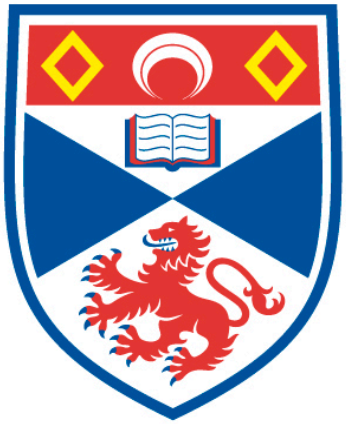
INCORPORATING ANIMAL MOVEMENT WITH DISTANCE SAMPLING AND SPATIAL CAPTURE-RECAPTURE

Richard Glennie

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews



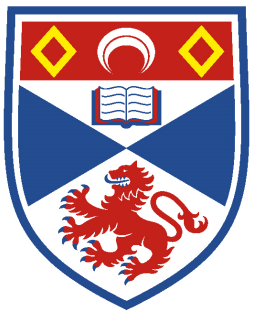
2018

Identifiers to use to cite or link to this thesis:

DOI: https://doi.org/10.17630/10023-16467
http://hdl.handle.net/10023/16467

Incorporating Animal Movement with Distance Sampling and Spatial Capture-Recapture

Richard Glennie



University of
St Andrews

This thesis is submitted in partial fulfilment for the degree of
Doctor of Philosophy (PhD)
at the University of St Andrews

March 2018

Candidate's declaration

I, Richard Glennie, do hereby certify that this thesis, submitted for the degree of PhD, which is approximately 50,000 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for any degree.

I was admitted as a research student at the University of St Andrews in September 2014.

I received funding from an organisation or institution and have acknowledged the funder(s) in the full text of my thesis.

Date: Signature of candidate:

Supervisor's declaration

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date: Signature of supervisor:

Permission for publication

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. We also understand, unless exempt by an award of an embargo as requested below, that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that this thesis will be electronically accessible for personal or research use and that the library has the right to migrate this thesis into new electronic forms as required to ensure continued access to the thesis.

I, Richard Glennie, confirm that my thesis does not contain any third-party material that requires copyright clearance.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

Printed copy

Embargo on all of print copy for a period of 2 years on the following ground(s):

- Publication would preclude future publication

Supporting statement for printed embargo request

The ability to publish work found within this thesis may be compromised if it is not embargoed.

Electronic copy

Embargo on all of electronic copy for a period of 2 years on the following ground(s):

- Publication would preclude future publication

Supporting statement for electronic embargo request

The ability to publish work found within this thesis may be compromised if it is not embargoed.

Title and Abstract

- I agree to the title and abstract being published.

Date:

Signature of candidate: Signature of supervisor:

Underpinning Research Data or Digital Outputs

Candidate's declaration

I, Richard Glennie, hereby certify that no requirements to deposit original research data or digital outputs apply to this thesis and that, where appropriate, secondary data used have been referenced in the full text of my thesis.

Date: Signature of candidate:

Acknowledgements

I would like to thank my supervisor Professor Stephen Buckland. I was lucky to have met Steve in my first year as an undergraduate at St Andrews. He inspired me to study Statistics, and his encouragement and motivation led to my pursuing a PhD. From the freedom he afforded me in my research, the confidence he has shown in me, and the advice he has given me, I have gained a lot from the experience.

I am further indebted to Steve as he introduced me to CREEM (Centre for Research into Ecological and Environmental Modelling). CREEM is a friendly, open, and collaborative research environment and I am thankful to everyone there who makes it a great place to study and work.

I would also like to thank Roland Langrock, my second supervisor, and Steve for pointing my research in the right direction and for the time they have both taken to help further the work contained in this thesis.

Finally, I would like to thank my family who I can always depend on for support, guidance, and, most importantly, laughter.

Data

I would like to acknowledge some people who provided data for analyses in this thesis and with whom I had many fruitful discussions:

- Tim Gerrodette and Lisa Ballance for taking the time to format and explain the distance sampling data collected for the spotted dolphin case study;
- Susan Chivers and Michael Scott for providing the dolphin tag data;
- Maite Louzao and José Manuel Arcos for providing the distance sampling and tag data on Balearic shearwaters. I am especially thankful to Maite for several useful discussions that contributed to my better understanding of this species;
- Bart Harmsen and Rebecca Foster who provided the data for the jaguar case study. Discussion with them both made it difficult not to be fascinated by the jaguar population.

Funding

This work was funded by the Carnegie Trust.

Abstract

Distance sampling and spatial capture-recapture are statistical methods to estimate the number of animals in a wild population based on encounters between these animals and scientific detectors. Both methods estimate the probability an animal is detected during a survey, but do not explicitly model animal movement.

The primary challenge is that animal movement in these surveys is unobserved; one must average over all possible paths each animal could have travelled during the survey. In this thesis, a general statistical model, with distance sampling and spatial capture-recapture as special cases, is presented that explicitly incorporates animal movement. An efficient algorithm to integrate over all possible movement paths, based on quadrature and hidden Markov modelling, is given to overcome the computational obstacles.

For distance sampling, simulation studies and case studies show that incorporating animal movement can reduce the bias in estimated abundance found in conventional models and expand application of distance sampling to surveys that violate the assumption of no animal movement. For spatial capture-recapture, continuous-time encounter records are used to make detailed inference on where animals spend their time during the survey. In surveys conducted in discrete occasions, maximum likelihood models that allow for mobile activity centres are presented to account for transience, dispersal, and heterogeneous space use.

These methods provide an alternative when animal movement causes bias in standard methods and the opportunity to gain richer inference on how animals move, where they spend their time, and how they interact.

Contents

Chapter 1

Introduction

Distance sampling (DS) (Buckland, Rexstad, Marques, & Oedekoven, 2015) and spatial capture-recapture (SCR) (Royle, Chandler, Sollmann, & Gardner, 2013) are statistical methods used to estimate how many animals there are in a wild animal population. Both survey methods involve one or more scientific detectors going out to where the animals are and recording those that they detect. This could be human observers in a ship at sea searching for seabirds (Jones et al., 2014), a plane flying over rough terrain surveying for bears (Aars et al., 2009), cameras placed in a thick forest photographing big cats that pass by (Harmsen, Foster, Silver, Ostro, & Doncaster, 2009), or microphones fixed to the bottom of the ocean recording whale song (Kyhn et al., 2012). In each case, a detector records data on any animals that are detected. DS and SCR can then be used to estimate how many animals were not detected; from which, population abundance can be deduced.

These population abundance surveys, when applied to wild animal populations, arise from a fundamental process: encounters between animals and scientific detectors. The encounters between the detectors and the animals produce the data used by DS and SCR. These encounters arise from the ability of the detectors to detect and the animals to move. Despite this, the statistical methods applied in DS and SCR do not explicitly model animal movement (Buckland et al., 2015; Royle, Fuller, & Sutherland, 2016); they are constructed around an assumption of immobility: animals are identified with static points in space. This can introduce bias in the inference obtained from these methods (Glennie, Buckland, & Thomas, 2015; Royle et al., 2016; Turnock & Quinn, 1991) and can limit the knowledge that is gained from the survey data. At present, there is no general statistical model to describe encounters that incorporates both detection and movement in continuous time (Gurarie & Ovaskainen, 2013).

The aim of this thesis is to create a general statistical model for the encounters that occur in DS and SCR surveys. This general model would explicitly incorporate animal movement with the existing methods used to describe detection.

1.1 Population Abundance Surveys

A key property of any population is its size. Abundance is a key indicator of the overall status of a population. It is one of the primary characteristics used to assess species conservation status and inform the decisions made to manage wild populations (Standart & Subcommittee, 2017). There is clear evidence that natural and anthropogenic forces are driving species to extinction and making balanced ecosystems more fragile and less diverse (IUCN, 2017). This has called for action across the world with continued monitoring schemes and targets such as those set by the Convention on Biological Diversity (Secretariat, 2010). To assess which species and regions are most at risk and to chart our progress toward improving biodiversity, sound measures of population size are a priority (Buckland & Johnston, 2017).

There are many possible methods to estimate population size (Borchers, Buckland, Zucchini, & Stephens, 2002). Distance sampling and capture-recapture are, arguably, the most widely-used. Distance sampling is applied to large study areas of unmarked animals: animals that cannot be individually identified. Capture-recapture is applied to marked individuals, either with natural, distinguishing marks or man-made marks attached, and is better suited to smaller study regions, as the method depends on observing re-encounters with individuals. Here is a brief overview of the two methods.

1.1.1 Distance Sampling

Distance sampling is a popular survey method because it is relatively inexpensive to apply over large areas and has proven to provide abundance estimates that are robust to many of the vicissitudes that can bias estimation by alternate methods (Buckland et al., 2001).

A DS survey is planned by a series of lines or points being placed randomly across the study region. An observer then visits each point or travels along each line, recording the location of each animal that is sighted. From this, the distance of the animal from the line (distance perpendicular to the line) or point can be computed. Consider the case of a line transect survey, as the idea is conceptually more intuitive for lines than points, the data often look similar to Figure 1.1. The idea is that the observed pattern of decline in the number of

animals seen at each distance is due to a decline in detection probability: animals further away are less likely to be recorded. The pattern is not due to the distribution of animals over space because lines and points are placed at random across the space, so animals are located on average at random with respect to the survey. A curve can be fit to these data to describe how the detection probability changes with distance (Buckland, 1992). The assumption is then made that this curve has an intercept of one: all animals on the line or at the point are detected with certainty. This is a necessary assumption, as the data from a single observer provide no information on the absolute value of the detection probability, only on how it changes with distance. After detection probability is computed, one can compute the proportion of animals that were sighted within a given distance and estimate the total number of animals. Overall, conventional DS relies upon three assumptions:

1. Animals on the line or point are detected with certainty;
2. Distances are measured without error;
3. Animals do not move.

The first assumption is sufficient to relate the observed data to detection probability. The second assumption is necessary as systematic or gross random error in distance measurement will affect the estimated curve. The third assumption is central to the entire approach: animals are assumed to be fixed to the location recorded for them.

The method has been applied to a variety of taxa across the world. Since the release of the software `Distance` (Laake, Buckland, Burnham, & Anderson, 1993; Laake, Burnham, & Anderson, 1979; Thomas et al., 2010) and the `R` package `Distance` (Miller, 2017), DS has become a standard method to estimate population size not only as a single number but to quantify how it changes over space, time, and other environmental variables such as habitat or season (Marques & Buckland, 2003; Miller, Burt, Rexstad, & Thomas, 2013). It is used extensively in the management of populations, such as sika deer populations in southern Scotland (Marques et al., 2001); it is used to assess conservation status of key species, such as polar bears (Aars et al., 2009). It is the method of choice for the UK Breeding Bird survey (Newson, Evans, Noble, Greenwood, & Gaston, 2008) and SCANS (Small Cetaceans in the European Atlantic and North Sea survey) (Hammond et al., 2017, 2013). Furthermore, as greater protection is demanded for marine wildlife, DS has been extended to acoustic surveys of whales in the SAMBAH (Static Acoustic Monitoring of the Baltic Sea Harbour Porpoise) (SAMBAH, 2016) and DECAF (Density Estimation for Cetaceans from Passive Acoustic Sensors) projects (Marques et al., 2013).

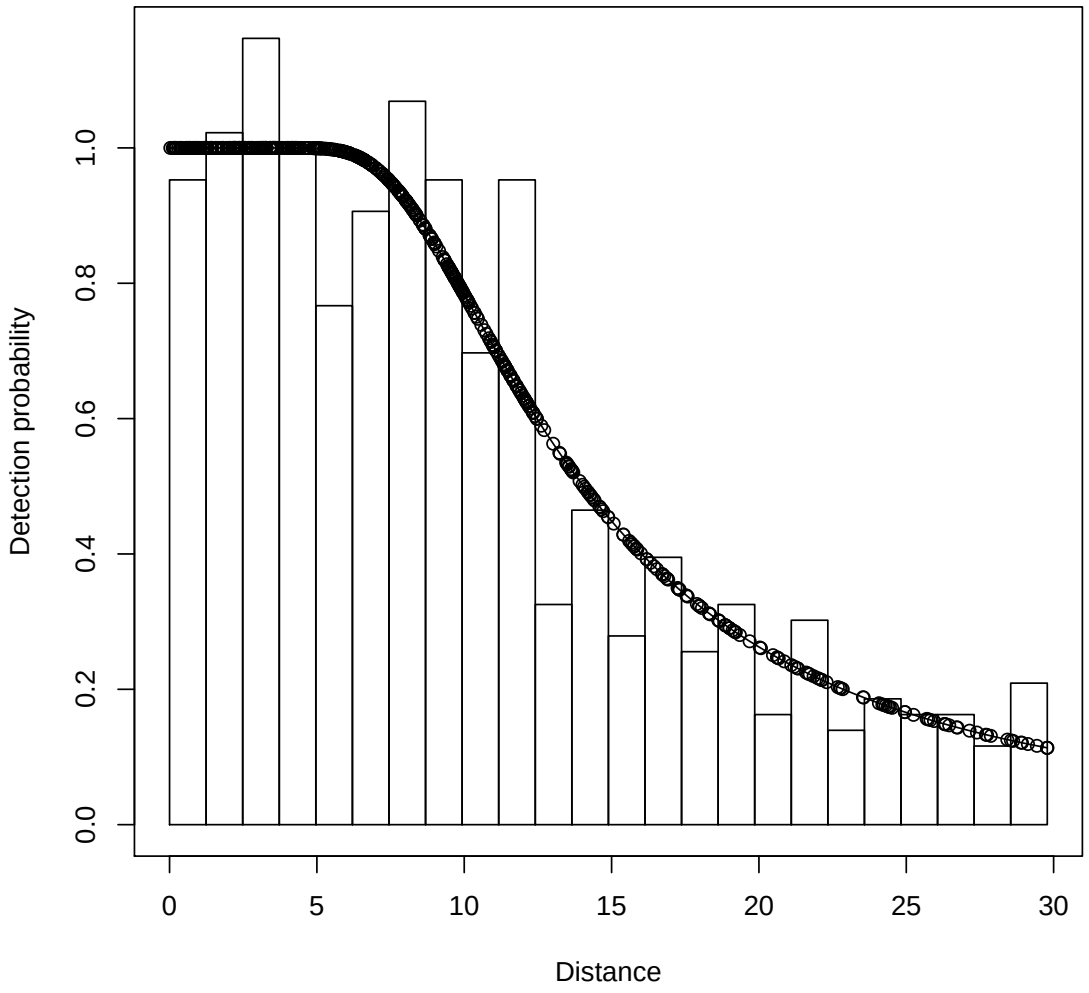


Figure 1.1 *Example of observations during a line transect survey: (scaled) number of animals seen at each distance with a fitted detection function*

As DS is applied in more contexts, violations of its core assumptions have arisen, and current research in DS has been driven by the desire to relax or remove each assumption (Buckland et al., 2015). In many surveys, the first assumption is violated: all animals are detected on the line or point. Cetaceans are a common example: they can dive for long periods of time and so be missed by shipboard surveys (Borchers, Zucchini, Heide-Jørgensen, Cañadas, & Langrock, 2013). This assumption was relaxed by the methods introduced by Borchers, Zucchini, and Fewster (1998), mark-recapture distance sampling (MRDS), where two or more observers can be used at once and their relative success at sighting the same animals used to quantify the probability of an animal being missed on the line or point. For the second assumption, that distance measurements are exact, improved technology, such as laser range-finders and anglometers, have improved accuracy; Marques (2004) and Borchers, Marques, Gunnlaugsson, and Jupp (2010) also developed statistical methods that can accommodate situations where measurement error is unavoidable. The third assumption can be violated in two ways and much more attention has been given to one compared to the other. Most focus has been on animals that move in response to the observer's presence (Borchers & Cox, 2017; Palka & Hammond, 2001; Turnock & Quinn, 1991). This can be directly observed during surveys and its effect on measured distances can be obvious, especially for evasive movement where a "hump" of recorded distances away from zero can be seen. Again, the recommendations of Buckland et al. (2001) for improved survey protocol are the first step to mitigating this violation. Furthermore, by having two or more observers, statistical models to account for responsive movement can be used to weaken the movement assumption (Borchers et al., 1998; Conn & Alisauskas, 2017). Alternatively, Palka and Hammond (2001) propose a method that accounts for animal movement by recording the observed orientation of detected individuals and comparing this to what one would expect when no responsive movement occurred. Interestingly, the movement of animals of their own volition has not been considered a problem in most DS surveys based on the guidance given in Buckland et al. (2001) and Hiby (1982). Nonetheless, non-responsive movement can cause substantial bias in density estimation (Glennie et al., 2015). Its effects on the observed data, however, are very difficult to discern (Glennie et al., 2015). This may explain why statistical methods to incorporate animal movement have not been a research priority.

The widespread use of distance sampling to mobile animal populations, its importance to conservation and management decision-making, and the possibly substantial, hidden bias in abundance estimation are motivations for incorporating animal movement with distance sampling methods.

1.1.2 Capture-Recapture

Capture-recapture (CR) is a classic population abundance method (Borchers et al., 2002). A modern statistical treatment is given by Seber (1982). The basic idea is that a sample of animals in the population are "captured" and a record of their identities made, making each animal uniquely identifiable either by a note of distinguishing marks they already possess or by making such a mark upon them. This first sample is then released back into the population. At a later occasion, another sample is captured and a record made of how many from the first sample have been re-captured. This proportion can then be used as an estimate of the detection probability assuming animals are equally catchable. The fewer re-captures, the smaller the probability of detection. Once the probability of detection is estimated, the population size can be estimated.

From this simple survey method, a wide variety of research methods have been developed. The early growth of capture-recapture is described by a series of review papers: Seber (1986), Seber (1992), and Schwarz and Seber (1999). The popularity of capture-recapture surveys is that they lead not only to estimates of population size, but can provide information on individuals across time. When the same individuals are re-captured, one can track their changing characteristics and, for longer studies, their life span. CR was first applied to closed populations (Otis, Burnham, White, & Anderson, 1978) where animals are assumed not to migrate from the study area, be born, or die during the survey. Yet, once it became apparent that the capture history of each individual provides information about population dynamics such as these, open population CR models were developed by Cormack (1964) and Jolly (1965) and Seber (1965) to provide estimates of survival, birth, and migration rates. This has further been incorporated within integrated population models (Besbeas, Freeman, Morgan, & Catchpole, 2002). There has been a greater drive toward using CR to gain a fuller picture of an animal population: its size and its dynamics. Software such as `CAPTURE` (Rexstad & Burnham, 1991) and `MARK` (White & Burnham, 1999) have made CR methods available to practitioners across Ecology and it has been applied to countless species. Its application also goes beyond Ecology, for example, when estimating drug users based on incomplete official registers (King, Bird, Hay, & Hutchinson, 2009).

During the growth of CR, the greatest concern has been on allowing for individual heterogeneity (Otis et al., 1978). The assumption of equal catchability is grossly violated in many animal populations (Chao, 1987): some animals are more likely to be captured and re-captured due to their behaviour or characteristics. To address violations of this assumption, a taxonomy of extensions of CR was introduced by Otis et al. (1978) where detection

probability varied with time, animal behaviour, heterogeneous animal characteristics, and their combinations. This, and many other approaches, were later unified by Pledger (2000) under the common theory of mixture modelling: the population is made up of a mixture of latent groups where the probability of detection for an animal depends on which group it is a member of.

One source of individual heterogeneity in capture probability had, until recently, been overlooked. When animals are captured, those that spend their time closer to the capture location are more likely to have been captured (and if the location is fixed, re-captured), than those further away. In other words, there is a spatial component to CR that in classical methods is not taken into account. A method for spatial capture-recapture (SCR) was introduced by Efford (2004) and the likelihood later developed by Borchers and Efford (2008). In this approach, each animal is associated with a single, static point in space known as their activity centre. The probability of an animal being captured depends on the distance between this activity centre and the capture location. Introducing this concept accounts for a source of heterogeneity. Furthermore, it provides a rigorous estimate of the space covered by the survey. Classic CR methods produce an estimate of abundance but do not attach an area to this estimate: one does not know, in theory, if these animals ranged over one hundred kilometres or one kilometre (Borchers & Efford, 2008). This made estimates of population density difficult to derive. SCR removes this issue by incorporating a spatial detection process. Since its introduction, SCR has been quickly adopted (Royle, Chandler, Sollmann, & Gardner, 2013). It has been extended within a maximum likelihood and Bayesian approach (Royle & Young, 2008) to accommodate many of the previous extensions of CR including mixture modelling (Borchers & Efford, 2008) and open population models (Gardner, Reppucci, Lucherini, & Royle, 2010). Furthermore, it is better suited to the new technology being used in CR surveys, for example, cameras, microphones, and DNA hair snares. Methods built around the original conception of scientists visiting a region and physically capturing animals are ill-suited to the reality of continuous technological surveys. SCR better matches the surveys as they are carried out. It can gain information not only from repeated captures of individuals over time, but also over space, at multiple detectors. With the development of continuous-time SCR models (Borchers, Distiller, Foster, Harmsen, & Milazzo, 2014), it can use the exact time of detections rather than the classical idea of discrete occasions.

