020). *Species-Habitat Associations: Spatial data, predictive models, and ecological insights*. University of Minnesota Libraries Publishing.  
Retrieved from the University of Minnesota Digital Conservancy,  
<https://hdl.handle.net/11299/217469>.

Matthiopoulos, Jason, Fieberg, J., Aarts, G., Barraquand, F., & Kendall, B. E. (2020). Within reach? Habitat availability as a function of individual mobility and spatial structuring. *The American Naturalist*, 195(6), 1009–1026. doi: 10.1086/708519

Matthiopoulos, Jason, Fieberg, J., Aarts, G., Beyer, H. L., Morales, J. M., & Haydon, D. T. (2015). Establishing the link between habitat selection and animal population dynamics. *Ecological Monographs*, 85(3), 413–436. doi: 10.1890/14-2244.1

Matthiopoulos, Jason, Hebblewhite, M., Aarts, G., & Fieberg, J. (2011). Generalized functional responses for species distributions. *Ecology*, 92(3), 583–589. doi: 10.1890/10-0751.1

Matthiopoulos, Jason, McConnell, B., Duck, C., & Fedak, M. (2004). Using satellite telemetry and aerial counts to estimate space use by grey seals around the British Isles. *Journal of Applied Ecology*, 41(3), 476–491. doi: 10.1111/j.0021-8901.2004.00911.x

McCloughlin, P. D., Morris, D. W., Fortin, D., Vander Wal, E., & Contasti, A. L. (2010). Considering ecological dynamics in resource selection functions. *Journal of Animal Ecology*, 79(1), 4–12. doi: 10.1111/j.1365-2656.2009.01613.x

Mendel, B., Schwemmer, P., Peschko, V., Müller, S., Schwemmer, H., Mercker, M., & Garthe, S. (2019). Operational offshore wind farms and associated ship traffic cause profound changes in distribution patterns of Loons (*Gavia* spp.). *Journal of Environmental Management*, 231(October 2018), 429–438. doi: 10.1016/j.jenvman.2018.10.053

Michelot, T., Blackwell, P. G., Chamaillé-Jammes, S., & Matthiopoulos, J. (2019). Inference in MCMC step selection models. *Biometrics*, 1(October 2018), 1–10. doi: 10.1111/biom.13170

Michelot, T., Blackwell, P. G., & Matthiopoulos, J. (2019). Linking resource selection and step selection models for habitat preferences in animals. *Ecology*, 100(1), 1–12. doi: 10.1002/ecy.2452

Miller, D. A. W. W., Pacifici, K., Sanderlin, J. S., & Reich, B. J. (2019). The recent past and promising future for data integration methods to estimate species' distributions. *Methods in Ecology and Evolution*, 10(1), 22–37. doi: 10.1111/2041-210X.13110

Nelli, L., Ferguson, H. M., & Matthiopoulos, J. (2019). Achieving explanatory depth and spatial breadth in infectious disease modelling: Integrating active and passive case surveillance. *Statistical Methods in Medical Research*, 29(5), 096228021985638. doi: 10.1177/0962280219856380

Nur, N., Jahncke, J., Herzog, M. P., Howar, J., Hyrenbach, K. D., Zamon, J. E., ... Stralberg, D. (2011). *Where the wild things are : predicting hotspots of seabird aggregations in the California Current System*. 21(6), 2241–2257.

Oedekoven, C. S., Mackenzie, M. L., Scott-Hayward, L. A. S., & Rexstad, E. (2012). *Marine Scotland Science Report 05/14 Statistical Modelling of Seabird and Cetacean Data: Literature Review*. Retrieved from <http://www.scotland.gov.uk/Topics/marine/Publications/publicationslatest/Science/MSSR/2014/0514>

Oppel, S., Meirinho, A., Ramírez, I., Gardner, B., O'Connell, A. F., Miller, P. I., & Louzao, M. (2012). Comparison of five modelling techniques to predict the spatial distribution and abundance of seabirds. *Biological Conservation*, 156, 94–104. doi: 10.1016/j.biocon.2011.11.013

Ovaskainen, O., Abrego, N., Halme, P., & Dunson, D. (2016). Using latent variable models to identify large networks of species-to-species associations at different spatial scales. *Methods in Ecology and Evolution*, 7(5), 549–555. doi: 10.1111/2041-210X.12501

Pacifici, K., Reich, B. J., Miller, D. A. W. W., Gardner, B., Stauffer, G., Singh, S., ... Collazo, J. A. (2017). Integrating multiple data sources in species distribution modeling: A framework for data fusion. *Ecology*, 98(3), 840–850. doi: 10.1002/ecy.1710