A further effect of the advent of SCR is the relationship found between CR surveys and animal movement. Within SCR, animals are assumed to move around their activity centre such that they spend time at each location according to a Gaussian distribution around this

centre (Borchers & Efford, 2008). SCR provides estimates of the movement range during the survey and this can be related to telemetry observed on these animals (Sollmann et al., 2013). Similar to the inferences made from telemetry data on resource selection functions (Royle, Chandler, Sun, & Fuller, 2013), including environmental and temporal variables with SCR gives a picture of how individuals use space. As this link between SCR and movement is explored, greater demand is made for SCR methods to provide inference not only on population size, but on how animals move around (Royle, Fuller, & Sutherland, 2017). The difficulty of this is that SCR does not explicitly model animal movement; the movement is implicit and animals are assumed to teleport around their centre to independently sampled locations (Borchers & Efford, 2008). Density estimation has proven to be very robust to violation of this implicit movement model (Royle et al., 2016). Yet, estimates of movement parameters, such as range, can be substantially biased (Royle et al., 2016). Furthermore, the lack of an explicit movement model limits the inference SCR methods could, theoretically, provide. Captures of individuals over time and space are very similar to the telemetry tracks observed from GPS tagged animals (Linden, Siren, & Pekins, 2017). By assuming locations to be independent, a lot of the information these captures have about movement is thrown away. SCR surveys provide an opportunity to collect information about how a cross-section of the population moves, rather than the few individuals often focused on in telemetry studies. The incorporation of telemetry and SCR would likely provide richer inference, but this is only entirely possible by the incorporation of an animal movement model. Additionally, as research shifts from only inferring population size and distribution to how animals move through space, the ability of SCR to allow for animals to use space heterogeneously is a burgeoning research topic (Ovaskainen, Rekola, Meyke, & Arjas, 2008; Royle et al., 2017; Sutherland, Fuller, & Royle, 2015); yet, instead of fully incorporating an explicit movement model, current approaches prefer less computationally demanding techniques such as assuming animals move across a graph network and select the ecologically least-cost path (Royle, Chandler, Gazenski, & Graves, 2013).

The development of SCR is driving toward the incorporation of animal movement. SCR data provide a tempting resource of information about how a large section of the population move around and how they use their natural environment over long time periods.

1.1.3 Unified Theory

Distance sampling and spatial capture-recapture share a common statistical theory of detection. Borchers, Stevenson, Kidney, Thomas, and Marques (2015) laid out a unifying

framework that contains as special cases both DS and SCR. It is described as a spectrum. At one end is SCR where one does not observe an animal's activity centre, its location, but instead only its captures over space and time. At the other end is distance sampling where the exact location of the individual is recorded. In between these are methods where noisy information is collected on an animal's location; this would include cases with measurement error in the recorded location, or for acoustic surveys, one may be able to deduce the direction a sound came from, but not the distance. Overall, one can describe the theory of detection prevalent in statistical ecology as based around a single idea: each animal has a location and, during the survey, data are collected that provide a noisy idea of where that location is.

Currently, the framework of Borchers et al. (2015) provides the most abstract description of a statistical theory of detection. Abstraction allows for more efficient research: the core problems met in many similar contexts can be identified clearly and effort directed toward their solution. Once solved, these abstract solutions can be translated into concrete ones for each special case. Further to this, abstraction unifies many similar veins of research, allowing for each to borrow and improve upon ideas long established in the others. This cross-pollination can strengthen many methods at once.

What both methods, and by extension this unified framework, are missing is an explicit model for animal movement. This returns to the call from Gurarie and Ovaskainen (2013) for a general statistical model for encounters: a theory of detection combined with models for movement. This general framework would be built on the idea that detections during a survey are not noisy impressions of a single animal location, but arise from each animal's hidden path.

1.2 Animal Movement

The study of animal movement is a historic topic and there is a vast literature on statistical models for animal movement (Hooten, King, & Langrock, 2017; Turchin, 1998). The nascent field of movement ecology is rapidly expanding as new technology allows for the movement of greater numbers of animals to be tracked over longer times than ever before (Cagnacci, Boitani, Powell, & Boyce, 2010). This increase in data and opportunity has led to further statistical innovation (Patterson et al., 2017). Animals are being affixed with radio-collared tags and GPS tags which record their locations periodically. From these observed tracks, inference is made on how animals move, where animals go, and how animals behave (Nathan et al., 2008). Furthermore, telemetry data are used to describe how animals interact with

their natural environment through resource selection and how animals interact with each other (Hooten, Johnson, McClintock, & Morales, 2017).

The most intuitive and simplest statistical model for movement is the random walk (Berg, 1993). Animals take one step after another, choosing each time how large a step to take and in which direction. From this, there are extensions that allow for steps and turns to be correlated over time; for steps to be biased toward or away from particular locations, for example, animals may have a home from which they do not venture far; and for measurement error recorded steps and turns to be accounted for by state space models (Patterson, Thomas, Wilcox, Ovaskainen, & Matthiopoulos, 2008). In all of these models, animals are envisioned as taking steps: time progresses in discrete chunks, often called time-steps. There are two difficulties with this. First, in many cases, animal tracks are not observed as regular steps, locations are recorded irregularly over time (McClintock, Johnson, Hooten, Ver Hoef, & Morales, 2014). Discrete-time models cannot accommodate this, and thus practitioners are forced to interpolate the observations to create a sequence of regularly-spaced steps. The effect of this interpolation on the final inferences is poorly understood. Second, for a study species, one may observe telemetry at vastly different temporal scales (McClintock et al., 2014). For example, one may have minute-by-minute locations of individuals over a single day, and on another set of individuals have day-to-day locations over a year. Even though both tracks are regularly observed, the movement over a minute and the movement over a day are vastly different processes. One cannot describe both using the same discrete-time model. In the future, when all technology provides accurate, regular records that last over long periods, this may not be an issue. Today, one can use a GPS tag to get minute-by-minute records, but it will only last a day due to limited battery power, or one can use the same tag for a much longer time, but must limit the periodicity of records. Thus, this inability to integrate the two forms of data is a problem of current concern (Schick et al., 2008).

Continuous-time movement models are a remedy to these issues. In continuous-time models, animals move in paths rather than steps. The basic building block of most continuous-time models is the diffusion model (Okubo & Levin, 2013). This can be seen as the continuous-time analog of the random walk. Similar extensions as in the discrete-time approach can be made: correlated velocity and measurement error (Johnson, London, Lea, & Durban, 2008), and biased movement around a home centre using the Ornstein-Uhlenbeck model (Uhlenbeck & Ornstein, 1930). Despite this, the use of continuous-time models has lagged behind that of discrete-time models (McClintock et al., 2014). Continuous-time models remove the two drawbacks of discrete-time, but this comes at the cost of making an important

extension more difficult to implement: state-switching (McClintock et al., 2014).

A state-switching, sometimes called behaviour-switching, model is one where the parameters that govern an animal's movement switch between different values depending on a hidden state (Morales, Haydon, Frair, Holsinger, & Fryxell, 2004). Animals exhibit different behaviours over time and these behaviours affect how they move. For discrete-time, the use of the hidden Markov model (HMM) has proved invaluable (Zucchini, MacDonald, & Langrock, 2016). A HMM is built from two time series, one hidden and the other observed. The observed time series is the steps and turns of the animal; the hidden series is the sequence of behaviours the animal exhibits. HMM methods provide an efficient computational algorithm to average over all the possible state histories that could have produced the observed series of steps and turns. Intuitively, it clusters these steps and turns according to their similarities, identifying periods of long steps compared to short, or periods of persistent motion compared to desultory. It does this efficiently and provides intuitive inference. Continuous-time state-switching movement models exist (Blackwell, Niu, Lambert, & LaPoint, 2016) within a Bayesian framework, but their application is computationally expensive and, arguably, requires statistical expertise that every practitioner cannot be expected to have. The difficulty stems from the fact that in discrete time the state can change only at each time-step, but in continuous time, the state could change at any time and do so an infinite number of times. An efficient way to average over all these possible state histories whilst accounting for measurement error and irregular observations is not found in the literature. A two-stage approach (Scharf, Hooten, & Johnson, 2017) has been applied where a continuous-time movement model is fit to observed data, accounting for error and irregular observation, and then, from this fitted model, movement tracks are simulated in discrete-time, allowing for discrete-time state-switching methods to be applied. Yet, this method does not allow for state-switching to affect the movement as estimated in the first stage, and requires a large number of paths to be simulated when information on animal location is poor. Alternative approaches are to model behaviour as a function of observed covariates (Johnson et al., 2008) or, by treating space as a discrete grid, to model movement rates across this grid as a function of space, time, and covariates (Hanks, Hooten, Alldredge, et al., 2015). These approaches have the potential to make continuous-time movement models with behaviour-switching tractable; however, both do so by avoiding the need to average over all possible state histories. In this thesis, the focus is on doing exactly that efficiently.

With that in mind, one method to fit state-switching continuous-time models does merit special consideration. Pedersen, Patterson, Thygesen, and Madsen (2011) used an HMM to approximate the continuous-time process. Space was discretised into cells and time into

time-steps; the path of each animal across this space and the state history of that animal then formed a hidden time series. The observations can then arise from this time series. The efficient algorithm that made the HMM popular can then be used to fit a model that allows not only for state-switching in continuous-time but also allows for the measurement error, since the animal's true location is hidden and what is observed can be a noisy version of this. The continuous-time model and the HMM, which operates in discrete-time, are linked by a partial differential equation that describes how the animal's location changes over time. As the time-step and discretisation of space are reduced, the HMM comes to approximate the continuous-time process.

The description of this model closely resembles the description given of the proposed general model for encounters: animals move around and their path travelled is hidden; a survey is conducted and a noisy version of this path observed. It follows that the method developed by Pedersen et al. (2011) provides the framework for a general encounter model. The difficulty is that the application of this method requires a large amount of computer memory and computational time. The smaller the time-step and discretisation of space, the more infeasible the computations become. This limitation has prevented the realistic use of these methods.

From a general viewpoint, the problem is that one must average over all the possible paths an animal has travelled. This is an integral over an infinite-dimensional space of continuous functions, termed a path integral (Daniell, 1919). It has been given a rigorous formulation in functional analysis, used to derive the properties of stochastic processes, and become a key concept in the formulation of quantum mechanics. It is seldom calculated. The Bayesian approach by Blackwell et al. (2016) is a Monte Carlo approximation of this integral. The HMM approach by Pedersen et al. (2011) is a quadrature approximation. They both fall foul of the common issues found when approximating an integral in high-dimensional space. The Monte Carlo method requires a large number of samples from the space to produce an accurate approximation, while quadrature must either be coarse and inaccurate or fine and infeasible.

1.3 Thesis

1.3.1 Overview

The aim of this thesis is to incorporate animal movement with distance sampling and spatial capture-recapture. To achieve this, the thesis will focus on five objectives:

1. **Develop a general statistical model for encounters**:

   Chapter 2 presents a general model for encounters where the common theory of detection found in DS and SCR are incorporated with a class of advection-diffusion continuous-time movement models. This is further extended to include state-switching to allow both the detection and movement of animals to depend on their state.

2. **Improve the computational algorithm to perform path integration**:

   Once the general formulation for the encounter model is stated, the need to average over all possible unobserved animal paths is apparent. Thus, Chapter 2 presents an adaptation of the original HMM quadrature approach of Pedersen et al. (2011) that can do so feasibly and accurately. Matrix sparsity and linear algebraic dimension-reduction techniques are used to greatly improve upon the computational time and memory requirements of the quadrature approach. Furthermore, the convergence of the HMM approximation is proven for a class of integrands and movement models.

3. **Incorporate movement into distance sampling**:

   Chapter 3 presents the existing theory of distance sampling and shows it to be a special case of the general model in Chapter 2. Thus, this general model is used to incorporate an explicit animal movement model with the distance sampling method. Abundance estimation is shown to be biased by animal movement and this bias is mitigated by employing a movement model. State-switching is incorporated to expand the application of distance sampling to animals that have disparate behaviours such as seabirds that can rest on the sea surface or be in flight. Finally, Chapter 3 shows that the general framework can accommodate the existing extensions of distance sampling and can provide the opportunity to improve them in future research.

4. **Incorporate movement into continuous-time spatial capture-recapture**:

   Chapter 4 presents the theory of continuous-time spatial capture-recapture and shows it to be a special case of the general model in Chapter 2. Continuous-time SCR is more similar to distance sampling and to the general model than discrete-time SCR, where captures occur in discrete occasions. For this reason, the continuous-time SCR model is considered first. Incorporating an explicit movement model is shown to expand the inferences that can be made from continuous-time SCR. For example, the method is applied to a camera trap study of jaguars in Belize where each individual's movement is related to the existing path and river network within their range. Along with estimation of population size, inferences can now be made on each individual's

range during the survey and the spatio-temporal overlap in the range of individuals with neighbouring territories.

5. **Incorporate movement into discrete-time spatial capture-recapture**:

   The general model of Chapter 2 is formulated in continuous-time. In Chapter 5, this is shown to be easily adapted to the discrete-time framework of standard spatial capture-recapture. Incorporating a movement model into this framework allows each animal's activity centre to move over time. This can account for short-term transience found in some populations and the dispersal of individuals over long-term surveys. Some of these problems have been addressed by a Bayesian approach. The maximum likelihood alternatives have not been developed due to the need to compute a path integral. The computational methods from Chapter 2 are thus used to achieve this. This also provides a good opportunity to compare the HMM quadrature approach developed in Chapter 2 to the Bayesian methods applied independently in the literature. To do so, two previously analysed surveys are considered and compared with the new method.

1.3.2 Notation

The notation in this thesis is introduced when used. Here, some key aspects of notation are highlighted.

- Random variables can be upper or lower case and are not explicitly distinct from fixed values. Parameters are denoted by Greek letters.
- The set of all real numbers is denoted $\mathbb{R}$, the set of all integers by $\mathbb{Z}$, and the set of all positive integers by $\mathbb{N}$.
- Square brackets are used to denote probability density functions (PDF) and probability mass functions (PMF). For clarity, the random variables to which each PDF and PMF pertain is not explicitly stated, but is clear from the variables introduced. For example, $[x]$ denotes the PDF/PMF of the random variable x. When dependence of this PDF/PMF on a parameter θ is important, it is denoted as $[x \mid \theta]$. Similarly, the conditional PMF/PDF of a random variable x given a random variable y is denoted $[x \mid y]$. Their joint PDF/PMF would be denoted $[x, y]$. The purpose of this notation is to aid exposition.
- The expectation operator is denoted $\mathbb{E}$ and the variance operator by Var.

- Vectors are denoted in lowercase, bold font. Matrices are denoted in uppercase, bold font. The Euclidean distance between vectors $\boldsymbol{x}, \boldsymbol{y}$ is denoted $\|\boldsymbol{x} - \boldsymbol{y}\|$.
- A path is defined as a function that associates a two-dimensional location with each moment in time. A function that represents a path is denoted in bold font with an arrow; this makes paths, a key concept in this thesis, easily discernable. For example, if $\vec{\boldsymbol{x}}$ is a path in 2D space, then $\vec{\boldsymbol{x}}(t)$ is the location an animal has on that path at time t.

Chapter 2

Encounters

2.1 Introduction

The statistical methods to analyse distance sampling (DS) and spatial capture-recapture (SCR) data can be described as special cases of a general statistical model (Borchers et al., 2015). Both surveys rely upon encounters between scientific detectors and individuals of a given study population.

Definition 2.1 (Encounter)**.** A practical definition of an **encounter** in a scientific survey with a member of a study population is the event that this member is observed for an interval of time and a record of this observation is made.

Defining what constitutes an encounter is difficult to address in general (Gurarie & Ovaskainen, 2013). The practical definition will depend on the type of survey and the population under study. Two characteristics of this practical definition will depend on the context of the survey:

- Should repeated observations of an individual over a short time period be considered separate encounters or constituents of a single encounter? Gurarie and Ovaskainen (2013) defines an encounter such that an infinite number of encounters occur when an individual is observed over an interval of time. The suitability of this will depend on how individuals are encountered. For human observers, continuous observation of an individual is best described as a single encounter; however, for a motion-triggered camera, multiple pictures of an individual over a short time could be considered separate encounters, as the camera, unlike the human observer, has no memory of previous encounters. It is likely that most technology used to record encounters (e.g.

acoustic devices) does so when some threshold is met (significant motion in front of a camera, or sound levels for an acoustic device) and so each record can be considered a new encounter. This is not the case for the human observer because continuous observation does not constitute a series of *independent* encounters. Much of the statistical theory that follows in this chapter requires encounters are independently recorded given the path travelled by individuals and so what constitutes an encounter should be determined with this in mind.

- The practical definition requires that a record of the encounter is made. What this record consists of will again depend on the type of survey and the study population. In distance sampling, this record consists of the location where the individual was seen and the time when the encounter took place. In SCR, the identity of the individual is recorded and the time of the encounter. The information recorded for each encounter must be sufficient to ensure the parameters governing the underlying process are identifiable.

Both distance sampling and SCR surveys can be described as one or a series of encounter surveys.

Definition 2.2 (Encounter Survey)**.** Let the study area, $\mathcal{A} \subset \mathbb{R}^2$, be a subset of the Euclidean plane and $T \in [0, \infty)$ be the total time spent surveying. An **encounter survey** has the following elements:

1. A study population of size $N \in \mathbb{N}$. Each member of the population is termed a **target** and one can identify when a target is encountered repeatedly.

2. Each target k follows a continuous path $\vec{\boldsymbol{x}}_k : [0, T] \to \mathcal{A}$ within the survey area during the survey time. The paths of targets may be completely or partially unobserved during the survey. The space of all continuous paths over $\mathcal{A}$ is denoted $\mathcal{C}(\mathcal{A})$.

3. There are $J \in \mathbb{N}$ **detectors** where the j^{th} detector follows a continuous path $\vec{\boldsymbol{z}}_j : [0, T] \to \mathcal{A}$. The paths of detectors are completely observed during the survey.

4. There are n targets encountered during the survey time. A target that was encountered at least once is termed an **encountered target**. Targets that were never encountered are termed **unencountered targets** or **missed targets**.

5. The **record** of the e^{th} encounter between the i^{th} encountered target and the j^{th} detector is denoted $\boldsymbol{r}_{i,j,e}$. Notice that $\boldsymbol{r}_{i,j,e}$ is a vector as the record may contain

auxiliary information. All records of encounters between encountered target i and detector j are denoted $\boldsymbol{r}_{i,j}$, all records of encounters with target i are denoted $\boldsymbol{r}_i$, and the collection of all records of encounters during the survey is denoted $\boldsymbol{r}$.

6. The number of encounters with encountered target i is denoted n_i and $n_{i,j}$ is the number of encounters between target i and detector j. Furthermore, let $n_{i,j}(I)$ be the number of encounters between detector j and target i in the time interval I.

7. The **encounter time** of the e^{th} encounter of detector j with the i^{th} encountered target is denoted $t_{i,j,e}$. It is assumed that encounter times are observed. A collection of all detection times for target i with detector j is denoted $\boldsymbol{t}_{i,j}$, the collection of all encounters with i is denoted $\boldsymbol{t}_i$, and the collection of all detection times denoted $\boldsymbol{t}$.

This chapter describes the statistical methods to analyse the data that arise from an encounter survey. The aim is to model how targets move around and the process by which detectors encounter targets. To achieve this, the joint probability density function (PDF) of the observed encounters $\boldsymbol{r}$ is constructed. Encounters depend upon two processes: detection and movement. Detectors and targets move around the survey area and their relative positions determine how likely or unlikely it is that a detector will detect a target. Current methods in DS and SCR focus on models to describe detection whilst idealising targets as immobile points in space. In this Chapter, it is shown that this existing theory of detection can be incorporated with statistical models for target movement. The full model is developed in three stages:

1. the statistical model for detection is constructed conditional on the path of every target being known;

2. a model for target movement is specified, associating with each possible target path a probability density;

3. the detection and movement models are incorporated together such that when paths are unobserved one can average over all possible paths the target could have taken, given the model for how targets move.

2.2 Detection

In an encounter survey, some targets are missed and some are encountered. Furthermore, some targets can be encountered more often than others. Conceptually, encounters can be

viewed as a point process over time. Encounter times $\boldsymbol{t}_{i,j}$ are the realisation of this process on $[0, T]$: a point is placed at each time where an encounter occurred.

Point processes are often described using a counting measure and intensity function which describes the number of points that occur in a given interval and the propensity for points to appear at particular times (Cox & Isham, 1980). In this context, the counting measure is the number of encounters that have occurred with that target and the intensity function is the propensity for encounters to occur at certain times, often driven by the location of the target with respect to the detectors. Each detector contributes toward the intensity of encounters that target may have. The intensity accumulated over a segment of path is simply the integral of the intensity over that path.

Definition 2.3 (Accumulated Intensity)**.** The **counting measure** or **accumulated intensity** for detector j and target i with path $\vec{\boldsymbol{x}}_i$ is the function $\Lambda_j : \mathcal{C}(A) \times [0, T]^2 \to [0, \infty]$ such that

$$\mathbb{E}(n_{i,j}(s, t)) = \Lambda_j(\vec{\boldsymbol{x}}_i, s, t).$$

This quantity will depend on the path taken by detector j, this is suppressed to improve readability. Also, for brevity, $\Lambda_j(\vec{\boldsymbol{x}}_i, t) = \Lambda_j(\vec{\boldsymbol{x}}_i, 0, t)$. In particular, note that $\mathbb{E}(n_{i,j}) = \Lambda_j(\vec{\boldsymbol{x}}_i, T)$.

The **intensity** function associated with Λ_j is the Radon-Nikodym derivative: the function $\lambda_j : \mathcal{A} \times [0, T] \to [0, \infty]$ such that

$$\Lambda_j(\vec{\boldsymbol{x}}_i, s, t) = \int_s^t \lambda_j(\vec{\boldsymbol{x}}_i(u), u)\, \mathrm{d}u$$

for all paths $\vec{\boldsymbol{x}}$ and time intervals (s, t) where $s < t$.

Intuitively, $\lambda_j(\boldsymbol{x}, t)$ is the density or intensity of encounters that occur at time t when the target is in location $\boldsymbol{x}$. This is the reason that Λ_j is termed the accumulated intensity.

A central assumption of the detection theory used in DS and SCR is that the encounter intensity or hazard of encounter of a target with a detector depends on the distance that separates them. In particular, it is natural to impose the constraint that the encounter intensity should decrease as the distance between target and detector lengthens.

Definition 2.4 (Encounter Intensity)**.** A **hazard** function (Hayes & Buckland, 1983) is a function $h : [0, \infty] \to [0, \infty]$ that is montonically decreasing: for $d_1, d_2 \in [0, \infty]$

$$d_1 < d_2 \implies h(d_1) \geq h(d_2)$$

An intensity function λ_j is an **encounter intensity** function when $\lambda_j(\boldsymbol{x}, t) = h(\|\vec{\boldsymbol{x}} - \vec{\boldsymbol{z}}_j(t)\|)$ for some hazard function h for all $\boldsymbol{x}, t$.

There are several common functional forms for the hazard function (Borchers & Cox, 2017; Hayes & Buckland, 1983).

- The simplest form is the **hard-core** constant rate where intuitively each detector is surrounded by a circle within which targets are encountered with a fixed rate irrespective of distance:

$$h(d) = \begin{cases} \alpha : d \leq \rho \\ 0 : d > \rho \end{cases}$$

  for parameters $\rho > 0$, the radius of the hard-shell around the detector, and $\alpha > 0$, the constant encounter rate.

- The **radial encounter rate** is given by

$$h(d) = \left(\frac{d}{\sigma}\right)^{\alpha}$$

  for a scale parameter $\sigma > 0$ and a shape parameter $\alpha > 1$.

- The **Gaussian encounter rate** is given by

$$h(d) = \exp\left(-\left(\frac{d}{\sigma}\right)^2\right).$$

After a hazard function is chosen, the counting measure can be computed. There are many point process models that could be used to describe the encounters. In this thesis, only the Poisson process model is used. A definition of the Poisson process is given in Diggle (2013).

Definition 2.5 (Poisson detection process)**.** Encounters between detector j and target i arise from a Poisson **detection process** when

(a) $n_{i,j}(a, b)$ given $\vec{\boldsymbol{x}}_i$ has a Poisson distribution with mean $\Lambda_j(\vec{\boldsymbol{x}}_i, a, b)$;

(b) for disjoint intervals $(a, b), (c, d)$, the random variables $n_{i,j}(a, b)$ and $n_{i,j}(c, d)$ are independent.

Specifying a model for the detection process determines the probability that in a given time interval each target will be detected or not. Assuming a Poisson detection process implies

that all dependence between encounters is induced by the target's chosen path; in other words, encounters are conditionally independent given the target's path.

Assuming the encounter process follows a Poisson point process leads to an important simplification of the model that makes this approach computationally practical. The key property used is that the Poisson encounter process is memoryless and the PDF of the encounter times can be written as product of functions on disjoint time intervals. This is similar to the probability density functions for transition in discrete space as derived by Hanks et al. (2015).

Theorem 2.1. For detection times $\boldsymbol{t}_{i,j}$ in a Poisson detection process:

(a) the PDF of the first encounter time given $\vec{\boldsymbol{x}}_i$ is

$$[t_{i,j,1} \mid \vec{\boldsymbol{x}}_i] = \lambda_j(\vec{\boldsymbol{x}}_i(t_{i,j,1})) \exp\left(-\Lambda_j(\vec{\boldsymbol{x}}_i, t_{i,j,1})\right);$$

(b) the Poisson encounter process is memoryless:

$$[t_{i,j,e+1} \mid t_{i,j,e}, \vec{\boldsymbol{x}}_i] = \lambda_j(\vec{\boldsymbol{x}}_i(t_{i,j,e+1})) \exp\left(-\Lambda_j(\vec{\boldsymbol{x}}_i, t_{i,j,e}, t_{i,j,e+1})\right);$$

(c) the joint PDF of $n_{i,j}$ and $\boldsymbol{t}_{i,j}$ given $\vec{\boldsymbol{x}}_i$ is

$$[\boldsymbol{t}_{i,j}, n_{i,j} \mid \vec{\boldsymbol{x}}_i] = \exp\left(-\Lambda_j(\vec{\boldsymbol{x}}_i, T)\right) \prod_{e=1}^{n_{i,j}} \lambda_j(\vec{\boldsymbol{x}}_i(t_{i,j,e})).$$

Proof.

(a) The PDF of $[t_{i,j,1}]$ is obtained by differentiating the corresponding cumulative distribution function.

$$\begin{aligned}
[t_{i,j,1} \leq t \mid \vec{\boldsymbol{x}}_i] &= 1 - [t_{i,j,1} > t \mid \vec{\boldsymbol{x}}_i] \\
&= 1 - [n_{i,j}(0,t) = 0 \mid \vec{\boldsymbol{x}}_i] \\
&= 1 - \exp\left(-\Lambda_j(\vec{\boldsymbol{x}}_i, 0, t)\right).
\end{aligned}$$

The last line follows from the Poisson distribution.

Thus, the fundamental theorem of calculus implies

$$[t_{i,j,1} \mid \vec{\boldsymbol{x}}_i] = \lambda_j(\vec{\boldsymbol{x}}_i(t_{i,j,1})) \exp\left(-\Lambda_j(\vec{\boldsymbol{x}}_i, 0, t_{i,j,1})\right)$$

as required.

(b) The second result is obtained similarly:

$$\begin{aligned}[t_{i,j,e+1} \leq t \mid t_{i,j,e}, \vec{\boldsymbol{x}}_i] &= 1 - [t_{i,j,e+1} > t \mid t_{i,j,e}, \vec{\boldsymbol{x}}_i] \\ &= 1 - [n_{i,j}(t_{i,j,e}, t) = 0 \mid t_{i,j,e}, \vec{\boldsymbol{x}}_i] \\ &= 1 - \exp\left(-\Lambda_j(\vec{\boldsymbol{x}}_i, t_{i,j,e}, t)\right).\end{aligned}$$

Thus, the result follows from (b) of Definition 2.5 and the fundamental theorem of calculus.

(c) The joint probability density of $\boldsymbol{t}_{i,j}$ and $n_{i,j}$ is the probability density of $n_{i,j}$ encounters in sequence at times $\boldsymbol{t}_{i,j}$ and no other encounters occurring afterward.

$$\begin{aligned}[\boldsymbol{t}_{i,j}, n_{i,j} \mid \vec{\boldsymbol{x}}_i] &= [t_{i,j,1}, \ldots, t_{i,j,n_{i,j}}, n_{i,j}(t_{i,j,n_{i,j}}, T) = 0 \mid \vec{\boldsymbol{x}}_i] \\ &= [n_{i,j}(t_{i,j,n_{i,j}}, T) = 0 \mid \vec{\boldsymbol{x}}_i][t_{i,j,1} \mid \vec{\boldsymbol{x}}_i] \prod_{e=1}^{n_{i,j}-1} [t_{i,j,e+1} \mid t_{i,j,e}, \vec{\boldsymbol{x}}_i]\end{aligned}$$

Since, $[n_{i,j}(t_{i,j,n_{i,j}}, T) = 0 \mid \vec{\boldsymbol{x}}_i] = \exp\left(-\Lambda_j(\vec{\boldsymbol{x}}_i, t_{i,j,n_{i,j}}, T)\right)$, the result follows from (a) and (b).

□

Theorem 2.1 provides the PDF for the detection times and the number of encounters between a single detector and target.

Multiple detectors in an encounter survey are assumed to be independent given the paths followed by each target $\vec{\boldsymbol{x}}_1, \ldots, \vec{\boldsymbol{x}}_n$. Given this, the set of all encounter times with target i, $\boldsymbol{t}_i$, are a realisation from a superposition of Poisson point processes with intensity functions $\lambda_1, \ldots, \lambda_J$; this is itself a Poisson point process with intensity $\lambda = \sum_{j=1}^{J} \lambda_j$ by the superposition property (Cox & Isham, 1980).

The joint PDF of all encounters with target i is

$$[\boldsymbol{t}_i \mid \vec{\boldsymbol{x}}_i] = \prod_{j=1}^{J} [\boldsymbol{t}_{i,j} \mid \vec{\boldsymbol{x}}_i].$$

The complete PDF of the observed encounters can be constructed by further assuming targets are independent.

Definition 2.6 (Independent Targets)**.** Targets in an encounter survey are **independent** when their encounter times $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n$ are mutually independent sets of random variables given their paths $\vec{\boldsymbol{x}}_1, \ldots, \vec{\boldsymbol{x}}_n$:

$$[\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n \mid \vec{\boldsymbol{x}}_1, \ldots, \vec{\boldsymbol{x}}_n] = \prod_{i=1}^{n} [\boldsymbol{t}_i \mid \vec{\boldsymbol{x}}_i].$$

Finally, the detection probability of each target can be derived.

Theorem 2.2 (Detection probability)**.** For a Poisson detection process (Borchers & Efford, 2008),

(a) the probability of target i being detected by detector j at least once in the encounter survey given target i travels path $\vec{\boldsymbol{x}}_i$ is

$$p_{i,j}(\vec{\boldsymbol{x}}_i) = 1 - \exp\left(-\Lambda_j(\vec{\boldsymbol{x}}_i, T)\right);$$

(b) if detectors are independent, the probability of target i being detected by any detector at least once in the encounter survey given target i travels path $\vec{\boldsymbol{x}}_i$ is

$$p_i(\vec{\boldsymbol{x}}_i) = 1 - \exp\left(-\sum_{j=1}^{J} \Lambda_j(\vec{\boldsymbol{x}}_i, T)\right).$$

Proof. The event target i is never encountered by detector j is equivalent to the event that $n_{i,j} = 0$. Thus,

$$p_{i,j}(\vec{\boldsymbol{x}}_i) = 1 - [n_{i,j} = 0 \mid \vec{\boldsymbol{x}}_i] = 1 - \exp\left(-\Lambda_j(\vec{\boldsymbol{x}}_i, T)\right)$$

as $n_{i,j}$ has a Poisson distribution with mean $\Lambda_j(\vec{\boldsymbol{x}}_i, T)$.

Similarly, target i is never encountered by any detector if and only if $n_{i,1} = \ldots = n_{i,J} = 0$, thus

$$\begin{aligned} p_i(\vec{\boldsymbol{x}}_i) &= 1 - [n_{i,1} = 0, \ldots, n_{i,J} = 0 \mid \vec{\boldsymbol{x}}_i] \\ &= 1 - \prod_{j=1}^{J} [n_{i,j} = 0 \mid \vec{\boldsymbol{x}}_i] \qquad \text{since detectors are independent} \\ &= 1 - \prod_{j=1}^{J} \exp\left(-\Lambda_j(\vec{\boldsymbol{x}}_i, T)\right) \end{aligned}$$

$$= 1 - \exp\left(- \sum_{j=1}^{J} \Lambda_j(\vec{\boldsymbol{x}}_i, T) \right).$$

□

This section provides the mathematical model for how encounters between detectors and targets arise given the movements of both agents are known.

2.3 Movement

In the previous section, a target's path $\vec{\boldsymbol{x}}_i : [0, T] \to \mathcal{A}$ was treated as known while in most encounter surveys (e.g. DS and SCR) the target's path will be unobserved. Indeed, one motivation for an encounter survey is to determine the properties of target movement. In this section, a target's path is treated as a stochastic process over time. A class of movement models is introduced and discussed. The primary purpose of the movement model is to allow a full model for encounters to be constructed in the next section. Further to this, this section discusses how the parameters of each movement model can be estimated from telemetry data.

Definition 2.7 (Telemetry)**.** A **telemetry** survey of m targets over a time period $[0, T]$ in a survey region $\mathcal{A}$ tracks the movements of each target by some means (e.g. visual tracking, GPS tagging, radio tagging) and records the location of each target at a finite number of times. In particular for target i with path $\vec{\boldsymbol{x}}_i : [0, T] \to \mathcal{A}$, the telemetry record is of the form $(\vec{\boldsymbol{x}}_i(t_1), \ldots, \vec{\boldsymbol{x}}_i(t_{R_i}))$ for times $t_1 < \ldots < t_{R_i}$ for some number of records $R_i \in \mathbb{N}$.

Targets move in continuous space and time. A stochastic differential equation (SDE) can be used to describe a model for how a target's path is generated (Preisler, Ager, Johnson, & Kie, 2004). In this section, the movement models considered are described by the same form of SDE. The class of movement models is constructed from a Wiener process. The Wiener process is a mathematical representation of Brownian motion.

Definition 2.8 (Wiener process)**.** The stochastic process $\vec{\boldsymbol{x}} : [0, T] \to \mathcal{A}$ is a **Wiener process** with a mean $\boldsymbol{\mu} : \mathcal{A} \times [0, T] \to \mathbb{R}^2$ and variance $\sigma : \mathcal{A} \times [0, T] \to (0, \infty]$ when

- $\vec{\boldsymbol{x}}(0)$ is known or has a specified distribution over space;
- for $s, t \in [0, T]$ where $s < t$, $\vec{\boldsymbol{x}}(t) - \vec{\boldsymbol{x}}(s)$ has a bivariate Gaussian distribution with

mean $\boldsymbol{\mu}(\vec{\boldsymbol{x}}(s), s)$ and diagonal covariance matrix with both diagonal entries equal to $(t-s)\sigma^2(\boldsymbol{x}(s), s)$;

- the process is a Markov process: for $s, t, u, v \in [0, T]$ where $s < t < u < v$, the increments $\vec{\boldsymbol{x}}(v) - \vec{\boldsymbol{x}}(u)$ and $\vec{\boldsymbol{x}}(t) - \vec{\boldsymbol{x}}(s)$ are independent random variables.

Furthermore, the Wiener process $\vec{\boldsymbol{w}} : [0, T] \to \mathcal{A}$ with zero mean and $\sigma = 1$ is termed the **standard Wiener process**.

A stochastic differential equation (SDE) is a convenient way to specify a Wiener process in terms of how a target's position changes over time (Brillinger, Preisler, Ager, & Kie, 2004; Preisler et al., 2004).

Definition 2.9 (Langevin)**.** For a Wiener process $\vec{\boldsymbol{x}} : [0, T] \to \mathcal{A}$ with mean $\boldsymbol{\mu}$ and variance σ, the **Langevin** equation of the process is the stochastic differential equation:

$$\mathrm{d}\vec{\boldsymbol{x}}(t) = \boldsymbol{\mu}(\vec{\boldsymbol{x}}(t), t)\ \mathrm{d}t + \sigma(\vec{\boldsymbol{x}}(t), t)\ \mathrm{d}\vec{\boldsymbol{w}}(t)$$

where $\vec{\boldsymbol{w}}$ is a standard Wiener process. The Langevin equation describes the movement of each individual target over an infinitely small time period.

This SDE is similar to the standardisation of a random normal variable: if Z is a standard Gaussian variable then $X = \mu + \sigma Z$ is a Gaussian variable with mean μ and standard deviation σ. Intuitively, the SDE states that over a small time-step Δt, $\vec{\boldsymbol{x}}(t + \Delta t)$ given $\vec{\boldsymbol{x}}(t)$ has a bivariate Gaussian distribution with mean $\boldsymbol{\mu}(\vec{\boldsymbol{x}}(t), t)$ and variance $\sigma(\vec{\boldsymbol{x}}(t), t)\Delta t$ in both components and zero covariance. Notice, that the covariance matrix of this bivariate Gaussian is diagonal and that both diagonal elements are equal. This has an intuitive justification (Johnson et al., 2008): a non-zero off-diagonal element would imply that a target's movement in one direction is correlated with its movement in another leading to diamond shaped movement patterns (e.g. for positive correlation, up correlated with right movement, and down correlated with left movement). This is unrealistic for most movements that are not constrained to a diamond-shaped survey region, thus the off-diagonals are constrained to be zero. The diagonal entries are constrained to be equal since there is often nothing to justify that target movement is related to the coordinate system used to record target movement. Constraining the diagonal entries to be equal ensures the movement model is invariant to rotational coordinate transformations.

The family of movement models presented here contains many of the standard continuous-time movement models used in telemetry analyses, bringing them together under a single framework.

In the remainder of this section, special cases of this SDE are discussed, considering briefly how their parameters can be estimated from telemetry data.

2.3.1 $\mathrm{d}\vec{\boldsymbol{x}}(t) = \boldsymbol{\mu}\,\mathrm{d}t + \sigma\,\mathrm{d}\vec{\boldsymbol{w}}(t)$

The simplest case is when the mean and variance do not depend on time or space, that is, they are constants.

Simple Diffusion

When $\boldsymbol{\mu} = \mathbf{0}$, the equation $\mathrm{d}\vec{\boldsymbol{x}}(t) = \sigma\,\mathrm{d}\vec{\boldsymbol{w}}(t)$ is the Langevin equation for **simple diffusion**, alternatively named Brownian motion, where direction and speed are uncorrelated. The only parameter is the diffusion constant σ. The parameter σ can be estimated from telemetry data (Okubo & Levin, 2013).

Ideal Gas Movement

Alternatively, when $\sigma = 0$, the SDE is no longer stochastic: $\mathrm{d}\vec{\boldsymbol{x}}(t) = \boldsymbol{\mu}\,\mathrm{d}t$. The differential equation describes straight line motion, often termed the **ideal gas** movement model (Hutchinson & Waser, 2007).

Advection-Diffusion

Advection-Diffusion models are a combination of the ideal gas model and simple diffusion (Preisler et al., 2004; Turchin, 1998). Targets continue to move in a straight line *on average* but their movement can diffuse randomly around this line. It is akin to linear regression where data points conceptually arise from a signal that lies on the line with Gaussian noise which displaces the data around the line. Advection is the term used to describe the drift of the target in a single direction; the diffusion describes the variability in the movement that is not linear.

The PDF for an advection-diffusion model is obtained by marginalising over all possible target directions of travel.

Theorem 2.3. If $\vec{\boldsymbol{x}} : [0, T] \to \mathcal{A}$ is an advection-diffusion process with parameters $\boldsymbol{\theta} = (\nu, \alpha, \sigma)$, $f_{\boldsymbol{\theta}}$ is the PDF of the bivariate Gaussian distribution with mean $\nu(\cos\alpha, \sin\alpha)$ and

covariance matrix $\sigma^2\boldsymbol{I}$, and $t_1, \ldots, t_m \in [0, T]$ such that $t_1 < \ldots < t_m$ then

$$[\boldsymbol{x}(t_1), \ldots, \boldsymbol{x}(t_m)] = [\boldsymbol{x}(t_1)] \frac{1}{2\pi} \int_0^{2\pi} \prod_{r=2}^{m} f_{\boldsymbol{\theta}} \left(\frac{\boldsymbol{x}(t_r) - \boldsymbol{x}(t_{r-1})}{\sqrt{t_r - t_{r-1}}} \right) \, \mathrm{d}\alpha$$

This PDF can then be computed numerically and the parameters ν, σ estimated by maximum likelihood.

2.3.2 $\mathrm{d}\vec{\boldsymbol{x}}(t) = \boldsymbol{\mu}(t)\,\mathrm{d}t + \sigma(t)\,\mathrm{d}\vec{\boldsymbol{w}}(t)$

The functions $\boldsymbol{\mu}$ and σ can depend on time in two ways. First, they could be a function of time with certain parameters (e.g. Kie, Ager, and Bowyer (2005)). Alternatively, $\boldsymbol{\mu}$ and σ can themselves be described by stochastic processes (Blackwell et al., 2016; Gurarie et al., 2017). This is different to the former approach as the value of $\boldsymbol{\mu}$ at any particular time is a random variable and not specified by a function that is known given the parameters of the process.

Once the functions $\boldsymbol{\mu}$ or σ are stochastic, the PDF in Theorem 2.3 cannot be used. That PDF depends only on the differences between locations and so may be called **advection-based**. A advection-based approach is applicable when the distribution of the increment $\vec{\boldsymbol{x}}(t_r) - \vec{\boldsymbol{x}}(t_s)$, for $s, r \in \mathbb{N}$ and $s < r$, depends only on $\vec{\boldsymbol{x}}(t_r), \vec{\boldsymbol{x}}(t_s), t_r, t_s$ and does not depend on any intermediate time or location. When advection or diffusion changes with time or space as a stochastic process, the distribution of this increment depends on the advection and diffusion at possibly all intermediate times between t_r and t_s. To account for this can be computationally expensive.

There are two ways that $\boldsymbol{\mu}$ and σ can be defined as a stochastic process. Both allow velocity and diffusion to change with time. How parameters can be estimated depends strongly on what approach is taken.

State-Switching

State switching is a common way to allow properties of a movement process to change over time such that those properties are temporally correlated (Morales et al., 2004; Parton, Blackwell, & Skarin, 2016; Zucchini et al., 2016). Conceptually, there are a finite number of possible "states" that a target can inhabit and within each state the target moves according to a different model. A model for how a target's state changes over time is most commonly described by a Markov chain (Morales et al., 2004).

Definition 2.10 (CTMC)**.** A stochastic process $b : [0, T] \to \mathbb{Z}_B$, where $\mathbb{Z}_B = \{1, \ldots, B\}$,

is described by a **continuous-time Markov chain** (CTMC) with transition rates $\rho_{i,j}$ for states $i, j \in \mathbb{Z}_B$ when

1. $b(t)$ is a Markov process: for $t, s \in [0, T]$ where $s < t$,

$$[b(t) \mid \{b(u) : u < s\}] = [b(t) \mid b(s)];$$

2. for $i, j \in \mathbb{Z}_B$, $\rho_{i,j} \geq 0$ when $i \neq j$ and $\rho_{i,i} < 0$;

3. for small $\Delta t > 0$ and $s \in [0, T)$, if $p_{i,j}(t) = [b(s + \Delta t) = i \mid b(s) = j]$ then

$$\frac{\mathrm{d}p_{i,j}}{\mathrm{d}t} = \begin{cases} \rho_{i,j} : i \neq j \\ -\rho_{i,i} : i = j. \end{cases}$$

   Thus, $\rho_{i,j}$ for $i \neq j$ quantifies the instantaneous rate that targets switch from state i to state j. Also, $\rho_{i,i}$ quantifies the rate that targets switch to another state such that $\rho_{i,i} = \sum_{j \neq i} \rho_{i,j}$.

Given a CTMC model (Hanks et al., 2015) for the state-switching process, it can be assumed that advection and diffusion only change through time due to their dependence on which state the target inhabits, thus $\boldsymbol{\mu}(t) = \boldsymbol{\mu}(b(t))$ and $\sigma(t) = \sigma(b(t))$. For brevity, these functions can be re-defined such that $\boldsymbol{\mu} : \mathbb{Z}_B \to \mathbb{R}^2$ and $\sigma : \mathbb{Z}_B \to [0, \infty]$. Thus each state has a fixed advection and diffusion and these quantities only vary by state-switching. This is represented by the CTMC and the SDE:

$$\mathrm{d}\vec{\boldsymbol{x}}(t) = \boldsymbol{\mu}(b(t))\ \mathrm{d}t + \sigma(b(t))\ \mathrm{d}\vec{\boldsymbol{w}}(t).$$

This model is termed a **state-switching advection-diffusion** process.