Paton, R. S., & Matthiopoulos, J. (2018). Defining the scale of habitat availability for models of habitat selection. *Ecology*, 97(July), 1113–1122. doi: 10.1002/ecy.2446

Peel, S. L., Hill, N. A., Foster, S. D., Wotherspoon, S. J., Ghiglione, C., & Schiaparelli, S. (2019). Reliable species distributions are obtainable with sparse, patchy and biased data by leveraging over species and data types. *Methods in Ecology and Evolution*, 10(7), 1002–

1014. doi: 10.1111/2041-210X.13196

Perrow, M. R., Harwood, A. J. P., Skeate, E. R., Praca, E., & Eglington, S. M. (2015). Use of multiple data sources and analytical approaches to derive a marine protected area for a breeding seabird. *Biological Conservation*, 191, 729–738. doi: 10.1016/j.biocon.2015.08.031

Petalas, C., Lazarus, T., Lavoie, R. A., Elliott, K. H., & Guigueno, M. F. (2021). Foraging niche partitioning in sympatric seabird populations. *Scientific Reports*, 11(1), 1–12. doi: 10.1038/s41598-021-81583-z

Pittman, S., Yates, K., Bouchet, P., Alvarez-Berastegui, D., Andréfouët, S., Bell, S., ... Young, M. (2021). Seascape ecology: identifying research priorities for an emerging ocean sustainability science. *Marine Ecology Progress Series*, 663, 1–29. doi: 10.3354/meps13661

Renner, I. W., & Warton, D. I. (2013). Equivalence of MAXENT and Poisson point process models for species distribution modeling in ecology. *Biometrics*, 69(1), 274–281.

Robinson, N. M., Nelson, W. A., Costello, M. J., Sutherland, J. E., & Lundquist, C. J. (2017). A systematic review of marine-based Species Distribution Models (SDMs) with recommendations for best practice. *Frontiers in Marine Science*, 4(DEC), 1–11. doi: 10.3389/fmars.2017.00421

Rue, H., Martino, S., & Chopin, N. (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(2), 319–392. doi: 10.1111/j.1467-9868.2008.00700.x

Sansom, A., Wilson, L. J., Caldow, R. W. G., & Bolton, M. (2018). Comparing marine distribution maps for seabirds during the breeding season derived from different survey and analysis methods. *PLoS ONE*, 13(8), 1–17. doi: 10.1371/journal.pone.0201797

- Signer, J., Fieberg, J., & Avgar, T. (2017). Estimating utilization distributions from fitted step-selection functions. *Ecosphere*, 8(4). doi: 10.1002/ecs2.1771
- Sinclair, S. J., White, M. D., & Newell, G. R. (2010). How useful are species distribution models for managing biodiversity under future climates? *Ecology and Society*, 15(1). doi: 10.5751/ES-03089-150108
- Steiner, U. K., & Gaston, A. J. (2005). Reproductive consequences of natal dispersal in a highly philopatric seabird. *Behavioral Ecology*, 16(3), 634–639. doi: 10.1093/beheco/ari035
- Tasker, M. L., Jones, P. H., Dixon, T., & Blake, B. F. (1984). Counting Seabirds at Sea from Ships: A Review of Methods Employed and a Suggestion for a Standardized Approach. *The Auk*, 101(3), 567–577. doi: 10.1093/auk/101.3.567
- Thaxter, C. B., Lascelles, B., Sugar, K., Cook, A. S. C. P., Roos, S., Bolton, M., ... Burton, N. H. K. (2012). Seabird foraging ranges as a preliminary tool for identifying candidate Marine Protected Areas. *Biological Conservation*, 156, 53–61. doi: 10.1016/j.biocon.2011.12.009
- Tikhonov, G., Opedal, Ø. H., Abrego, N., Lehikoinen, A., de Jonge, M. M. J., Oksanen, J., & Ovaskainen, O. (2020). Joint species distribution modelling with the r-package Hmsc. *Methods in Ecology and Evolution*, 11(3), 442–447. doi: 10.1111/2041-210X.13345
- Votier, S. C., Fayet, A. L., Bearhop, S., Bodey, T. W., Clark, B. L., Grecian, J., ... Patrick, S. C. (2017). Effects of age and reproductive status on individual foraging site fidelity in a long-lived marine predator. *Proceedings of the Royal Society B: Biological Sciences*, 284(1859), 0–6. doi: 10.1098/rspb.2017.1068
- Waggitt, J. J., Evans, P. G. H., Andrade, J., Banks, A. ., Boisseau, O., Bolton, M., ... Hiddink, J.