Parameter estimation of $\boldsymbol{\mu}(1), \ldots, \boldsymbol{\mu}(B), \sigma(1), \ldots, \sigma(B)$ and $\rho_{i,j}$ for $i, j \in \mathbb{Z}_B$ from telemetry data is non-trivial. Between any two times $s, t \in [0, T]$ where $s < t$, there is a non-zero probability that for any $K \in \mathbb{N}$ that the target switched states K times. In other words, the distribution of $\vec{\boldsymbol{x}}(t)$ given $\vec{\boldsymbol{x}}(s)$ does not only depend on $\vec{\boldsymbol{x}}(s)$ but also on all intermediate state switches. An approximate method to fit this full model is discussed in Section 2.4.5 and makes use of the encounter model framework constructed in the next section. Besides this, there are two current approaches taken (Gurarie et al., 2017):

- Approximate the velocity process by computing the average velocity between teleme-

try records. For target i, let

$$\hat{\boldsymbol{v}}_{i,r} = \frac{\vec{\boldsymbol{x}}(t_r) - \vec{\boldsymbol{x}}(t_{r-1})}{t_r - t_{r-1}}$$

for records $r = 2, \ldots, R_i$. This quantity $\hat{\boldsymbol{v}}_{i,r}$ is an estimate of the instantaneous velocity at time t_r. This estimate is best when the temporal resolution of the telemetry data is high. These estimates can be treated as a realisation from the velocity process and can be used to estimate the parameters using a hidden Markov model (Langrock et al., 2012) where in state b, $\hat{\boldsymbol{v}}_{i,r}$ has a Gaussian distribution with mean $\boldsymbol{\mu}(b)$ and variance $\sigma(b)^2(t_r - t_{r-1})$.

- Bayesian data augmentation: Bayesian data augmentation can be used to average over an unobserved stochastic process when the model for the observed data conditional on the unobserved process is known. In this example, if the times of the state switches and the state of the target at each time where known, the PDF of the observed data is Gaussian and can be computed similarly to Theorem 2.3 — effectively splitting the target's path into sections where the state is known not to change and applying this Theorem on each segment, as disjoint segments are conditionally independent. This approach is described in more detail by Blackwell et al. (2016).

Correlated Velocity

There is an alternative approach. Rather than have parameters switch between a finite set of values, they can vary continuously by specifying a further SDE model. In particular, the Ornstein-Uhlenbeck model (Gillespie, 1996; Uhlenbeck & Ornstein, 1930) is preferred in many cases.

Definition 2.11 (OU process)**.** A stochastic process $\boldsymbol{\mu} : [0, T] \to \mathbb{R}^2$ is an **Ornstein-Uhlenbeck** process with mean $\boldsymbol{m}$, diffusion $s > 0$, and attraction α when it has the following Langevin SDE:

$$\mathrm{d}\boldsymbol{\mu}(t) = \alpha(\boldsymbol{m} - \boldsymbol{\mu}(t))\ \mathrm{d}t + s\ \mathrm{d}\vec{\boldsymbol{w}}(t).$$

The Ornstein-Uhlenbeck process is an example of an advection-diffusion model where the process diffuses randomly but also drifts toward a mean $\boldsymbol{m}$. The strength of the attraction of the process to the mean is controlled by α, where in particular if $\alpha = 0$ there is no attraction and the process is a simple diffusion process.

A process where advection follows an OU process and movement given this follows an

advection-diffusion process is termed a correlated velocity model (Gurarie et al., 2017). The state-switching model can be seen as a discrete approximation to the correlated velocity model where the continuous velocity space is split into discrete regions and state switches occur as velocity continuously passes between these regions. Unlike the state-switching approach, a correlated velocity model, surprisingly, has the nice property that for $s, t \in [0, T]$, the joint conditional distribution of $\boldsymbol{x}(t), \boldsymbol{\mu}(t)$ given $\boldsymbol{x}(s), \boldsymbol{\mu}(s)$ can be written in terms of the $\boldsymbol{x}(s)$ and $\boldsymbol{\mu}(s)$ only, in other words, the required integration over all possible intermediate velocities is analytically tractable (Hooten & Johnson, 2017; Hooten, Johnson, et al., 2017). This is due to the velocity and movement processes being Wiener. In particular, maximum likelihood estimation is possible using a Kalman Filter (Johnson et al., 2008). Alternatively, the two approaches mentioned in the last section can be adopted. One can treat $\hat{\boldsymbol{v}}$ as a realisation of the advection OU process and estimate the parameters using Theorem 2.3 (as the OU process is an advection-diffusion process). A Bayesian data augmentation approach has been used for this model (Parton et al., 2016), but found to be computer intensive.

2.3.3 $\mathrm{d}\vec{\boldsymbol{x}}(t) = \boldsymbol{\mu}(\vec{\boldsymbol{x}}(t))\ \mathrm{d}t + \sigma(\vec{\boldsymbol{x}}(t))\ \mathrm{d}\vec{\boldsymbol{w}}(t)$

The functions $\boldsymbol{\mu}$ and σ may be explained by a spatial smooth (Preisler et al., 2004), termed potential functions in the literature (Hooten, Johnson, et al., 2017). The implies that target speed and mean direction of travel are dictated by their position in space where some region may motivate high speeds and more persistent motion compared to regions where targets may prefer to spend more time and so have lower speeds and more tortuous movement.

This approach is intrinsically linked to the survey area $\mathcal{A}$. If the study was solely interested in the relationship between movement dynamics and this particular survey area, then this approach may lead to useful inference on hotspots and centres of attraction. Yet, if the aim were to extend inference to targets in other spaces, then this approach is less useful.

2.3.4 $\mathrm{d}\vec{\boldsymbol{x}}(t) = \boldsymbol{\mu}(\vec{\boldsymbol{x}}(t), t)\ \mathrm{d}t + \sigma(\vec{\boldsymbol{x}}(t), t)\ \mathrm{d}\vec{\boldsymbol{w}}(t)$

This returns us to the full model as specified at the beginning of this section where $\boldsymbol{\mu}$ and σ vary with time and space. Naturally, this includes the combined effects of the models discussed in previous sections. For example, behaviour-switching can depend on target location or a velocity OU process and have parameters that vary with space.

One model that has not been covered by the previous sections is the **biased** advection-diffusion process or home-range model (Okubo & Levin, 2013). This is equivalent to a

movement process being an OU process:

$$\mathrm{d}\vec{\boldsymbol{x}}(t) = \alpha(\boldsymbol{\mu} - \vec{\boldsymbol{x}}(t))\,\mathrm{d}t + \sigma\,\mathrm{d}\vec{\boldsymbol{w}}(t)$$

where $\boldsymbol{\mu}$ is a position in space that the target is attracted to if $\alpha > 0$ and avoids if $\alpha < 0$. This is the mathematical representation of the intuitive concept that a target may have a home or centroid to which it returns periodically. The parameters $\boldsymbol{\mu}$ and σ can be estimated using Theorem 2.3.

Overall, the SDE provides a rich class of models to describe animal movement.

2.4 Encounters

The objective of this Chapter is to estimate the parameters that govern the encounter process: the detection and movement parameters. To do this, sufficient information on these two processes is required. In Theorem 2.1, the PDF for the detection process was constructed for a given target's path. Thus, if the entire path each target followed during the survey is known, this PDF can be used to estimate the detection parameters and the known paths can be treated as telemetry data and so used to estimate the movement parameters as described in Section 2.3. However, encounter surveys are applied to large populations where it is difficult, if not impossible, to track the movements of all targets; thus, this approach is seldom possible. To estimate the encounter parameters, one must construct the PDF of the observed data by averaging over all possible target paths. Importantly, the observed data must contain sufficient information to estimate the encounter parameters. Considering what information is needed naturally leads to survey methods such as distance sampling and spatial capture-recapture.

2.4.1 Encounter Times

For this thesis, at a minimum, it is assumed that the time of every encounter is observed. Times can be observed exactly (continuously) or in intervals. For example, an encounter at time $t = 1.5$ can be recorded exactly as 1.5 or recorded to the nearest unit and so be in the interval $(1, 2)$. The latter is termed **interval** data or **grouped** data, as the times are grouped into intervals. Both types of data can be handled similarly.

Estimating the encounter process from a single detector survey where only detection times are recorded is difficult due to parameter non-identifiability. The difficulty is that encounter times from a single detector are not sufficient to estimate the detection process when target

locations are unknown and targets can move. To illustrate this, suppose targets were immobile. It is then natural to propose that targets that took longer to be encountered and were encountered less often are further from the detector than those that were encountered quickly and often. In other words, the encounter rate for each target provides information on where in space that target resides. In this case, the encounter parameters are identifiable. If, in reality, targets do move assuming that they do not will lead to biased estimates of detection parameters as the detection range and target movement will be confounded. On the other hand, a model with target movement incorporated leads to parameters that are poorly identified. For any proposed detection range and target location, one could propose the target were twice as far away and the detection range twice as wide such that the likelihood of this new scenario is equal to the likelihood of the former. As there is no information on the target's path, the likelihood is flat and has no unique maximum. Thus, for a single detector survey, auxiliary information on the target's location is necessary. In particular, recording the target's location at each encounter is sufficient to avoid this problem. This is covered in the next section. In short, this observation leads naturally to the practice of distance sampling.

Contrary to this, the encounter parameters can be estimated from encounter times alone when multiple detectors are involved in a survey. This is because the location of all detectors is known throughout the entire survey and so detection of the same target by multiple detectors provides indirect information on the target's location. This observation leads to the practice of spatial capture-recapture surveys.

Given the observed encounter times, the marginal PDF of this data is obtained by averaging over all possible target paths. To do this, the space of all paths $\mathcal{C}(\mathcal{A})$ must have an associated probability measure. This probability measure is obtained by specifying a movement model. For this thesis, a Wiener process is the assumed movement model. The Wiener process induces a **Wiener measure** $\mathcal{W}$ on the space $\mathcal{C}(\mathcal{A})$. The Wiener measure is a probability measure and associates a probability with each set of continuous paths. The existence of the Wiener measure follows from the Kolmogorov extension theorem (Tao, 2011).

Definition 2.12. The **Wiener measure** $\mathcal{W}_{\boldsymbol{\mu},\sigma}$ on a space $\mathcal{A}$ is the probability measure induced on $\mathcal{C}(\mathcal{A})$ by a Wiener process with mean $\boldsymbol{\mu}$ and diffusion σ. An integral with respect to this measure is called a **Wiener path integral**. The dependence on $\boldsymbol{\mu}$ and σ will often be omitted.

Thus, the PDF of observed encounter times $\boldsymbol{t}_i$ for target i is given by the path integral:

$$\int [\boldsymbol{t}_i \mid \vec{\boldsymbol{x}}_i] \ \mathrm{d}\mathcal{W}(\vec{\boldsymbol{x}}_i).$$

The path integral is usually not analytically tractable and is difficult to compute numerically. In Section 2.5, a computational algorithm is developed to perform path integration.

2.4.2 Observed Location

In the previous section, the marginal PDF given was for the observed detection times only and so the entire target's path was presumed to be unknown. In some encounter surveys, the target's location may be observed at certain times, for example, when an encounter occurs. The previous section discussed that the target's location must be partially observed when only a single detector is used, as otherwise there is insufficient information to estimate the encounter parameters.

Suppose for the e^{th} encounter between detector j and target i that record $\boldsymbol{r}_{i,j,e}$ is the observed location of the target at the time of the encounter $t_{i,j,e}$. The PDF of the observed data is again a path integral, except only over those paths that pass through the locations $\boldsymbol{r}_{i,j,e}$ at the times $t_{i,j,e}$:

$$\int [\boldsymbol{r}_i, \boldsymbol{t}_i \mid \vec{\boldsymbol{x}}_i] \ \mathrm{d}\mathcal{W}(\vec{\boldsymbol{x}}_i)$$

Since $[\boldsymbol{r}_i, \boldsymbol{t}_i \mid \vec{\boldsymbol{x}}_i] = [\boldsymbol{t}_i \mid \vec{\boldsymbol{x}}_i][\boldsymbol{r}_i \mid \vec{\boldsymbol{x}}_i]$ the path integral can be equivalently written as

$$\int_{\chi_i} [\boldsymbol{t}_i \mid \vec{\boldsymbol{x}}_i] \ \mathrm{d}\mathcal{W}(\vec{\boldsymbol{x}}_i)$$

where $\chi_i \subset \mathcal{C}(\mathcal{A})$ is defined by

$$\chi_i = \{\vec{\boldsymbol{x}}_i \in \mathcal{C}(\mathcal{A}) : \vec{\boldsymbol{x}}_i(t_{i,j,e}) = \boldsymbol{r}_{i,j,e} \text{ for all } j, e\}$$

This approach assumes that the target's location is observed without error. In Section 6.2.1, it is discussed how a measurement error model can be incorporated.

2.4.3 First Encounters

Theorem 2.1 gives the PDF for $t_{i,j,1}$, the time of the first encounter between detector j and target i. For two reasons, one may wish to restrict attention to only the first encounters:

1. The encounter process of interest may fundamentally change after the first encounter, for example, when a human observer encounters a target animal in a wild population, the observer is more likely to encounter that animal again soon after because the observer has foreknowledge of that animal's location and behaviour. For this reason, the encounter process changes. Nevertheless, the interest may remain in the initial encounter rate since often interest is in how encounters with previously unseen targets arise. Hence, only the first encounters between detectors and targets provide the relevant data on this process.

2. Repeated encounters with a target are not always identifiable: a detector cannot ascertain whether any two encounters where with the same target or different targets unless targets are continuously tracked. This means that the framework presented here cannot be fully applied as it relies on encounters being linked by their target ID, i. Nevertheless, some information can be obtained by making the requirement that targets be tracked after their first encounters. Similar to the first point, subsequent "encounters" are no longer relevant to the encounter process of interest, but this tracking ensures that any new encounters are indeed first encounters with new targets and not repeated encounters with a previously encountered target.

In either case, the PDF of all first encounters, denoted $\boldsymbol{t}_{i,\cdot,1}$, is again a path integral with the exception that only the target's movement up to the first encounter is of relevance:

$$[\boldsymbol{t}_{i,\cdot,1}] = \int [\boldsymbol{t}_{i,\cdot,1} \mid \vec{\boldsymbol{x}}_i] \ \mathrm{d}\mathcal{W}(\vec{\boldsymbol{x}}_i).$$

The analogous PDF for observed locations with first encounters can be derived as in the previous section.

2.4.4 State-switching

In Section 2.3.2, state-switching (or behaviour-switching) movement models were discussed. For encounter models, the state can affect both the movement and detection of targets; the Wiener measure, $\mathcal{W}_s$, and encounter intensity, λ_s, depends on the state. Here, not only is the target's movement path unobserved, but also the state history of the target. Hence, one must integrate over all possible movement paths and state histories jointly. This is termed a state-switching path integral. For notational brevity, let $\vec{\boldsymbol{x}}$ be a three-dimensional process such that $\vec{\boldsymbol{x}}(t) = (\boldsymbol{x}, s)$ when at time t the two-dimensional location of the target is $\boldsymbol{x}$ and the state of the target is s. The PDF of observed encounter times is then analogous to that

given in Section 2.4.1:

$$\int [\boldsymbol{t}_i \mid \vec{\boldsymbol{x}}_i] \ \mathrm{d}\mathcal{W}_s(\vec{\boldsymbol{x}}_i).$$

The same adjustment can be made for the PDFs of observed locations and first encounters. Furthermore, as with observed locations, if the state of the target is observed at some times, the path integral can be restricted to those state histories that conform with the observed state sequence.

2.4.5 Telemetry

A telemetry survey was defined in Definition 2.7: at a collection of, possibly irregularly spaced, times the location of a target is observed. Often in telemetry studies, the target's location is presumed to be observed with certainty: the reason why the location was observed is ignored. For example, GPS tagging of cetaceans often records the animal's locations when the cetacean is at the sea surface and not when diving; thus, the records of location are irregularly spaced due to the animal's behaviour. This can be viewed, conceptually, as an encounter process: the animal moves and the time and location of the animal is recorded whenever the tag is available. An encounter corresponds to a telemetry recording. In other words, telemetry surveys are a type of encounter survey where the target's location is recorded during each encounter. The difference is that the intensity of the encounter process often does not depend on distance but rather on target behaviour or the technology used to record target locations. Often, telemetry studies ignore the detection process: telemetry records are presumed to occur with certainty, that is, with a fixed intensity. The effect of this on inference is unknown.

The path taken by a target between two telemetry records is unknown and so a path integral is required. In Section 2.3, parameter estimation of the movement process given telemetry data was discussed. In some cases, a advection-based approach was possible as the path integral was analytically tractable and reduced to a Gaussian PDF (Hooten, Johnson, et al., 2017). Yet, in more complex cases, this was not possible and an approximate advection-based approach was discussed (Gurarie et al., 2017). A location-based approach using a Kalman filter was used in some cases (Johnson et al., 2008), but the discussion highlighted that when behaviour-switching was involved only an approximate method (e.g. multiple imputation Scharf et al. (2017)) or a fully Bayesian data augmentation method was possible (Blackwell et al., 2016). Here, as a telemetry survey is a special case of the encounter survey, the PDF can be written as a path integral and so the computational algorithm in the next section provides an alternative to the data augmentation approach.

The PDF of telemetry records $\boldsymbol{r}_1, \ldots, \boldsymbol{r}_R$ on target i at times $t_1, \ldots, t_R$ is

$$\int [\boldsymbol{r}_1, \ldots, \boldsymbol{r}_R \mid \vec{\boldsymbol{x}}_i] \, \mathrm{d}\mathcal{W}(\vec{\boldsymbol{x}}_i)$$

Notice that this is equivalent to the Wiener measure of the set

$$\chi_i = \{\vec{\boldsymbol{x}}_i \in \mathcal{C}(\mathcal{A}) : \vec{\boldsymbol{x}}_i(t_k) = \boldsymbol{r}_k \text{ for all } k\}$$

Furthermore, the state-switching path integral in Section 2.4.4 can be used to formulate the PDF of a state-switching continuous-time movement model. Computational algorithms to fit such models by maximum likelihood are computationally costly (Pedersen et al., 2011); thus, the computational algorithms in the next section can contribute to their improvement.

2.4.6 Covariates

Parameters may vary across time, space, or some other dimension. All parameters in the encounter model can change through time and space, furthermore they can change according to other variables. For example, a survey area may cover land with varying elevation and the target movement and detection process may change according to the elevation. Variables that are suspected to effect the detection or movement process are called **covariates**. Other common covariates are weather, visibility, season, vegetation cover, and detector.

The treatment of covariates in encounter models can be similar to generalised linear models (McCullagh & Nelder, 1989). Consider a parameter θ and covariates $c_b : \mathcal{A} \times [0, T] \to \mathbb{R}$ for $b = 1, \ldots, B$ where B is the number of covariates. There is often no *a priori* justification for assuming a specific functional relationship between θ and any covariates, thus for mathematical convenience, a polynomial relationship is assumed. This can be somewhat justified by arguing that the true functional relationship between θ and c_b can be approximated by its Taylor series expansion.

In short, it is proposed that the parameter θ in spatial location $\boldsymbol{x}$ at time t has a value such that

$$g(\theta) = \beta_0 + \beta_1 c_1(\boldsymbol{x}, t) + \ldots + \beta_B c_B(\boldsymbol{x}, t)$$

where $g : \Omega \to \mathbb{R}$ is a function that maps **onto** $\mathbb{R}$, termed a **link function**, and Ω is the set of values θ can obtain. Thus, for instance, if θ were strictly positive, $\Omega = (0, \infty)$, a common choice for g would be the log function. The right hand side is termed the **linear predictor** and $\beta_1, \ldots, \beta_B$ are new parameters that quantify the strength of relationship

between θ and each covariate. Notice, θ itself is no longer directly estimated, it is derived from the estimates of $\beta_0, \ldots, \beta_B$.

Including covariates not only often improves inference by accounting for known realities, but it also links the encounter process to processes occurring in the real world, a task that is also often of interest. Investigating why targets move to particular areas in space or determining in which conditions detection is most likely can be addressed by including relevant covariates.

2.5 Path Integration

In the previous section, the PDF of the encounter data involved a Wiener path integral. This path integral often is not analytically tractable. In this section an approximation to the path integral is developed and an efficient computational algorithm discussed that makes the encounter model computationally tractable. For this section, consider a function $f : \mathcal{C}(\mathcal{A}) \to \mathbb{R}$ that maps target paths to real numbers. In the encounter PDF this function is the exponential of the accumulated intensity or a product of this exponential with the intensity function. Here, the path integration of an integrable function f is considered:

$$\int f(\vec{\boldsymbol{x}}) \, \mathrm{d}\mathcal{W}(\vec{\boldsymbol{x}}).$$

The simplest "Wiener" process is the ideal free gas process: $\mathrm{d}\vec{\boldsymbol{x}} = \boldsymbol{\mu} \, \mathrm{d}t$ where $\boldsymbol{\mu} = \nu(\cos(\theta), \sin(\theta))$ for speed parameter ν and unknown direction θ. This is a deterministic movement process. Every path is uniquely determined by its starting point $\vec{\boldsymbol{x}}(0)$ and direction of travel θ. The path integral of the function f can then be reduced to a finite-dimensional integral:

$$\int f(\vec{\boldsymbol{x}}) \, \mathrm{d}\mathcal{W}(\vec{\boldsymbol{x}}) = \frac{1}{2\pi|\mathcal{A}|} \int_{\mathcal{A}} \int_0^{2\pi} f(p(\boldsymbol{x}_0, \theta)) \, \mathrm{d}\theta \, \mathrm{d}\boldsymbol{x}_0$$

where $p(\boldsymbol{x}_0, \theta)$ is the path that starts at the point $\boldsymbol{x}_0$ and travels in direction θ. This integral can then be computed using standard numerical integration techniques (Gonzalez, 2018).

For a Wiener process where $\sigma > 0$, the movement involves a diffusive component, the path integral cannot be easily simplified. One approach to this would be a Monte Carlo approximation to the path integral where many possible paths are simulated from the movement process (Scharf et al., 2017). This may be a reasonable approach when the primary goal is to compute the path integral a few times. Looking ahead, to estimate the

parameters of the encounter process the likelihood of the encounter parameters will need to be computed many times and each computation of the likelihood may contain multiple path integrations. Approximating each path integral by a Monte Carlo method will be prohibitively expensive in computational time. Therefore, here, we adopt a quadrature style method that reduces the path integral to a finite dimensional integral.

The idea can be described in two steps:

1. identify a set of *simple* paths where the path integration can be reduced to a finite-dimensional integration;
2. show that every path can be approximated by a *simple* path and that the path integral can be approximated by the finite-dimensional integral of a corresponding simple path.

2.5.1 Simple Paths

Intuitively, a simple path is a path on a grid that is placed over space such that targets jump from one position on the grid to the next at a finite number of times.

Definition 2.13 (Simple path)**.** A path $\vec{\boldsymbol{x}} : [0, T] \to \mathcal{A}$ is a **simple path** at **time-step** Δt when for $\tilde{T} = \lfloor \frac{T}{\Delta t} \rfloor$ there exists $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{\tilde{T}} \in \mathcal{A}$ where $\vec{\boldsymbol{x}}(t) = \boldsymbol{x}_i$ when $(i-1)\Delta t < t < i\Delta$. In short, the target only changes position at each time-step and not between. Let $\mathcal{S}(\Delta t)$ be the set of all simple paths at time-step Δt across the space $\mathcal{A}$.

Theorem 2.4 (Simple path integral)**.**

$$\int_{\mathcal{S}(\Delta t)} f(\vec{\boldsymbol{x}}) \, \mathrm{d}\mathcal{W}(\boldsymbol{x}) = \int_{\mathcal{A}^{\tilde{T}}} f(\vec{\boldsymbol{x}})[\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{\tilde{T}}] \, \mathrm{d}\boldsymbol{x}_1 \ldots \boldsymbol{x}_{\tilde{T}}.$$

This reduces the path integral over all simple paths to a finite-dimensional integral. Thus, in theory, the survey area $\mathcal{A}$ could be divided into a grid $\mathcal{G}(\Delta x)$ of points that are separated by a distance Δx and the integral approximated by a sum.

The difficulty is that this finite-dimensional integral is over a very high-dimensional space. The curse of dimensionality makes a quadrature approach infeasible. The integral can, however, be further simplified when f is of a particular form, termed Markov-separable.

Definition 2.14 (Markov-separable)**.** An integrable function $f : \mathcal{S}(\Delta t) \to \mathbb{R}$ is **Markov-separable** when there exists integrable functions $f_i : \mathcal{A} \to \mathbb{R}$ for $i = 1, \ldots, \tilde{T}$ such that

$$f(\vec{\boldsymbol{x}}) = \prod_{i=1}^{\tilde{T}} f_i(\vec{\boldsymbol{x}}(i\Delta t))$$

That is, the function can be separated into components that are each functions of only a single target location at a particular time.

Theorem 2.5. For a Markov-separable function $f : \mathcal{S}(\Delta t) \to \mathbb{R}$,

$$\int_{\mathcal{S}(\Delta t)} f(\vec{\boldsymbol{x}}) \, \mathrm{d}\mathcal{W}(\vec{\boldsymbol{x}}) = \int_{\mathcal{A}^{\tilde{T}}} [\boldsymbol{x}_0] f_0(\boldsymbol{x}_0) \prod_{i=1}^{\tilde{T}} f_i(\boldsymbol{x}_i)[\boldsymbol{x}_i \mid \boldsymbol{x}_{i-1}] \, \mathrm{d}\boldsymbol{x}_0 \ldots \boldsymbol{x}_{\tilde{T}}$$

The path integral of a Markov-separable function has the same form as the likelihood of a hidden Markov process in continuous time with a continuous state space (Cappé, Moulines, & Rydén, 2005). The Wiener process is the hidden Markov process and the function f is akin to the PDF of the "observation process". Here, f need not be a PDF, but the analogy is convenient. Computing the finite-dimensional integral by quadrature is equivalent to approximating the continuous-time, continuous-space process with a discrete-time, discrete-space hidden Markov model. Recognising this means that the efficient **forward algorithm** (Zucchini et al., 2016) that exists to compute the likelihood of a HMM can be used to compute an approximation to the Wiener path integral of a Markov-separable function over a given grid and time-step.

2.5.2 Hidden Markov Model Quadrature

An encounter process is not an example of a hidden Markov process because the detection process depends on the entire path a target takes. Nonetheless, the previous section shows that the path integral of a Markov-separable function over the space of simple paths can be written as the likelihood of a hidden Markov model. If targets only moved in simple paths, then the encounter process would be a hidden Markov process because the encounter intensity between any two time-steps would depend only on the target's location in that time-step, which is a constant location. The idea is that if the time-step and spatial grid resolution is reduced to zero, then the integration over simple paths will approximate the integration over all continuous paths.

Here, the path integral over all simple paths of a Markov-separable function f is computed by quadrature using the methodology of hidden Markov models. For this reason, this approach is termed **hidden Markov model quadrature**. Consider a given time-step Δt and let the survey area $\mathcal{A}$ be covered in a grid $\mathcal{G}(\Delta x)$. For this approximation, targets jump, at each time-step, from one point on the grid to another. This discrete-time jump process is what will approximate the continuous space-time process as Δt and Δx tend to zero.

A hidden Markov model likelihood has three components (Zucchini et al., 2016):

1. an initial distribution represented by a $1 \times |\mathcal{G}|$ row vector $\boldsymbol{\gamma}_0$ with i^{th} entry given by $[\boldsymbol{x}(0) = \boldsymbol{g}_i]$ where $\boldsymbol{g}_i$ is the location of grid point i;
2. a $|\mathcal{G}| \times |\mathcal{G}|$ diagonal matrix $\boldsymbol{F}_{\tilde{t}}$ with i^{th} diagonal given by $f_{\tilde{t}}(\boldsymbol{g}_i)$ where $\boldsymbol{g}_i$ is the i^{th} grid point. This is termed here the **integrand** matrix;
3. a **transition rate matrix** $\boldsymbol{R}_{\tilde{t}}$ for time-step $\tilde{t}$ of dimension $|\mathcal{G}| \times |\mathcal{G}|$ where the $(i,j)^{\text{th}}$ entry is the rate, denoted $\rho_{i,j,t}$, that a target moves from grid point i to grid point j.

The initial distribution vector and the integrand matrix are convenient objects for computing the HMM likelihood, and their contents have been described in the previous section. The transition rate matrix, $\boldsymbol{R}_{\tilde{t}}$, is a new concept.

Definition 2.15 (Transition Rate)**.** The **transition rate** (Hanks et al., 2015) of a target between locations $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ is ρ when

$$\frac{\partial[\vec{\boldsymbol{x}}(t) = \boldsymbol{y}_2 \mid \vec{\boldsymbol{x}}(u) = \boldsymbol{y}_1]}{\partial t} = \rho.$$

Thus, intuitively, ρ quantifies the instantaneous rate targets move from one position to another.

Theorem 2.6 (Wiener Transitions)**.** Let $\phi : \mathcal{A} \times [0,T] \to [0,\infty]$ be a function such that $\phi(\boldsymbol{x},t)$ is the PDF of the random variable $\vec{\boldsymbol{x}}(t)$ evaluated at the point $\boldsymbol{x}$.

(a) The **Eulerian** equation (Turchin, 1998) for the Wiener process is given by

$$\frac{\partial \phi}{\partial t} = -\nabla \cdot \boldsymbol{\mu}(\boldsymbol{x},t)\phi + \frac{\sigma(\boldsymbol{x},t)^2}{2}\nabla^2\phi$$

where ∇ is the gradient operator on $\mathbb{R}^2$ and ∇^2 is the Laplacian on $\mathbb{R}^2$.

(b) For a time t and constants $\Delta x > 0$ and $\Delta t > 0$, when $\boldsymbol{\Delta} = (\Delta x, 0)$ or $(0, \Delta x)$,

$$\begin{aligned}\phi(\boldsymbol{x}, t+\Delta t) = &(s(\boldsymbol{x},t) - m(\boldsymbol{x},t))\phi(\boldsymbol{x}+\boldsymbol{\Delta},t)\Delta t \\ &+ (s(\boldsymbol{x},t) + m(\boldsymbol{x},t))\phi(\boldsymbol{x}-\boldsymbol{\Delta},t)\Delta t \\ &+ (1 - 2s(\boldsymbol{x},t))\phi(\boldsymbol{x},t)\Delta t\end{aligned}$$

where $s = \dfrac{\sigma(\boldsymbol{x},t)^2}{2\Delta x^2}$ and $m = \dfrac{\mu(\boldsymbol{x},t)}{2\Delta x}$.

Proof.

(a) The **Eulerian** equation can be seen as describing the PDF for a single target's path; alternatively, it can be seen to describe the entire population of targets and where they spend their time. This equation is also called the **Fokker-Planck** equation. A full derivation of the Fokker-Planck equation from the Langevin SDE is given in Méndez, Campos, and Bartumeus (2013).

(b) This is derived by performing second-order central finite differences on the Eulerian equation in (a). A proof is given by Mitchell and Griffiths (1980) and the approximation used by Hooten, Johnson, et al. (2017); Hooten and Wikle (2008); Pedersen et al. (2011).

□

This gives the approximate transition rates for $\boldsymbol{R}_{\tilde{t}}$ in the negative direction $(s-m)$ and positive direction $(s+m)$ for each dimension. Given this, the HMM quadrature can be stated.

Theorem 2.7 (HMM quadrature)**.** For a Markov-separable function $f : \mathcal{S}(\Delta t) \to \mathbb{R}$,

$$\int_{\mathcal{S}(\Delta t)} f(\vec{\boldsymbol{x}}) \, \mathrm{d}\mathcal{W}(\vec{\boldsymbol{x}}) = \lim_{\Delta x \to 0} \boldsymbol{\gamma}_0 \boldsymbol{F}_0 \left(\prod_{\tilde{t}=1}^{\tilde{T}} \boldsymbol{\Gamma}_{\tilde{t}} \boldsymbol{F}_{\tilde{t}} \right) \boldsymbol{\epsilon}_1$$

where $\boldsymbol{\epsilon}_1$ is a $|\mathcal{G}| \times 1$ vector of ones and $\boldsymbol{\Gamma}_{\tilde{t}} = \exp\left(\boldsymbol{R}_{\tilde{t}} \Delta t\right)$ is the **transition probability matrix** over a time-step Δt.

Proof. By Theorem 2.5,

$$\int_{\mathcal{S}(\Delta t)} f(\vec{\boldsymbol{x}}) \, \mathrm{d}\mathcal{W}(\vec{\boldsymbol{x}}) = \int_{\mathcal{A}^{\tilde{T}}} [\boldsymbol{x}_0] f_0(\boldsymbol{x}_0) \prod_{i=1}^{\tilde{T}} f_i(\boldsymbol{x}_i) [\boldsymbol{x}_i \mid \boldsymbol{x}_{i-1}] \, \mathrm{d}\boldsymbol{x}_0 \ldots \boldsymbol{x}_{\tilde{T}}$$

For a given grid $\mathcal{G}(\Delta x)$, this finite-dimensional integral can be approximated by quadrature with the multi-dimensional sum

$$\sum_{\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{\tilde{T}}} [\boldsymbol{x}_0] f_0(\boldsymbol{x}_0) \prod_{\tilde{t}=1}^{\tilde{T}} f_{\tilde{t}}(\boldsymbol{x}_i) [\boldsymbol{x}_{\tilde{t}} \mid \boldsymbol{x}_{\tilde{t}-1}].$$

This is the likelihood of a hidden Markov model (Zucchini et al., 2016) and so can be written as the matrix product given in the result. □

The computation of this matrix product is achieved in HMM methods using the forward algorithm. The algorithm iterates the following steps $\tilde{T}$ times beginning with $t = 0$:

1. $\boldsymbol{\psi} \leftarrow \boldsymbol{\gamma}_t \boldsymbol{F}_t$
2. $\boldsymbol{\gamma}_{t+1} \leftarrow \boldsymbol{\psi} \boldsymbol{\Gamma}_t$
3. $t \leftarrow t + 1$

The value of the approximation to the path integral is then given by the sum of $\boldsymbol{\gamma}_{\tilde{T}}$.

In many applications of HMMs, the forward algorithm is an efficient and practical algorithm. Here, the approximation to the Wiener path integral is best when Δx is small and thus the grid $\mathcal{G}$ is large. This requires that the transition probability matrices and integrand matrices be of very large dimension. Thus, large amounts of computer memory and computation time is required to execute the forward algorithm. For the levels of accuracy necessary in many applications, this obstacle can prohibit practical use of this approximation (Pedersen et al., 2011). Therefore, in the next two sections, adjustments to the forward algorithm are introduced that reduce the computational burden of the high-dimensional matrices involved, and makes this approximation a practical approach.

2.5.3 Matrix Sparsity

There are two matrices required in each iteration of the forward algorithm: $\boldsymbol{F}_t$ and $\boldsymbol{R}_t$. Each of these matrices have a sparsity property that can be exploited to speed computation and reduce computer storage requirements. Sparsity is the property that a large proportion of matrix entries are zero.

The matrix $\boldsymbol{F}_t$ is by definition **diagonal**, an extreme form of sparsity. Thus, only the diagonal of the matrix need be stored. Furthermore, multiplication of a vector with a diagonal matrix is equivalent to the element-wise product of the vector and the diagonal itself. Thus, matrix multiplication by $\boldsymbol{F}_t$ is reduced to element-wise multiplication.

The sparsity of the transition rate matrix $\boldsymbol{R}_t$ depends on the parameters of the target movement model. Intuitively, over a short time, there is a limit to how far targets can move. Hence, in practice, instantaneous movement rates decay quickly and so $\boldsymbol{R}_t$ is sparse. Thus, in many applications, it is efficient to store $\boldsymbol{R}$ as a sparse matrix, most commonly in column-major form. Furthermore, if it is assumed that targets cannot move more than one grid cell instantaneously, that is, transitions can only occur between neighbouring cells in small time steps, then the matrix $\boldsymbol{R}_t$ is highly sparse. Multiplication of a vector with a

sparse matrix is more efficient when sparsity is high. This is the same approach taken in the approximation of point process models by Gauss Markov random fields (Illian et al., 2013).

2.5.4 Krylov Approximation

The forward algorithm requires the multiplication of a vector with the matrix exponential of the transition rate matrix: $\boldsymbol{\psi} \exp(\boldsymbol{R}_t \Delta t)$. Computation of the matrix exponential is a costly operation for large matrices whether sparse or not. In this section, an approximation to this matrix-vector product is introduced that removes the need to explicitly compute the matrix exponential. This approximation is termed the **Krylov** approximation (Saad, 1992a; Sidje, 1998).

Consider the space of all $n \times n$ matrices, $\mathcal{M}_n$, and a scalar function $f : \mathbb{R} \to \mathbb{R}$. By Taylor's theorem, this scalar function can be written as

$$f(x) = a_0 + a_1 x + \frac{1}{2} a_2 x^2 + \ldots = \sum_{k=0}^{\infty} \frac{a_k}{k!} x^k$$

A corresponding matrix function $f : \mathcal{M}_n \to \mathcal{M}_n$ can be defined as

$$f(\boldsymbol{X}) = a_0 + a_1 \boldsymbol{X} + \frac{a_2}{2} \boldsymbol{X}^2 + \ldots = \sum_{k=0}^{\infty} \frac{a_k}{k!} \boldsymbol{X}^k$$

for a matrix $\boldsymbol{X} \in \mathcal{M}_n$. In particular, the matrix exponential is $\exp(\boldsymbol{X}) = \sum_{k=0}^{\infty} \frac{\boldsymbol{X}^k}{k!}$.
The idea is to approximate these matrix functions in a vector space of low dimension m. In particular, the interest lies in the product $\boldsymbol{w} = \boldsymbol{v} f(\boldsymbol{X})$ for a matrix $\boldsymbol{X}$ and row vector $\boldsymbol{v}$. A natural approach would be to use a truncated Taylor series:

$$\boldsymbol{w} = \boldsymbol{v} f(\boldsymbol{X}) \approx a_0 \boldsymbol{v} + a_1 \boldsymbol{v} X + a_2 \boldsymbol{v} \boldsymbol{X}^2 + \ldots a_m \boldsymbol{X}^m = \tilde{\boldsymbol{w}}.$$

Effectively, the vector $\tilde{\boldsymbol{w}}$ lies in a vector space with at most m dimensions and is an approximation of the vector $\boldsymbol{w}$. As the dimension m is increased, the approximation will improve.

Definition 2.16 (Krylov Subspace)**.** For a $n \times n$ matrix $\boldsymbol{X}$ and $1 \times n$ vector $\boldsymbol{v}$, the Krylov

subspace of dimension m, denoted $\mathcal{K}_m(\boldsymbol{v}, \boldsymbol{X})$ is the vector space spanned by the vectors

$$\boldsymbol{v}, \boldsymbol{v}\boldsymbol{X}, \boldsymbol{v}\boldsymbol{X}^2, \ldots, \boldsymbol{v}\boldsymbol{X}^m$$

.

In practice the basis vectors $\boldsymbol{v}\boldsymbol{X}^k$ for $k = 1, \ldots, m$ are a poor choice for a vector basis because they are likely to be almost linearly dependent. Thus, the Arnoldi process (Saad, 1992b) is used to create a linearly independent basis for the Krylov subspace. By doing so, for any given level of accuracy, fewer matrix-vector products are required (Moler & Van Loan, 2003). This process depends only on matrix-vector products of $\boldsymbol{X}$ and $\boldsymbol{v}$.

Theorem 2.8 (Krylov)**.** If $\boldsymbol{v}$ is $1 \times n$ vector, $\boldsymbol{X}$ is a $n \times n$ matrix, and $f_m(\boldsymbol{X}) = \sum_{k=1}^{m} \frac{a_k}{k!} \boldsymbol{X}^k$, then

$$\boldsymbol{v} f(\boldsymbol{X}) \approx \|\boldsymbol{v}\| \boldsymbol{\epsilon}_1 f_m(\tilde{\boldsymbol{X}}) \boldsymbol{P}_m$$

where

- $\boldsymbol{P}_m$ is a matrix with columns $\mathcal{B} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_m)$ where $\mathcal{B}$ is the basis produced by the Arnoldi process for the Krylov subspace $\mathcal{K}_m(\boldsymbol{v}, \boldsymbol{X})$. In particular, $\boldsymbol{b}_1 = \|\boldsymbol{v}\|^{-1} \boldsymbol{v}$;
- $\tilde{\boldsymbol{X}}$ is the approximation of $\boldsymbol{X}$ in the Krylov subspace, that is $\tilde{\boldsymbol{X}} = \boldsymbol{P}_m \boldsymbol{X} \boldsymbol{P}_m^{-1}$;
- $\boldsymbol{\epsilon}_1$ is a vector with first entry 1 and all others zero.

Proof. It is assumed the matrix function f has a corresponding scalar form with Taylor series expansion: $f(x) = a_0 + a_1 x + a_2 x^2 + \ldots$ for $a_0, a_1, a_2, \ldots \in \mathbb{R}$. Given this,

$$\begin{aligned}
\boldsymbol{v} f(\boldsymbol{X}) &\approx \sum_{k=0}^{m} a_k \boldsymbol{v} \boldsymbol{X}^k \\
&= \sum_{k=0}^{m} \|\boldsymbol{v}\| a_k \boldsymbol{\epsilon}_1 \boldsymbol{P}_m \boldsymbol{X}^k \qquad \text{since } \boldsymbol{v} = \|\boldsymbol{v}\| \boldsymbol{\epsilon}_1 \boldsymbol{P} \\
&= \sum_{k=0}^{m} \|\boldsymbol{v}\| a_k \boldsymbol{\epsilon}_1 \boldsymbol{P}_m \boldsymbol{X}^k \\
&= \sum_{k=0}^{m} \|\boldsymbol{v}\| a_k \boldsymbol{\epsilon}_1 \boldsymbol{P}_m (\boldsymbol{P}_m^{-1} \tilde{\boldsymbol{X}} \boldsymbol{P}_m)^k \\
&= \sum_{k=0}^{m} \|\boldsymbol{v}\| a_k \boldsymbol{\epsilon}_1 \boldsymbol{P}_m \boldsymbol{P}_m^{-1} \tilde{\boldsymbol{X}}^k \boldsymbol{P}_m
\end{aligned}$$

$$= \sum_{k=0}^{m} \|\boldsymbol{v}\| a_k \boldsymbol{\epsilon_1} \tilde{\boldsymbol{X}}^k \boldsymbol{P}_m$$
$$= \|\boldsymbol{v}\| \boldsymbol{\epsilon}_1 f_m(\tilde{\boldsymbol{X}}) \boldsymbol{P}_m.$$

□

Notice, the Krylov approximation is built upon the vector $\boldsymbol{v}$ and not only the matrix $\boldsymbol{X}$ and function f. Thus, a separate basis must be computed for every vector $\boldsymbol{v}$ one wishes to compute the matrix-vector product for.

The accuracy of the Krylov approximation has been investigated for the matrix exponential in particular (Saad, 1992a). An important note is that the accuracy of the method depends on the spectral radius of matrix $\boldsymbol{X}$. Matrices with larger eigenvalues are approximated less well than those with smaller eigenvalues. A simple strategy to overcome this is to scale the matrix $\boldsymbol{X}$ by a constant, thus reducing its eigenvalues; after the Krylov approximation is complete, the solution can be re-scaled to the original magnitudes.