G. (2019). Distribution maps of cetacean and seabird populations in the North- East Atlantic. *Journal of Applied Ecology*, 205(4969), 1365-2664.13525. doi: 10.1111/1365-2664.13525

Wakefield, E. D., Bodey, T. W., Bearhop, S., Blackburn, J., Colhoun, K., Davies, R., ... Hamer, K. C. (2013). Space partitioning without territoriality in gannets. *Science*, 341(6141), 68–70. doi: 10.1126/science.1236077

Wakefield, E. D. E. D., Phillips, R. A. R. A., Trathan, P. N. P. N., Arata, J., Gales, R., Huin, N., ... Matthiopoulos, J. (2011). Habitat preference, accessibility, and competition limit the global distribution of breeding Black-browed Albatrosses. *Ecological Monographs*, 81(1), 141–167. doi: 10.1890/09-0763.1

Wakefield, E. D., Miller, D. L., Bond, S. L., le Bouard, F., Carvalho, P. C., Catry, P., ... Matthiopoulos, J. (2021). The summer distribution, habitat associations and abundance of seabirds in the sub-polar frontal zone of the Northwest Atlantic. *Progress in Oceanography*, 198(April). doi: 10.1016/j.pocean.2021.102657

Wakefield, E. D., Owen, E., Baer, J., Carroll, M. J., Daunt, F., Dodd, S. G., ... Bolton, M. (2017). Breeding density, fine-scale tracking, and large-scale modeling reveal the regional distribution of four seabird species: *Ecological Applications*, 27(7), 2074–2091. doi: 10.1002/eap.1591

Warton, D. I., Shepherd, L. C., & others. (2010). Poisson point process models solve the “pseudo-absence problem” for presence-only data in ecology.” *The Annals of Applied Statistics*, 4(3), 1383–1402.

Weimerskirch, H., Louzao, M., De Grissac, S., & Delord, K. (2012). Changes in wind pattern alter albatross distribution and life-history traits. *Science*, 335(6065), 211–214. doi:



10.1126/science.1210270

Yates, K. L., Bouchet, P. J., Caley, M. J., Mengersen, K., Randin, C. F., Parnell, S., ... Sequeira, A. M. M. (2018). Outstanding Challenges in the Transferability of Ecological Models. *Trends in Ecology and Evolution*, 33(10), 790–802. doi: 10.1016/j.tree.2018.08.001

Zeller, K. A., Vickers, T. W., Ernest, H. B., & Boyce, W. M. (2017). Multi-level, multi-scale resource selection functions and resistance surfaces for conservation planning: Pumas as a case study. *PLoS ONE*, 12(6), 1–20. doi: 10.1371/journal.pone.0179570

## BOX 1: SIMPLE FORMULATIONS OF SEABIRD COLONIALITY FOR STATISTICAL MODELS OF SPACE USE

Different features of seabird biology can be incorporated into mathematically simple expressions of incremental sophistication for modelling the expected intensity of use  $\lambda(\mathbf{s})$  of a point at sea. We aim to keep the number of parameters and levels of nonlinearity low, so that model fitting is feasible.

**Accessibility** of any point at sea  $\mathbf{s}$ , from the  $i^{\text{th}}$  colony located at a point  $\mathbf{s}_i$  may be treated as a function of distance  $d_i(\mathbf{s}) = |\mathbf{s}_i - \mathbf{s}|$ . Euclidean distance may be used, or distance measures based on at-sea travel (Jason Matthiopoulos, 2003). Unlike other model covariates, these distances should be calculated as spatial layers, specific to each colony and could be replaced by more elaborate measures of accessibility, incorporating landscape resistance (e.g. due to prevailing wind fields)(Zeller, Vickers, Ernest, & Boyce, 2017). We can use the following simple model, closely linked to log-linear regression.

$$\lambda(\mathbf{s}) = \exp(\alpha_0 - c_0 d_i(\mathbf{s})) \quad (1)$$

Graphs of this function for different parameterisations ranging from unconstrained (dark green) to highly constrained (yellow to white) ranges of flight are shown in Fig. 1a. Note that the estimated value of  $c_0$  will depend on the openness afforded to each colony by the surrounding coastline (small island colonies will have a higher value than colonies at inlets).

**Intra-colony competition** may be modelled by making the decline of  $\lambda(\mathbf{s})$  with distance slower for larger colonies. Building on eq. (1), we may introduce an interaction term between distance

and colony size ( $N_i$ ) (Fig. 1b)

$$\lambda(\mathbf{s}) = \exp(\alpha_0 - c_0 d_i(\mathbf{s}) + c_1 d_i(\mathbf{s}) N_i) \quad (2)$$

This function lacks biological realism at high values of  $c_1$  where it becomes a positive relationship with distance, implying that individuals lose the central-place constraint. A non-monotonic relationship would allow total usage to initially increase with distance, and then eventually decay (Fig. 1c). One way to achieve this is by a nonlinear extension of eq. (2)

$$\lambda(\mathbf{s}) = \exp(\alpha_0 - c_0 d_i(\mathbf{s}) + \frac{c_1 d_i(\mathbf{s}) N_i}{1 + c_2 d_i(\mathbf{s})}) \quad (3)$$

The parameter  $c_2$  controls how far from the colony the effect of density dependence is surpassed by the commuting costs of flying that far.

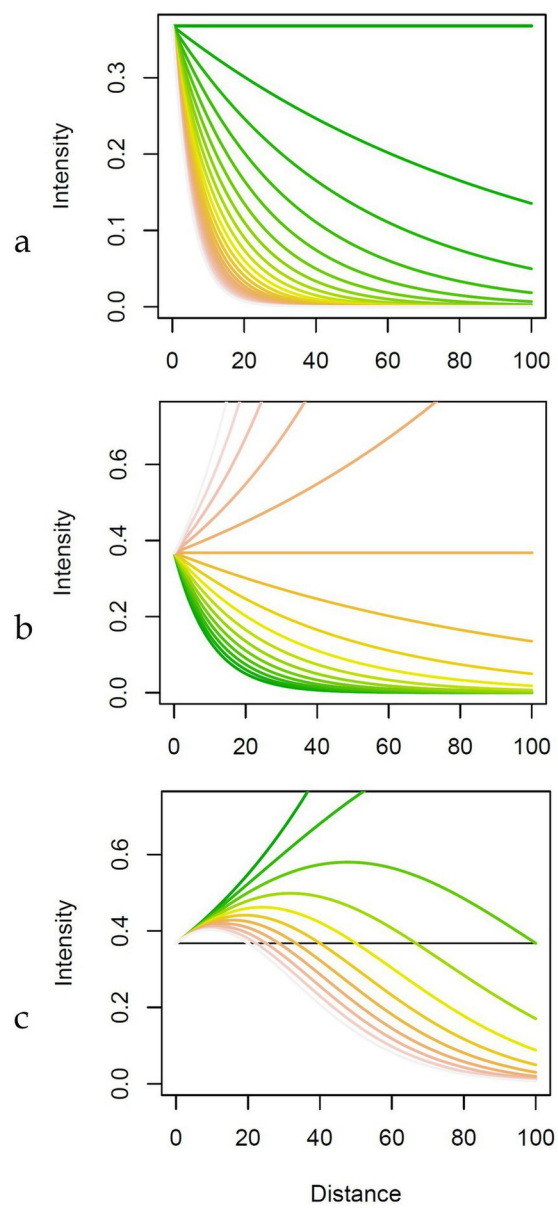
**Inter-colony competition.** Eq. (3) can be extended to account for the effects of other colonies.

The effect on the usage of a given marine point by a focal colony will depend on the distance of that point from the competing  $j^{\text{th}}$  colony (as well as the competing colony's size  $N_j$ ).

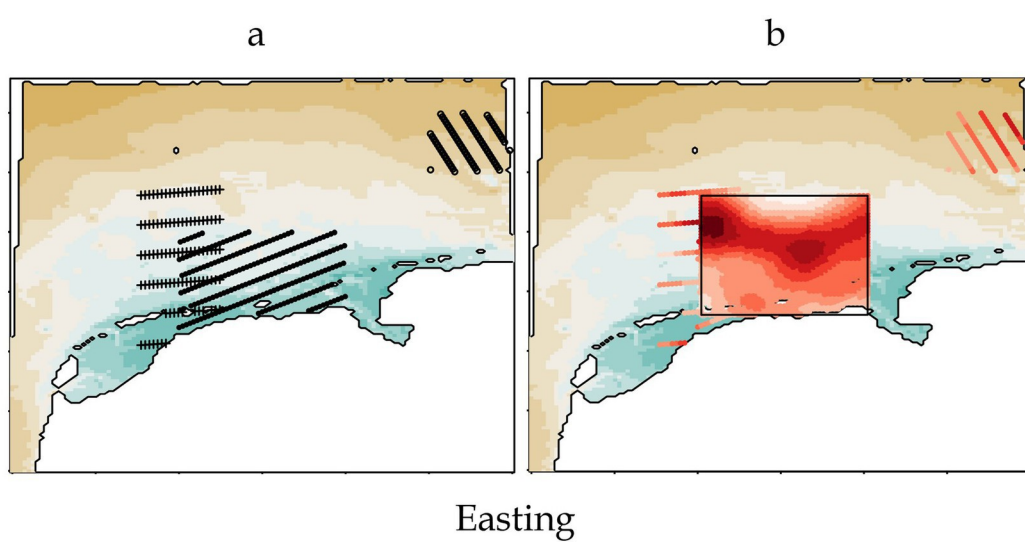
$$\lambda(\mathbf{s}) = \exp(\alpha_0 - c_0 d_i(\mathbf{s}) + \sum_j \frac{c_1 d_j(\mathbf{s}) N_j}{1 + c_2 d_j(\mathbf{s})}) \quad (4).$$