When approximating the matrix exponential, another strategy is commonly used to improve the error of the approximation: time-stepping (Sidje, 1998). Time stepping is the same as the scaling strategy described above, but arises in a particular situation that is of relevance here. Time-stepping is used when approximating the function $\exp(\boldsymbol{X}\Delta t)$ for a matrix $\boldsymbol{X}$ and a scalar $\Delta t > 0$. This functional form arises when computing the transition probability matrix over a given time-step Δt. When $\boldsymbol{X}$ has large spectral radius or Δt is large, the Krylov approximation can be poor. Time-stepping uses the property that $\exp(\boldsymbol{X}t_1)\exp(\boldsymbol{X}t_2) = \exp(\boldsymbol{X}(t_1 + t_2))$. Hence, the matrix exponential can be computed at smaller time-steps δt and cumulatively multiplied. The value of δt can be chosen such that the error in the approximation is low (Saad, 1992a).

2.5.5 Toeplitz Matrices

The Krylov approximation reduces computationally intensive matrix operations, such as the matrix exponential, to matrix-vector products. Sparse matrix storage and computation make these matrix-vector products calculable efficiently. A further efficiency can be made when the form of the matrices involved is known beforehand. In particular, matrices describing target movement are often of Toeplitz or block-Toeplitz form.

Definition 2.17 (Toeplitz)**.** A matrix is **Toeplitz** when every diagonal is constant, that

is, for an $n \times n$ matrix $\boldsymbol{X}$ with (i,j)th entry $a_{i,j}$:

$$a_{i,j} = a_{i+1,j+1}$$

for $i, j = 1, \ldots, n-1$.

For target movement, the transition rate matrix is Toeplitz when the movement model is **translation invariant**: neither the advection nor diffusion depends on space. Recognising the matrix in a matrix-vector product is Toeplitz brings two efficiencies. First, the first row and column of the matrix contain all the information required and so only these must be stored, making use of computer memory more efficient. For very large matrices, this could be the difference between being able to use the matrix or not. The second efficiency is that the matrix-vector product can be performed quickly using a fast Fourier transform (Marple & Marple, 1987).

Theorem 2.9 (Toeplitz FFT)**.** If $\boldsymbol{X}$ is an $n \times n$ Toeplitz matrix with first row $\boldsymbol{r}$ and first column $\boldsymbol{c}$ and $\boldsymbol{v}$ is a $1 \times n$ vector, then

$$\boldsymbol{v}X = \mathcal{F}^{-1}(\mathcal{F}(\boldsymbol{v}_0) \cdot \mathcal{F}(\boldsymbol{w}))\boldsymbol{\epsilon}$$

where $\mathcal{F}$ is the discrete two-dimensional Fourier transform, $\boldsymbol{w}$ is $1 \times 2n$ vector of the form $(\boldsymbol{c}, 0, \boldsymbol{s})$, $\boldsymbol{s}$ is the reverse of the first row of $\boldsymbol{X}$ after its first element is excluded, $\boldsymbol{v}_0 = (\boldsymbol{v}, 0)$, and $\boldsymbol{\epsilon}$ is a diagonal matrix with first n entries on the diagonal equal to one and all other entries zero.

Proof. Define a matrix $\boldsymbol{Y}$ such that it has first column $\boldsymbol{w}$ and each subsequent column is a circular convolution of the last, that is, entry $y_{i,j} = y_{i',j'}$ where $i' = i - 1 \bmod n$ and $j' = j - 1 \bmod n$. This is termed a **circulant matrix**. By the circular convolution theorem,

$$\boldsymbol{v}_0\boldsymbol{Y} = \mathcal{F}^{-1}(\mathcal{F}(\boldsymbol{v_0})\mathcal{F}(\boldsymbol{w}))$$

The result follows from the observation that $\boldsymbol{X}$ is a sub-matrix of $\boldsymbol{Y}$ (it is the first n rows and columns) and so the first n entries of $\boldsymbol{v}_0\boldsymbol{Y}$ equal $\boldsymbol{v}\boldsymbol{X}$. □

2.5.6 Convergence

Section 2.5.1 defined the concept of a simple path and recent sections presented how path integration over all simple paths can be computed numerically for functions on simple paths.

In this section, it is shown that path integrals over continuous paths can be approximated by path integration over simple paths. Hence, integration over all possible target movements can be approximated using the HMM quadrature. To do this, it is shown that every integrable function on continuous paths can be approximated by a simple function: a function defined on simple paths.

Lemma 1. For every continuous function $f : \mathcal{C}(\mathcal{A}) \to \mathbb{R}$, there exists a sequence of simple functions, $f_1, f_2, \ldots$, such that

$$\lim_{n\to\infty} f_n(\vec{\boldsymbol{x}}) = f(\vec{\boldsymbol{x}}).$$

.

Proof. Let $\Delta x = \Delta t = \frac{1}{n}$ and $\mathcal{G}(\Delta x)$ be a grid over $\mathcal{A}$ with grid points spaced Δx apart. Define the path-reduction function $r : \mathcal{C}(\mathcal{A}) \to \mathcal{S}(\Delta t)$ such that $r(\vec{\boldsymbol{x}})(s)$ is the grid point immediately below and to the left of $\vec{\boldsymbol{x}}(t)$ for all $s \in [t, t + \Delta t)$. Effectively, the x-y coordinates of $\vec{\boldsymbol{x}}(t)$ are reduced down to the nearest multiple of Δx and the entire path in the time-step $[t, t + \Delta t)$ reduced to the single location. The path $r(\vec{\boldsymbol{x}})$ is a simple path.

Define $f_n(\vec{\boldsymbol{x}}) = f(r(\vec{\boldsymbol{x}}))$. Thus, f_n is a simple function. As $n \to \infty$, $\Delta x, \Delta t \to 0$. Thus, $r(\vec{\boldsymbol{x}}) \to \vec{\boldsymbol{x}}$. Since f is continuous, this proves the result. □

The Lemma shows how to computationally approximate any continuous function of continuous paths using a sequence of simple functions. In practice, one such simple function f_N will be used for large N. The path integration computations discussed above can then be used to compute the integral of f_N. It remains to show that this integral is an approximation of the integral of f.

Theorem 2.10 (Convergence of Path Integrals)**.** For a function $f : \mathcal{C}(\mathcal{A}) \to \mathbb{R}$ where f is bounded above, if there exists a sequence of integrable functions $f_1, f_2, \ldots$ such that $f_n \to f$ pointwise, then $\int f_n(\vec{\boldsymbol{x}})\ \mathrm{d}\mathcal{W}(\vec{\boldsymbol{x}}) \to \int f(\vec{\boldsymbol{x}})\ \mathrm{d}\mathcal{W}(\vec{\boldsymbol{x}})$.

Proof. This result follows directly from Lebesgue's dominated convergence theorem (Rudin, 1987). □

This theorem shows that the approximate path integral will converge to the true value. Notice, this requires that f is bounded, that is, either by a constant or by an integrable function. For encounter modelling, this is equivalent to assuming the likelihood function is bounded.

Little is known about the rate of convergence of this approximation. Quantifying the error in the approximation is difficult due to the generality of the dominated convergence theorem. Avigad, Dean, and Rute (2011) have shown that in many cases the convergence has a meta-stability, that is, eventually the sequence of approximate integral values will change more and more slowly over time; however, this result does not quantify the absolute distance between the limit value and the current value of the sequence.

2.5.7 State-switching Path Integration

This section has focussed on Wiener path integrals. In Section 2.4.4, a state-switching path integral was introduced where target movement and detection can depend on an unobserved finite-state Markov process. A development similar to that given above can be derived for state-switching path integration. The difference is that now the unknown path $\vec{\boldsymbol{x}}$ is a three-dimensional process such that $\vec{\boldsymbol{x}}(t) = (\boldsymbol{x}, s)$ when the target is in location $\boldsymbol{x}$ and state s at time t. Let $\rho_{s,r}$ be the transition rate between states s and r.

The necessary adjustments to the theory above for state-switching are given here.

- Simple paths can be defined similarly to Definition 2.13 where targets change location and state only at each time-step and not between. The path integral over all simple paths is then a high-dimensional finite integral as in Theorem 2.4.

- As the state-switching process is Markov, the finite-dimensional integral of a Markov-separable function can be written as in Theorem 2.7 where integration is with respect to the state-switching Wiener measure $\mathcal{W}_s$.

- Thus, this finite-dimensional integral can be approximated by HMM quadrature. Again, this requires the transition rate matrix $\boldsymbol{R}_t$ for each time t to be derived by finite-difference. The analogue of Theorem 2.6 is Theorem 2.11.

  Theorem 2.11. Let $\phi_s : \mathcal{A} \times [0, T]$ be a function such that $\phi_s(\boldsymbol{x}, t)$ is the probability density of a target being in location $\boldsymbol{x}$ and state s at time t. The **Eulerian** equation for the state-switching Wiener process is given by Pedersen et al. (2011):

  $$\frac{\partial \phi_s}{\partial t} = -\nabla \cdot \boldsymbol{\mu_s}(\boldsymbol{x}, t)\phi_s - \frac{\sigma_s(\boldsymbol{x}, t)^2}{2}\nabla^2\phi_s + \sum_{r=1}^{S} \rho_{r,s}\phi_r(\boldsymbol{x}, t)$$

  where there are S total states, ∇ is the gradient operator on $\mathbb{R}^2$ and ∇^2 is the Laplacian on $\mathbb{R}^2$.

This partial differential equation can then be approximated by finite-difference methods as in Theorem 2.6. This leads to a transition rate matrix $\boldsymbol{R}_{s_1,s_2,t}$ at each time t for each pair of states s_1, s_2. The entries of this transition rate matrix describe the movement of a target across space when the target switches from state s_1 to state s_2. The full transition rate matrix is a block-matrix with $(i,j)^{\text{th}}$ block given by $\boldsymbol{R}_{i,j,t}$. The transition probability matrix is again derived as the matrix exponential of this full transition rate matrix.

- Given this, HMM quadrature can be used to approximate the state-switching path integral:

  Theorem 2.12. For a Markov-separable function $f : \mathcal{S}(\Delta t) \to \mathbb{R}$, a spatial grid $\mathcal{G}$ and a set of S states:

  $$\int_{\mathcal{S}(\Delta t)} f(\vec{\boldsymbol{x}}) \, \mathrm{d}\mathcal{W}_s(\vec{\boldsymbol{x}}) = \lim_{\Delta x \to 0} \boldsymbol{\gamma}_0 \boldsymbol{F}_0 \left(\prod_{\tilde{t}=1}^{\tilde{T}} \boldsymbol{\Gamma}_{\tilde{t}} \boldsymbol{F}_{\tilde{t}} \right) \boldsymbol{\epsilon}_1$$

  where

  1. $\boldsymbol{\gamma}_0$ is an initial distribution represented by a $1 \times S|\mathcal{G}|$ row vector with i^{th} entry given by $[\vec{\boldsymbol{x}}(0) = (\boldsymbol{g}_i, s_i)]$ where $\boldsymbol{g}_i$ is the grid point corresponding to entry i and s_i is the state corresponding to entry i;
  2. a $S|\mathcal{G}| \times S|\mathcal{G}|$ diagonal matrix $\boldsymbol{F}_{\tilde{t}}$ with i^{th} diagonal given by $f_{\tilde{t}}(\boldsymbol{g}_i, s_i)$;
  3. a **transition rate matrix** $\boldsymbol{R}_{\tilde{t}}$ for time-step $\tilde{t}$ of dimension $S|\mathcal{G}| \times S|\mathcal{G}|$;
  4. $\boldsymbol{\Gamma}_{\tilde{t}} = \exp(\boldsymbol{R}_{\tilde{t}} \Delta t)$ is the **transition probability matrix** over a time-step Δt;
  5. ϵ_1 is a $|\mathcal{G}|S \times 1$ vector of ones.

- With state-switching path integration, the Krylov approximation, matrix sparsity, and Fourier transform efficiencies can be made in a similar way to the above.

2.6 Parameter Estimation

There are two statistical aims: to identify a statistical model that describes the data-generating process well and to estimate the parameters of this model. This amounts to specifying a detection model (a hazard) and a target movement model (an advection-diffusion model). The selection of an appropriate model is discussed in the next section. Here, the procedure of estimating the parameters of a selected model is presented.

2.6.1 Data

Before fitting the model, one must ensure the problem is well-posed and the parameters are identifiable. For encounter surveys, it is important to consider what data are sufficient to estimate both the detection and movement processes. Intuitively, one would expect the parameters of these two processes to be correlated. If the detection range of each detector in a survey is increased, then it is possible to maintain a constant encounter rate by decreasing the rate of target movement. Targets cover less space and so are less likely to be seen which counterweights the increased coverage of the detectors. Due to this relationship, it is critical that data are collected that distinguish the detection and movement processes. Otherwise, the model will be non-identifiable: an infinite collection of movement and detection parameters satisfying this counterbalancing relationship will be equally likely with respect to the data.

In most cases, this can be avoided by collecting a sample of the movement process or auxiliary data on the detection range. Telemetry data on targets can provide samples of target movement paths independent of the detection process and so when incorporated with the encounter model, can isolate a unique set of most likely parameters. Similarly, if a field test of the detectors were conducted where movement paths were known or instead detection was tested at a variety of distances, then again this would isolate a unique parameter estimate.

2.6.2 Likelihood

Recall from Definition 2.2 that in an encounter survey a target population of size N is studied. A record is made of each encounter with a target. In particular, $\boldsymbol{r}_i$ is the set of records for encounters with the i^{th} target. Each record must contain the time of each encounter, but may also include the observed location of the target (or any other auxiliary information). Section 2.4 gives the PDF of these records given parameters for the encounter process $\boldsymbol{\theta}$. Targets are assumed to be independent and so the likelihood of these parameters for the full data is a simple product:

$$\mathcal{L}(\boldsymbol{\theta}) = \prod_{i=1}^{n} [\boldsymbol{r}_i \mid \boldsymbol{\theta}].$$

The maximum likelihood estimate (MLE) $\hat{\boldsymbol{\theta}}$ is found by maximising this function numerically. Notice, that $[\boldsymbol{r}_i]$ requires the evaluation of a path integral for each individual. Each

path integral is approximated by a sequence of integrals computed by HMM quadrature. Denote $\mathcal{L}_n$ to be the approximate likelihood computed for $\Delta x = \Delta t = n^{-1}$.

This formulation is equivalent to the misspecified HMM framework discussed by Mevel and Finesso (2004). These authors show that the limit of the maximum likelihood estimators $\hat{\theta}_n$ of the approximate likelihoods $\mathcal{L}_n$ is a consistent estimator of a parameter $\boldsymbol{\phi}$ where $\boldsymbol{\phi}$ minimises the relative entropy between the true underlying continuous process and the space of all HMM models. In this context, Theorem 2.10 shows that the path integrals, and so the HMM likelihood, converges to the true likelihood; thus, $\boldsymbol{\phi} = \hat{\boldsymbol{\theta}}$. Hence, the maximum likelihood estimator is consistent. Furthermore, Mevel and Finesso (2004) shows that this estimator is asymptotically normal for the Wiener process.

In short, for the methods considered here, standard maximum likelihood theory applies to the encounter model.

2.7 Model Selection and Fit

There are many different detection and movement models that can be used. Models can be selected based on Akaike Information Criterion (AIC) (Akaike, 1973). For movement models that are nested (for example comparing an advection-diffusion model with a diffusion model), the likelihood ratio statistic (Casella & Berger, 2001) could be used to compare models by the asymptotic chi-squared distribution.

Once a model is selected, the model's fit can be assessed. In the framework presented here, models are restricted to a Poisson detection process with an advection-diffusion target movement. In reality, detections may be driven by a process unamenable to adequate description by a Poisson process or target movement may have qualities that are overlooked by the selected advection-diffusion model. Hence, assessing the fit of the model can determine what confidence one ought to have in any inference derived from the model and where the gap between the model and reality lies. In this section, two goodness-of-fit (GoF) methods are discussed: residual analysis and simulation-based methods.

2.7.1 Residuals

Residuals can be computed using the probability integral transform. Consider the observed encounter times between target i and detector j, $t_{i,j,1}, \ldots, t_{i,j,n_{i,j}}$. Given the model, these times arise from a Poisson process and so their differences are independent random variables: the times between encounters are independent. Thus, by the probability integral transform,

the cumulative distribution function evaluated for each such encounter should result in a random sample from a uniform distribution.

As in Zucchini et al. (2016), a pseudo-residual can be defined.

Definition 2.18 (Encounter Residual)**.** For the e^{th} encounter between target i and detector j, the **encounter residual** is defined as

$$\epsilon_{i,j,e} = [t \leq t_{i,j,e} \mid \boldsymbol{t}_{i,j}]$$

where t is random variable equal to the time of the e^{th} encounter between target i and detector j.

If the location of an encounter is recorded, then the encounter residual can include this:

$$\epsilon_{i,j,e} = [t \leq t_{i,j,e}, \boldsymbol{x} \leq \boldsymbol{x}_i(t_{i,j,e}) \mid \boldsymbol{r}_{i,j}]$$

where $\boldsymbol{x}$ is a random variable equal to the target's position at time $t_{i,j,e}$.

Intuitively, the encounter residual is the probability a target would have evaded encounter for the duration that was observed for this model given all previous and future encounters with that target. Extreme residuals allow one to identify anomalous encounters. Furthermore, the residuals are expected to be independent uniform random variables when the model fits well (Zucchini et al., 2016). (These can be transformed into independent Gaussian random variables using the inverse probability integral transform, for no other reason than to aid human interpretation.) Thus, this distributional knowledge can be used to assess the model's fit. Quantile-quantile plots or formal statistical tests (e.g. Shapiro and Wilk (1965)) can be used to detect patterns or deviations from the model. Similarly, a Kolmogorov-Smirnov test can be used to compare the empirical distribution of the residuals to the theoretical expected distribution. Residuals in the lower tail would imply individuals are encountered too quickly and so perhaps are attracted to the detectors; conversely, residuals in the upper tail suggests that individuals are repelled systematically from detectors. Other patterns in the residuals, for example a temporal correlation, could indicate that encounters are clustered not only due to the relative locations of detectors and individuals but because of another unobserved process.

2.7.2 Simulation-Based

An alternative way to measure a model's fit is to simulate new data assuming the model to be an accurate description of the true data-generating process. This has become a common way to study the fitness of a model (Gelman et al., 2014). From this, the distribution of one or more summary statistics conditional on the model being a true depiction of reality can be approximated. Comparing the observed summary statistic to this empirical distribution provides a formal test as to whether the observed data have any characteristics considered anomalous when compared to the model.

Let $\boldsymbol{r}^{(1)}, \ldots, \boldsymbol{r}^{(D)}$ be $D \in \mathbb{N}$ sets of encounter records from D simulated encounter surveys with parameters equal to the estimated parameters of the fitted model. Let $S(\boldsymbol{r}) \in \mathbb{R}$ be a summary statistic for records $\boldsymbol{r}$. Common choices of S may be:

- Mean encounter rate of a target with all detectors
- Mean encounter rate of a detector with all targets
- Average time between re-encounters with a target
- Total number of unique targets that were encountered
- (If locations were observed) average number of encounters at a given radial distance from a detector

Let the observed summary statistic be $S^* = S(\boldsymbol{r})$. An approximate p-value p^* can be used to quantify the strength of evidence against the hypothesis that the fitted model is a good description of the true data-generating process:

$$p^* = \frac{1}{D} \sum_{d=1}^{D} \mathbb{I}(|S(\boldsymbol{r}^{(d)})| > |S^*|)$$

where $\mathbb{I}(A) = 1$ when A is a true statement and 0 when false.

Comparing the model to the observed data using several summary statistics can provide insight into what characteristics of the data-generating process the model fails to capture.

2.8 Conclusion

Encounters between scientific detectors and members of a target population arise from their relative movement and their mutual detectability. This chapter presents a mathematical

model for these encounters. This theoretical framework is not complete but provides the necessary essentials to begin developing more complex models.

The computational algorithms described for path integration, in particular, introduce a novel way to compute an analytically intractable likelihood. HMM quadrature and the adjusted forward algorithm are the first steps toward an efficient alternative to Monte Carlo methods when integrating over correlated paths in high-dimensional spaces.

In Chapter 6, future research directions are outlined for both the theoretical and computational development of these models. Before this, in Chapters 3 – 5, distance sampling and spatial capture-recapture are shown to be special cases of the general encounter model presented in this chapter. Recognising this is shown to improve upon the inference obtained from these methods and to widen their application.

Chapter 3

Distance Sampling

Distance sampling (DS) is one of the most popular methods to estimate abundance of a wild animal population (Buckland et al., 2001, 2015). In this chapter, the statistical theory of distance sampling is presented, tracing its origins from the method of plot sampling. Doing so reveals distance sampling to depend on the fundamental assumption that animals do not move whilst they are surveyed. This assumption is removed by the novel approach discussed in this chapter. Distance sampling is shown to be a special case of the encounter model as developed in Chapter 2. Thus, the statistical theory of encounter models can be applied to achieve three goals:

1. incorporate animal movement into distance sampling theory and investigate the effect movement has on the estimation of animal population density;
2. show that incorporating movement generalises distance sampling and does not exclude any of the most important extensions of distance sampling, for example, multiple-covariate distance sampling or mark-recapture distance sampling;
3. extend the use of distance sampling methods to species and technology that currently cannot be reconciled with the assumption that animals do not move.

3.1 Conventional Distance Sampling

The objective of distance sampling is to estimate the total number of individuals in a population. It is a generalisation of the simple method of plot sampling (Yapp, 1956).

3.1.1 Plot Sampling

Suppose a study region of area A contains N animals. A plot sampling survey consists of placing a series of S circular or rectangular plots across the study region, each plot i covering a known area a_i. Observers then go to each plot i and record the number of individuals seen n_i.

There are two key assumptions made when plot sampling (Seber, 1982) is used:

- no individuals within the plot are missed by the observers;
- animals do not leave or enter a plot whilst it is being surveyed.

Plot sampling was originally intended for populations of static objects, such as plants (Gleason, 1920). Hence, the above assumptions were reasonable in many applications. For animal populations, these assumptions are likely to be violated especially when plots are large or animals difficult to detect Buckland et al. (2001). Distance sampling removes the first assumption; the methods presented in this chapter remove both assumptions.