**Inter-colony, interspecific competition** could be dealt with in the same way as intraspecific competition between colonies, allowing for asymmetric effects due to differences in species, as well as differences in colony size. Although biologically, the differences between species are important, from a mathematical viewpoint, all that is required to capture these effects is a reparameterization of eq. (4) for the  $j^{\text{th}}$  colony of the  $k^{\text{th}}$  species:

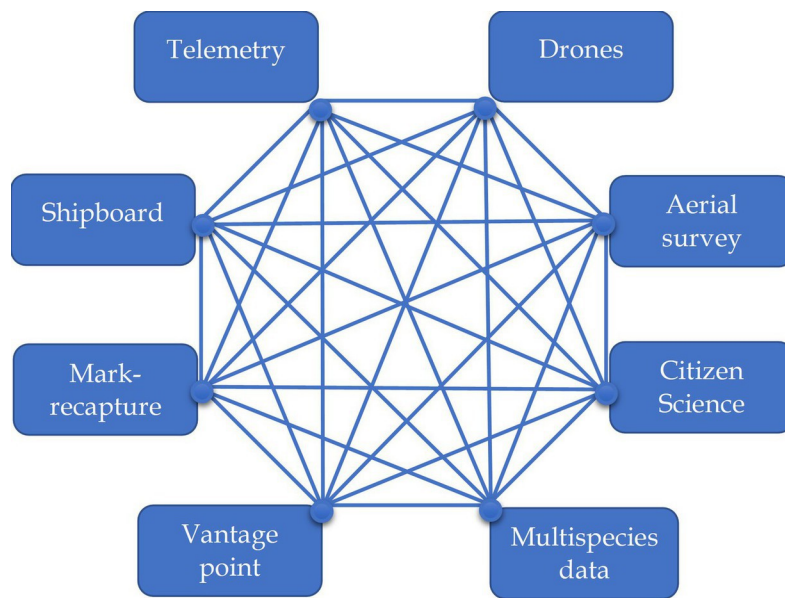
$$\lambda(\mathbf{s}) = \exp(\alpha_0 - c_0 d_i(\mathbf{s}) + \sum_k \sum_j \frac{c_{1,k} d_{k,j}(\mathbf{s}) N_{k,j}}{1 + c_{2,k} d_{k,j}(\mathbf{s})}) \quad (5)$$



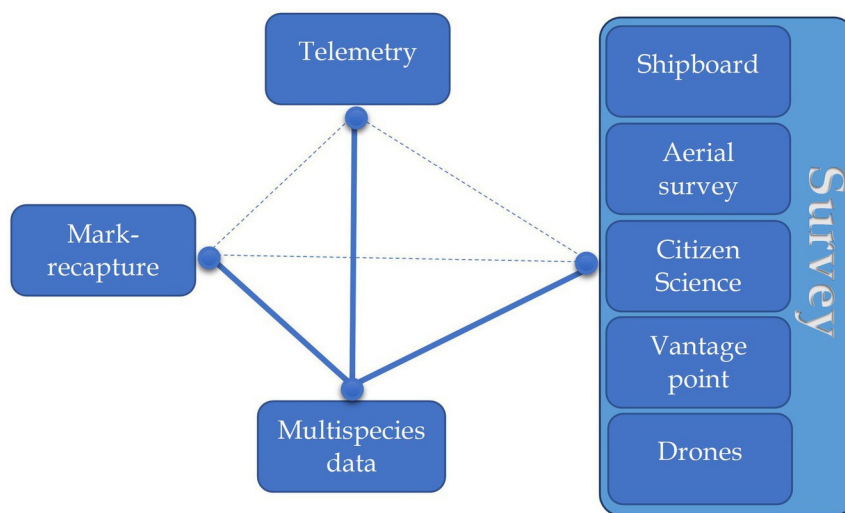
JPE\_14114\_Figure 1.JPG



JPE\_14114\_Figure 2.JPG



a. Types of data



b. Types of inference