Given the data, $n_1, \ldots, n_S$, the density of the population can be estimated. There are two approaches to density estimation: design-based and model-based (Buckland et al., 2015). The former approach is termed design-based because it relies on a further assumption about the design of the plot sampling survey: it assumes that plot locations are, on average, uniformly distributed across the study area: plot locations are representative of the study region. This can be achieved either by randomly locating each plot or placing plots systematically but with a random starting position. With this design assumption, the average number seen per unit area searched is an estimate of density:

$$\hat{D} = \frac{\sum_{i=1}^{S} n_i}{\sum_{i=1}^{S} a_i}.$$

Absolute abundance can then be estimated as $\hat{N} = \hat{D}A$. Variance of the estimator can be estimated by the sample variance of the plot densities or by bootstrap (Fewster et al., 2009).

The alternative approach is termed model-based Buckland, Oedekoven, and Borchers (2016). An explicit model is specified for the counts observed in each plot. The most common choices are the Poisson or negative binomial distributions. For the Poisson, $n_i \sim \text{Po}(Da_i)$, the maximum likelihood estimator of D is equivalent to the design-based approach. The model-based approach, however, can be extended to account for environmental covariates

using the standard generalised linear model framework. This can be used to account for sampling that is not randomised if the relevant covariates are recorded.

3.1.2 Distances

When observers visit plots they may miss individuals and so population density is underestimated. Conventional distance sampling (CDS) is a generalisation of plot sampling that estimates the proportion of individuals missed per plot. In a distance sampling survey, observers record a distance for each individual that is detected. In circular plots, this distance is the radius from the central point of the plot to the individual's location; for rectangular plots, it is the perpendicular distance of the individual to the central line of the plot. For this reason, circular plots are termed point transects and rectangular plots are named line transects.

The idea is that individuals are less likely to be detected at greater distances. The recorded distances to detected individuals can be used to estimate the number of individuals missed at each distance. Let y_i be the distance of the i^{th} individual from the transect and let $d_i = 1$ if the individual was detected and zero otherwise. Observers visit each transect and only record the distances of those animals that were seen, i.e, where $d_i = 1$. Thus, by Bayes theorem, the probability density of the recorded distances has the form:

$$[y_i \mid d_i = 1] = \frac{[d_i = 1 \mid y_i][y_i]}{\int [d_i = 1 \mid y][y] \ \mathrm{d}y}.$$

Hence, the distribution of the observed distances is specified when the components $[y_i]$ and $[d_i = 1 \mid y_i]$ are specified. The first component, $[y_i]$, is the distribution of individuals within each transect. The design-based assumption used in plot sampling is also assumed in distance sampling, thus, individuals are distributed, on average, uniformly within the transect. For a line transect of width w, this implies the perpendicular distance is uniformly distributed also; for point transects, as the area covered increases with radial distance, the distribution is triangular.

The second component, $[d_i \mid y_i]$, is termed the detection function and often denoted $g(y_i)$. It is the probability of an individual being detected given it is at a certain distance. Two assumptions are made about the detection function:

- individuals at distance zero are detected with certainty: $g(0) = 1$;
- the detection function is monotonically decreasing and has a shoulder around zero, that is, $g'(0) \approx 0$.

The first assumption is crucial to distance sampling: if one assumes all individuals were seen at distance zero, then the relative proportion seen at further distances, since $[y_i]$ is known, can be used to estimate the relative change in the detection function. The second assumption is not necessary but is common sense and is used to determine what functions could be used to represent the detection function.

In CDS, a functional form is chosen for g that satisfies both assumptions. The most common choice is the half-normal function with scale parameter σ:

$$g(y) = \exp\left(-\frac{y^2}{2\sigma^2}\right).$$

This choice is not theoretically justified, though often it fits well to real data. After a functional form for the detection function is chosen, it can be adjusted by adding a series of trigonometric or Hermite polynomials. These adjustments improve the fit to the real data.

An alternative approach is to explicitly model the search process that occurs in each transect (Buckland et al., 2004). When a transect is surveyed, some individuals are detected and some can be said to survive without being detected. Survival analysis (Miller Jr, 2011) is applied to such situations where one or more individuals are at risk of death and the aim is to estimate the hazard of each individual the instantaneous risk of failing to survive. In distance sampling, this risk will depend on the location of the individual within the transect. Let $\lambda(\boldsymbol{x}_i, t)$ be the hazard of individual i at two-dimensional location $\boldsymbol{x}_i$ within the transect being detected at time t. A common form for this hazard is $\alpha r(\boldsymbol{x}_i)^{-\beta}$ where $r(\boldsymbol{x})$ is the distance between location $\boldsymbol{x}$ and the transect and α, β are parameters. The probability of an individual surviving, that is, not being detected, for time t is termed the survival function, $S(\boldsymbol{x}_i, t)$. It is easily shown (Buckland et al., 2004) that

$$S(\boldsymbol{x}_i, t) = \exp\left(-\int_0^t \lambda(\boldsymbol{x}_i, s)\mathrm{d}s\right).$$

Consider a transect that is surveyed for a total time T. It is possible to define a two-dimensional detection function as $g(\boldsymbol{x}) = 1 - S(\boldsymbol{x}, T)$. The one dimensional detection function can be obtained by integrating over the second spatial domain. Conventionally, one dimensional detection functions have been used in distance sampling, but there are advantages to the use of a two-dimensional detection process; Borchers and Cox (2017) have shown that a two-dimensional process can reduce bias in some cases. In this thesis, as in Chapter 2, the focus will be on two-dimensional detection processes.

Ultimately, the detection function can be derived from a chosen hazard. If the hazard is

infinite at zero radius and monotonically decreasing, then the resultant detection function satisfies the required assumptions. The detection function derived in this way is termed a hazard-rate detection function.

Given both components, $[y_i]$ and $g(y_i)$, the density of the observed distance, denoted $f(y_i)$, is given:

$$f(y_i) = \frac{g(y_i)[y_i]}{p_i}$$

where $p_i = \int g(y)[y]\,\mathrm{d}y$ is the detection probability within the transect that individual i was detected on. Recall, division by p_i is necessary as only the distances of detected individuals are recorded.

The joint density of the observed distances for all n detected individuals is obtained by assuming that detections are independent:

$$\mathcal{L}_{y\,|\,n}(\boldsymbol{\theta}_d) = \prod_{i=1}^{n} f(y_i).$$

This joint density is conditional on the value of n and so is termed the conditional likelihood for the detection parameters $\boldsymbol{\theta}_d$. The conditional likelihood $\mathcal{L}_{y\,|\,n}$ can be used to estimate the detection function from the observed distances. This can be achieved by maximum likelihood and the form of the detection function selected using AIC. Model fit can be assessed by the Kolmogorov-Smirnov, Cramer-Von-Mises, or chi-squared tests.

3.1.3 Density

As with plot sampling, density can be estimated in two ways: design-based and model-based. Recall, S transects (plots) are surveyed and a count of the number of individuals in each plot recorded, $n_1, \ldots, n_S$. By design, it is assumed that transects are placed, on average, uniformly across the study area. The design-based approach makes use of this assumption. It is also called a two-stage approach because first detection parameters are estimated from the conditional likelihood. This provides an estimate of the proportion of individuals seen in transect s, $\hat{p}_s$. Hence, the estimated density within transect s is $\dfrac{n_s}{a_s\hat{p}_s}$. As with plot sampling, the population density is estimated as

$$\hat{D} = \frac{\sum_{s=1}^{S} n_s}{\sum_{i=s}^{S} a_s\hat{p}_s}.$$

The variance of this estimator is computed in two stages: the delta method (Ver Hoef, 2012) is used to estimate the uncertainty in $\hat{p}_s$ and observed plot counts are used to estimate the uncertainty in n_s (Fewster et al., 2009).

An alternative approach is model-based (Buckland et al., 2016; Hedley & Buckland, 2004) where the transect counts, $n_1, \ldots, n_S$, are given a parametric distribution, for example, a Poisson where

$$n_s \sim \text{Po}(Da_s p_s)$$

for transect s. Estimation of detection and density parameters $\boldsymbol{\theta}$ can then be achieved in a single step with the full likelihood:

$$\mathcal{L}_{y,n}(\boldsymbol{\theta}) = [n] \prod_{i=1}^{n} f(y_i)$$

where $[n]$ is the PDF for the selected density model. With this method, variance estimates can be computed using the estimated Hessian (Buckland et al., 2001) without need of the delta method or bootstrapping. In practice, the model-based approach is less robust to violations of any further assumptions made about the model for density. The design-based is more robust but is less flexible when extensions to distance sampling are made.

3.1.4 Assumptions

In summary, distance sampling relies upon four key assumptions.

- **Transects are placed at random or systematically at random** This assumption can be satisfied in many applications unless logistical or geographical constraints make visiting certain places infeasible.
- **Detection of individuals at zero distance is certain,** $g(0) = 1$ This may not be satisfied in some surveys, for example, in shipboard cetacean surveys where cetaceans that dive directly under the ship, at zero distance, are missed. This can lead to severe underestimation of density. Furthermore, the collected distances do not provide any evidence of when this underestimation may have occurred. When $g(0)$ is known to likely be less than 1, a mark-recapture distance sampling (MRDS) (Borchers et al., 1998) survey can be conducted, requiring at least two observers. The additional information recorded during a MRDS survey allows for $g(0)$ to be estimated rather than its value be assumed. MRDS models are discussed in Section 3.9. An alternative solution is to directly model the process that determines when animals are available

or unavailable for detectionat zero distance, for example, when cetaceans dive or are near the surface. Models for availability (Borchers et al., 2013) can then account for $g(0) < 1$. These models are discussed in Section 3.6.1.

- **Measurement error is zero** The detection function is estimated from the distances. Any systematic or random error in these measurements can bias the detection function estimator. Extensions of CDS that account for measurement error exist and can be used when necessary (Marques, 2004). This is discussed in Section 3.4.4.

- **Individuals do not move** This is the fundamental assumption inherited from plot sampling. This is termed the snapshot assumption as it envisages that observers survey a transect in a snapshot of time, thus individuals cannot move. For many CDS surveys, this assumption is violated. Movement of individuals in response to the observer, avoidance or attraction, is well-known to cause bias in density estimation (as discussed in (Au & Perryman, 1982; Hammond et al., 2017; Palka & Hammond, 2001; Turnock & Quinn, 1991)). Survey methods and MRDS models can be used to mitigate this bias (Section 3.10) and there is widespread awareness of the impact responsive movement can have. Conversely, the effect of non-responsive movement of individuals is not routinely investigated despite the possibility it can have a substantial effect on density estimation (Glennie et al., 2015). Furthermore, there is no general method available to account for this bias. Incorporating non-responsive animal movement into distance sampling is the focus of this chapter.

3.2 DS with Movement

The process that generates distance sampling data does not match with the models used to analyse these data. The snapshot assumption, inherited from plot sampling, treats the individuals being surveyed as static immobile entities. Most distance sampling research is based upon this paradigm, in particular, density surface modelling (Miller et al., 2013) and point process formulations of distance sampling (Yuan et al., 2016) are strongly dependent on the snapshot assumption. There is no distance sampling method that accounts for the movement of individuals.

Distance sampling is an example of an encounter survey, as described in Chapter 2. The data arise due to encounters between the observer and individuals. These encounters depend on the detectability and movement of each individual. In particular, they depend on the entire unobserved path an individual has travelled whilst the transect is surveyed. Thus, the

statistical and computational theory presented in Chapter 2 generalises distance sampling, allowing for individuals to move over time. This new class of models is termed MDS models: distance sampling with animal movement.

3.2.1 Encounters

In particular, CDS is an encounter survey with a single detector, the observer, that records the first encounter with each individual seen. Only the first encounter is recorded because often once an individual has been seen, re-encounters arise from a different encounter process. For example, in many cases, the observer is able to track the individual after it has been sighted or may approach the individual so as to investigate it further.

In Chapter 2, an encounter model was constructed from two components: a detection process and a movement process. The detection process is similar to CDS where individuals further from the observer are less likely to be detected. For an encounter model, detection is described by the encounter intensity function (Definition 2.3). This function is synonymous with the hazard function introduced in Section 3.1.2. Unlike CDS, the encounter model constructed requires the time of each detection to be recorded. This information is necessary if movement is to be incorporated as one must know how long each individual survived detection. The location and time of each detection is commonly recorded in distance sampling surveys. For surveys where exact location or exact times have not been recorded, there are alternative approaches that can be used, see Section 3.4.1.

The detection model in CDS is constructed in two parts: a probability for detection d_i given the distance y_i an individual resides at, $[d_i = 1 \mid y_i]$ and a distribution for the distance $[y_i]$. The MDS model has a similar structure: the PDF of the observed first encounter with individual i at time t_i in location $\boldsymbol{y}_i$ given the path the individual travelled $\vec{\boldsymbol{x}}_i$, $[t_i \mid \vec{\boldsymbol{x}}_i]$, and the distribution of possible paths $[\vec{\boldsymbol{x}}_i]$. Theorem 2.1 gives the form of the first component:

$$[\boldsymbol{y}_i, t_i \mid \vec{\boldsymbol{x}}_i] = \lambda(\boldsymbol{y}_i, t_i) S(\vec{\boldsymbol{x}}_i, t_i)$$

where λ is the encounter intensity function and S is the survival function:

$$S(\vec{\boldsymbol{x}}, t) = \exp\left(- \int_0^t \lambda(\vec{\boldsymbol{x}}(s), s) \, \mathrm{d}s \right).$$

This is a generalisation of CDS. If individuals did not move, then $\vec{\boldsymbol{x}}(t) = \boldsymbol{y}$ for all t, and this detection function reduces to a hazard-rate detection function as in CDS with hazard equal to the encounter intensity function λ.

The second component is a model for the movement paths $[\vec{\boldsymbol{x}}_i]$. In CDS, the distribution of distances was known *a priori* to be uniform, for line transects, or triangular, for point transects. Furthermore, the distance of each detected individual was observed. Distance sampling data, records of first encounters, provide no information about how the individuals move. Thus, the parameters of any movement model cannot be estimated from the survey data; they must be estimated independently. Telemetry (e.g. GPS tagging) is a popular tool for tracking the movements of individuals in a population. The observed tracks of these individuals can be used to estimate the parameters of the movement model. Theoretically, these individuals need not be members of the surveyed population nor their movements be tracked at the same time as the distance sampling survey. The sole assumption is that the movements of these tracked individuals must be representative of the movements of those individuals within the surveyed population. The best practice to ensure this is to track a large number of individuals within the survey population at the same time as the distance sampling survey. This could be achieved by tracking detected individuals after their first detection and recording their movements. When individual telemetry data are not available movement parameters could be estimated by expert elicitation where point estimates or distributions for the movement parameters are specified and then used as known constants, often termed "plug-in" estimation. This is not an ideal approach, but will allow for individual movement to be accounted for rather than ignored. In Section 2.3, movement models and the methods to fit such models are discussed. Let $\mathcal{L}_m$ be the likelihood of the movement parameters $\boldsymbol{\theta}_m$ for a given movement model and observed telemetry data. Once the movement parameters are estimated, the PDF of any movement path can be computed, $[\vec{\boldsymbol{x}}_i]$.

Both components, $[\boldsymbol{y}_i, t_i \mid \vec{\boldsymbol{x}}_i]$ and $[\vec{\boldsymbol{x}}_i]$, give the joint PDF of $\boldsymbol{y}_i$ and t_i:

$$f(\boldsymbol{y}_i, t_i) = \frac{\int [\boldsymbol{y}_i, t_i \mid \vec{\boldsymbol{x}}_i][\vec{\boldsymbol{x}}_i] \, \mathrm{d}\vec{\boldsymbol{x}}_i}{p_i}$$

where $p_i = 1 - \int S(\vec{\boldsymbol{x}}, T_i)[\vec{\boldsymbol{x}}] \, \mathrm{d}\vec{\boldsymbol{x}}$ is the probability of being detected whilst the observer surveys the transect that individual i was detected within for total time T_i. Notice, the entire path the individual travels is unknown and so one must average over all possible paths the individual could have travelled whilst a plot was surveyed. The movement model is used to quantify the probability of each possible path occurring. This integration over all possible paths is a path integral. Section 2.5 covers the statistical theory behind these integrals and the computational algorithms required to compute them.

As in CDS, individuals are assumed to be independent. Thus, the joint density of all

observed encounters is

$$\mathcal{L}_{e \mid n}(\boldsymbol{\theta}_d \mid \boldsymbol{\theta}_m) = \prod_{i=1}^{n} f(\boldsymbol{y}_i, t_i)$$

Similar to CDS, this joint density is conditional on the value of n and so termed the conditional likelihood for the detection parameters $\boldsymbol{\theta}_d$. The conditional likelihood depends on the movement parameters $\boldsymbol{\theta}_m$. In a two-stage approach, these parameters can be estimated first and then, conditional on these estimates, detection parameters can be estimated from the conditional likelihood given the estimated movement parameters. With this approach, the uncertainty in the movement parameters will not be included in the uncertainty of the detection parameters. Uncertainty can be propagated by a parametric bootstrap, simulating from the fitted movement model, however, this is likely to be computationally intensive as it requires the MDS model to be fit a large number of times. The alternative approach is to estimate the detection and movement parameters jointly by maximising the joint likelihood: $\mathcal{L}_{m,e \mid n} = \mathcal{L}_m \mathcal{L}_{e \mid n}$. This joint likelihood is constructed by assuming that the telemetry and distance sampling observations are independent. The advantage of this approach is that uncertainty in the movement parameters is reflected in the uncertainty in the detection parameters.

Finally, the goodness-of-fit of MDS models can be assessed using the methods described in Section 2.7. The PDF of the observed distances can be computed from $f(\boldsymbol{y}, t)$ by integrating over time. As in CDS, this can be compared to the empirical distribution of the observed distances using a Kolmogorov-Smirnov, Cramer-Von-Mises, or Chi-squared test.

3.2.2 Density

Again, there are two approaches to estimating density: model-based and design-based. For the design-based approach, the conditional likelihood is maximised to estimate the detection parameters and movement parameters. The estimated density is then given by

$$\hat{D} = \frac{\sum_{s=1}^{S} n_s}{\sum_{s=1}^{S} A\hat{p}_s}$$

where n_s is the number of animals seen in transect s and $\hat{p}_s$ is the estimated probability of an animal being seen on transect s. Notice, this is the same form as the estimator used for CDS, the difference is that the estimated probability of detection $\hat{p}_s$ now accounts for movement. Unlike in CDS, the area of the transect a_s is not included in the denominator. Individuals can move into the transect whilst it is surveyed, thus the density of individuals

at risk of detection is not the density of individuals in the transect; individuals initially outside the transect can move in and be detected. Hence, $\hat{p}_s$ is the estimated probability of an individual anywhere in the survey region being detected. If individuals did not move and detection within the transect was a certainty, then $\hat{p}_s = \frac{a_s}{A}$ as in plot sampling.

The variance of this density estimator can be approximated by bootstrapping or by the delta method. The squared coefficient of variation for the estimated density from transect s is:

$$\frac{\widehat{\text{Var}}(n)}{n^2} + \frac{\widehat{\hat{p}_s}}{\hat{p}_s^2}.$$

The estimation of $\widehat{\text{Var}}(n)$ is discussed in detail by Fewster et al. (2009). The variance of the detection probability $\hat{p}_s$ is

$$\widehat{\text{Var}(\hat{p}_s)} = \boldsymbol{J}^{-1}\hat{\boldsymbol{V}}\boldsymbol{J}.$$

where $\hat{\boldsymbol{V}}$ is the covariance matrix of the estimated detection and movement parameters, estimated as the inverse of the negative Hessian of the conditional likelihood, and $\boldsymbol{J}$ is the Jacobian of the function that transforms the parameters to the estimate $\hat{p}_s$. The Jacobian is routinely computed using numerical differentiation.

Model-based density estimation is also similar to the CDS case. For example, a Poisson density model would have the form $n_s \sim \text{Po}(DAp_s)$. Again, given an explicit model for density, the full likelihood can be used:

$$\mathcal{L}_{e,m,n}(\boldsymbol{\theta}) = [n]\mathcal{L}_m(\boldsymbol{\theta}_m)\prod_{i=1}^{n} f(\boldsymbol{y}_i, t_i)$$

for all parameters $\boldsymbol{\theta}$.

Maximising the full likelihood would lead to joint estimation of the detection and density parameters. Furthermore, with animal telemetry data, the likelihood for the movement model can be integrated also, meaning all parameters would be estimated jointly and uncertainty propagated appropriately to the final density estimator.

3.2.3 Assumptions

The assumptions made in the theory presented are synonymous with those made in CDS. Violations of these assumptions will cause bias in the inference obtained.

1. **Sampling is representative and independent**. For the distance sampling survey, this assumption requires transects be placed according to a randomised design, that

transects be independent, and that individuals be independent. Assuming transects are independent, makes the assumption that if an individual were seen in two or more transects, these events would be considered independent. In reality, it is not possible to know whether an individual was seen on multiple transects, thus transects are considered separately.

For the movement model, it is assumed that tagged animals move independently and that the sample of tagged animals be representative of the surveyed population. One can use telemetry from tagged animals who are not members of the surveyed population, but only with the assumption that these animals exhibit movement patterns similar to those animals surveyed by distance sampling. For animals that travel in groups, treating groups as the independent unit to be sampled may be a better choice.

2. **Animals at zero radius are detected**. This assumption is equivalent to the canonical assumption in CDS that $g(0) = 1$. It is possible to use hazard functions that allow for $g(0) < 1$ (Borchers & Cox, 2017) but this approach has been found to not work well in practice. Alternatively MRDS or availability models can be used with MDS, discussed in Sections 3.9 and 3.6.1.

3. **Location measurements are exact**. This assumption applies to observed locations recorded on the distance sampling survey and the recorded telemetry. Observation error in telemetry data is common and can be accounted for (e.g. Johnson et al. (2008); Pedersen et al. (2011)). Incorporating measurement error into MDS is discussed in Section 3.4.4.

4. **The path an individual travels is independent of the observer**. Individuals do not respond to the observer and their movement is independent of the transect placement, that is, surveying does not preferentially take place in areas individuals would avoid or be attracted to. The former is mitigated by use of a two-dimensional detection process (Borchers & Cox, 2017) and can be accounted for explicitly as in Section 3.10. The latter is avoided by random placement of the transects.

5. **Individuals move according to the specified movement model**. The simple model that individual movement is a spatially-invariant, isotropic diffusion process is violated by many populations; more realistic movement models can be considered at the cost of greater computational burden. No matter what movement model is incorporated, one assumes that all individuals in the survey move according to the

specified model. Departures from the movement model could cause estimated detection probability to be biased.

3.2.4 Plot Sampling

CDS is shown to be a generalisation of plot sampling. MDS is a generalisation of CDS that allows for animal movement. Thus, MDS can be used to generalise plot sampling to allow for movement. In plot surveys, detection within a plot is certain and is impossible outside the plot. This could be described by a hard-core encounter intensity where $\lambda(\boldsymbol{x}, t) = \infty$ for all times t and locations $\boldsymbol{x}$ within the plot and is zero otherwise. The probability of an individual being included in the count of plot s is then simply the probability the individual's path enters the plot at some time:

$$\int_{\chi_s} [\vec{\boldsymbol{x}}_s] \, \mathrm{d}\vec{\boldsymbol{x}}_s$$

where χ_s is the set of all paths that enter the plot s at some time during the survey.

Once a movement model has been specified and its parameters estimated either by expert elicitation or from telemetry data, the value of p_s can be computed. For ideal free gas movement, p_s can be computed analytically for circular and strip plots, as shown in Glennie et al. (2015) and Gonzalez (2018). Density can then be estimated from the plot counts as before using the corrected detection probability p_s.

3.3 Simulation Study

Glennie et al. (2015) showed that density estimation from distance sampling can be substantially biased when individuals move and that this bias increases with movement speed. Here, a simulation study is conducted to compare density estimation using CDS and MDS models. A study population of 100 animals in 100 square kilometres is simulated. Throughout the survey, animals move according to a diffusion process with mean speed ν.

Two distance sampling surveys are considered: a line transect survey and a point transect survey. The line transect survey consisted of 50 transects placed at random across the survey area with fixed length 1km and width 60 metres. The observer traversed each line with speed 1 metre per second. For the point transect survey, a series of 100 point transects of radius 100 metres are surveyed for 5 minutes each.

In both surveys, the encounter intensity was of the form $\alpha r^{-\beta}$ where r was the distance between the individual and the observer. The detection parameters α, β were chosen such

that the effective search area of each transect was 0.015 square kilometres.

Independent animal telemetry data were simulated for ten tracked individuals where their location was recorded every minute for one hour. Individuals move according to a diffusion process and their mean speed ν is estimated using maximum likelihood. During the survey, animals also moved by a diffusion process. A wrap-around method was used when animals came into contact with the boundary of the simulated study area: any animal that crossed the boundary was replaced by a new animal that entered on the opposing boundary. This ensures the density in the simulated area is constant.

For each survey, one hundred simulations were performed. For each simulated data set, a two-dimensional CDS model and a MDS model were fit by maximum likelihood. For the MDS model, the detection and movement parameters were estimated jointly. Density is estimated for both CDS and MDS using the design-based approach.

The relative bias and mean square error of the density estimator is estimated from the simulations.

3.3.1 Line Transects

CDS estimators of detection probability and abundance are biased when individuals move at large speeds compared to the observer. For this scenario, bias exceeds 10% when individuals move faster than 150% observer speed; bias exceeds 100% for speeds exceeding 300% observer speed. This bias is caused by two effects. First, individuals originally located outside the transect move into the transect and are detected; thus, the total number detected is larger than it would have been for an immobile population. Second, individuals are most likely to be seen when they are closest to the observer, thus the distance recorded for an individual is, on average, smaller than what would be recorded if the individual were immobile during the survey. The first effect leads to a count n that overestimates the number seen that originally resided in the transect; the second effect leads to an estimated detection function that is more steep than the true function, resulting in an underestimated detection probability, $\hat{p}$. Both effects cause positive bias in the estimated abundance $\frac{n}{\hat{p}}$.

In contrast, MDS has $< 5\%$ bias for all scenarios (Figure 3.1). Mean square error showed a similar comparison. Confidence interval coverage across all parameters for CDS was less than 40% for speeds over 100% observer speed and fell to 0% for speeds over 200% observer speed. MDS coverage was nominal within 2% for all parameters and across all simulation scenarios.

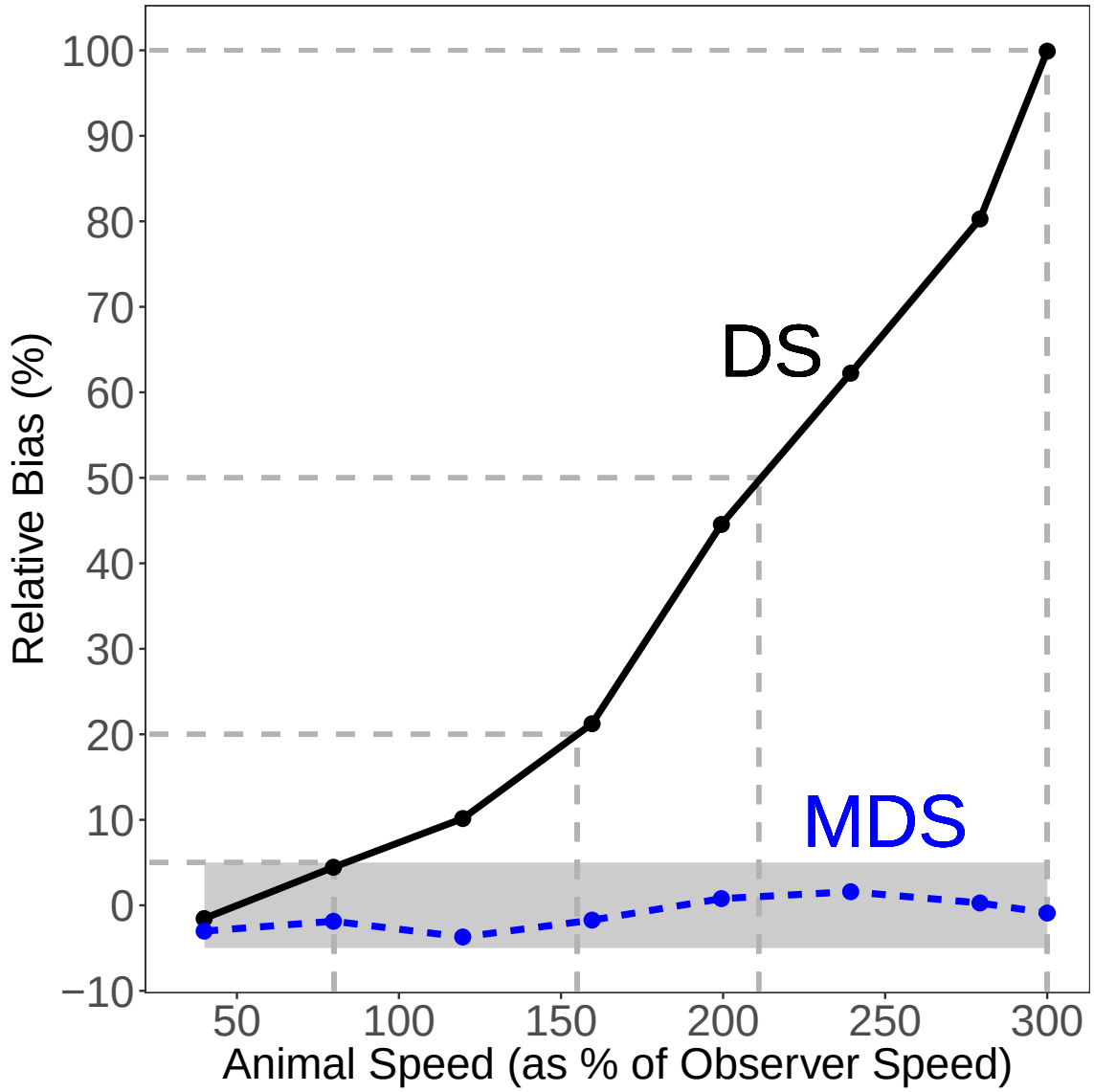


Figure 3.1 *Percentage relative bias in estimated density for conventional distance sampling (solid black line) and distance sampling with movement incorporated (dotted blue line) against animal speed (as % of observer speed) estimated from* 100 *simulations of a line transect survey of* 50 *transects with truncation width* 30 *metres and observer speed* 1 *metre per second. Animals move according to diffusion. Shaded region marks* ±5% *relative bias.*

3.3.2 Point Transects

In point transect surveys, the observer is stationary and, thus, any individual movement has the potential to cause substantial bias. For this scenario, CDS produced density estimates with bias $> 10\%$ for speeds > 2 metres per second and bias reaching 90% for speeds around 4 metres per second (Figure 3.2).

Incorporating movement reduced bias to $< 5\%$ across all scenarios and MSE varied negligibly. CDS 95% confidence interval coverage was poor ($< 45\%$) for all parameters when speed exceeded 2 metres per second, while coverage was nominal for all parameters when movement was incorporated.

3.3.3 Robustness

In the above simulation study, individuals moved according to a diffusion process and the MDS models assumed this. In reality, individuals move with correlated velocity; however, fitting such models with MDS is currently computationally prohibitive. The higher the correlation in animal movement, the more biased density estimation from CDS becomes (Glennie et al., 2015).

A simulation study was conducted to assess how MDS performed when the assumption of diffusive animal movement was extremely violated. The line transect survey from the previous simulation study was repeated with the same parameters, except that animals moved according to an ideal free gas model, in randomly-oriented straight lines, with a constant speed ν. This model describes animal movement with the highest correlation possible (perfect correlation in velocity).

Both MDS and CDS overestimated abundance when animal speed exceeded observer speed (Figure 3.3). Nevertheless, the bias in abundance estimation from MDS was half that from CDS. Both methods produced poor nominal coverage of confidence intervals for higher animal speeds. This extreme example shows that current MDS methods applied to highly transitory or migratory species can be biased when diffusive animal movement is assumed. Currently, however, MDS provides a better alternative to CDS and the potential for further development to incorporate movement models that can account for this animal behaviour.

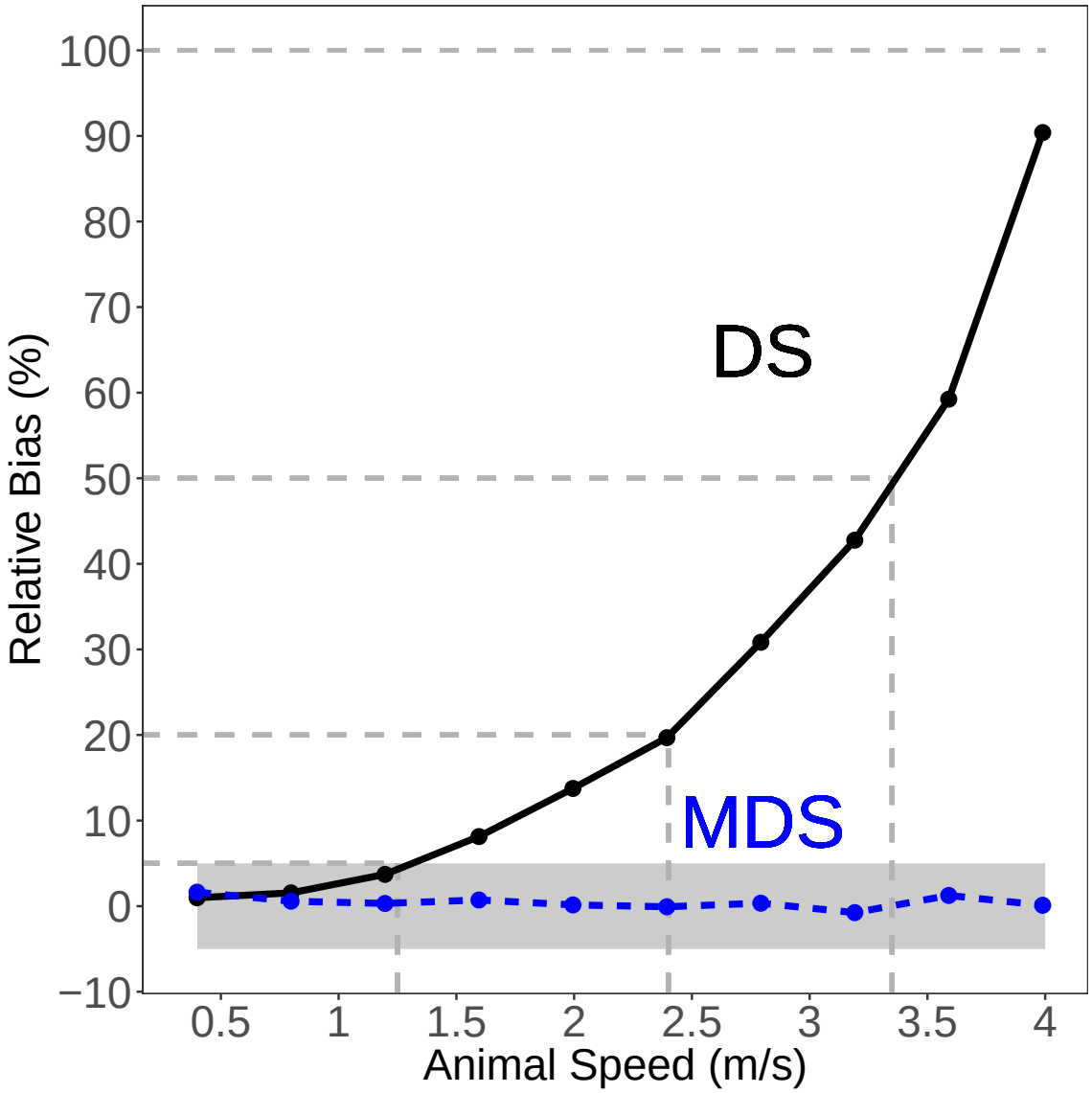


Figure 3.2 *Percentage relative bias in estimated density for conventional distance sampling (solid black line) and distance sampling with movement incorporated (dotted blue line) against animal speed (metres per second) estimated from* 100 *simulations of a point transect survey with* 100 *transects of radius* 100 *metres, surveyed each for* 5 *minutes. Animals move according to diffusion. Shaded region marks* ±5% *relative bias.*

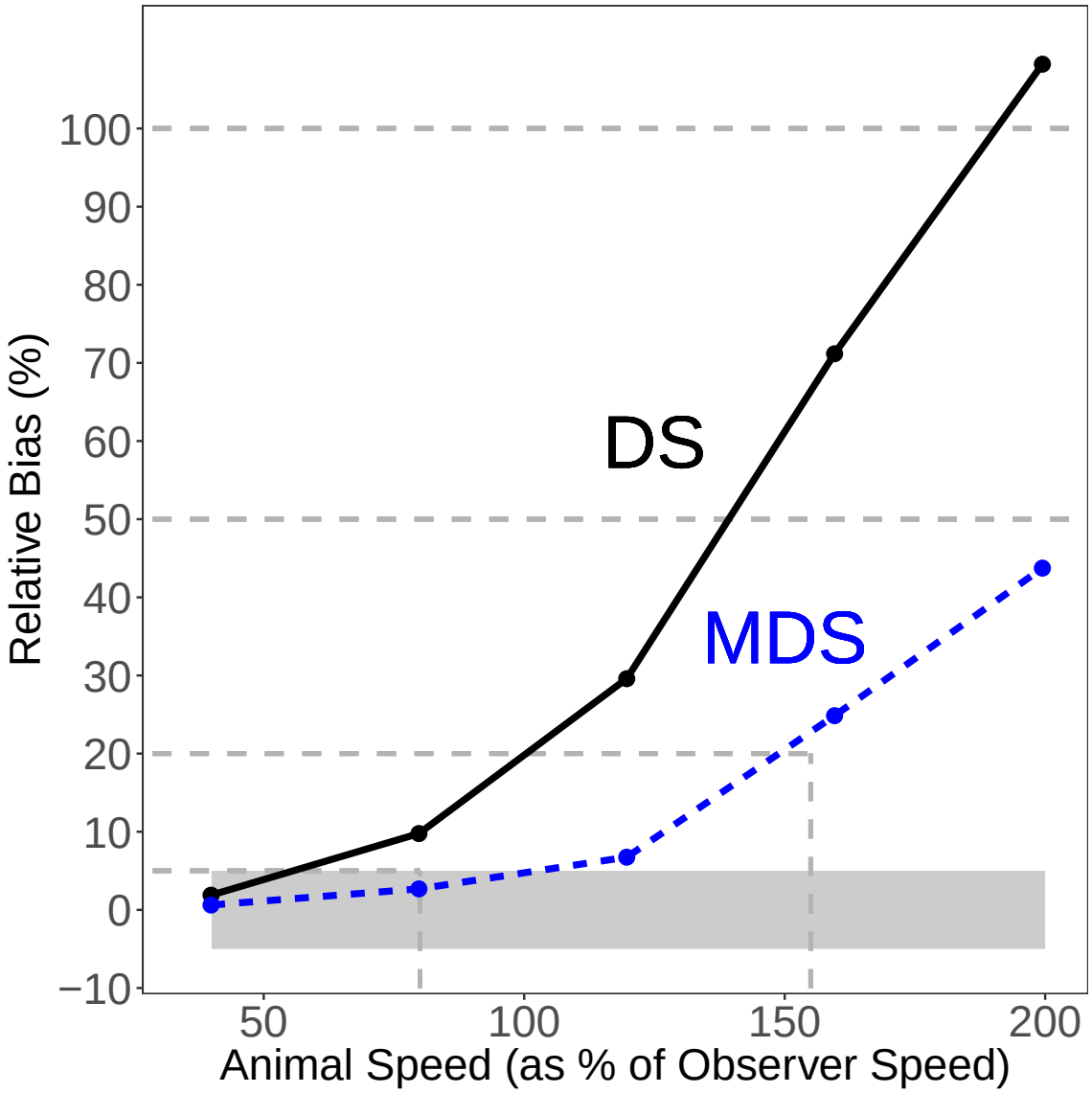


Figure 3.3 *Percentage relative bias in estimated density for conventional distance sampling (solid black line) and distance sampling with movement incorporated (dotted blue line) against animal speed (as % of observer speed) estimated from* 100 *simulations of a line transect survey of* 50 *transects with truncation width* 30 *metres and observer speed* 1 *metre per second. Animals move in randomly-orientated straight lines. Shaded region marks* ±5% *relative bias.*

3.4 Common Issues in Distance Sampling

In this section, three common issues encountered in distance sampling analyses are discussed. These issues arise due to the way distance sampling surveys occur in practice.

3.4.1 Marginalisation

In Section 3.2, the exact location in two-dimensions of each detected individual was assumed to be observed. The use of a two-dimensional detection process is less common in the application of distance sampling. One-dimensional detection processes are considered more robust Hayes and Buckland (1983). For MDS, inference can be based on the observed distances and detection times, rather than two-dimensional location. This is achieved by simple marginalisation over the appropriate distance levels.

For line transects, the perpendicular distance is measured; thus, the marginal PDF of a perpendicular distance y at time t is obtained from the PDF of the two-dimensional location $\boldsymbol{x} = (y, z)$ by integration:

$$f(y,t) = \int_{-\infty}^{\infty} f(\boldsymbol{x},t)\ \mathrm{d}z.$$

For point transects, the radial distance is measured; thus, the marginal PDF of this distance r at time t is easiest to obtain by changing to polar coordinates. Thus, let $f(r,\phi,t)$ be the PDF obtained by polar transformation of the Euclidean PDF $f(\boldsymbol{x},t)$, where $f(r,\phi,t) = rf(\boldsymbol{x},t)$. The radial distance PDF is then obtained by marginalisation:

$$f(r,t) = \int_{0}^{2\pi} f(r,\phi,t)\ \mathrm{d}\phi.$$

Inference based on marginal likelihoods has less information than using the two-dimensional locations of individuals. For many distance sampling surveys, records of the 2D locations is commonplace or easily made relative to measuring distances alone; thus, it is recommended that 2D locations be recorded.

Marginalisation over detection time t is also possible, providing the density of recorded locations or distances (if further marginalisation is performed) for the survey. In practice, detection time is routinely recorded in distance sampling surveys and so there is no need to marginalise over time. Furthermore, the time it takes for an individual to be detected provides, given an assumed movement model, a large amount of information about where the individual was located when surveying began. It is this distance moved by the individual

in the meanwhile that is responsible for the bias in CDS and so is the critical inference required for MDS models.

3.4.2 Grouped Data

In Sections 3.1 and 3.2, it is assumed that during the survey the exact distance or location and the exact detection time are recorded for each detected individual. In practice, distances and locations may be grouped together on an ordinal scale. For example, for point transects, rather than a radial distance measured, space may be segmented into bands with boundaries, often termed cut-off points, $[0, 100), [100, 200), [200, 300), \ldots$.

When an individual is detected, the radial distance may be measured and then later assigned to the appropriate group or only the group may be recorded. Data grouping is used when distances cannot be recorded reliably and so reducing the resolution of the data (from exact measurements to grouped measurements) is thought to provide more robust estimation (Buckland et al., 2001). In some cases, measurement of exact locations is not possible during the survey and so an ordinal scale is necessary. Grouped data in CDS leads to a count of how many individuals were seen in each distance band; these counts are described by a multinomial distribution. A similar development is possible with MDS.

For MDS, the space around the observer can be segmented in any way. Common choices may be a polar segmentation or a Euclidean segmentation.

Time could also be segmented into blocks or could remain an exact measurement. When detection time is recorded exactly, each detection consists of the time t_i and the cell in the detection grid that the individual was seen in, c_i. The PDF of this observation is obtained by integrating over the appropriate cell. For a Euclidean cell spanning perpendicular distances w_1, w_2 and forward distances l_1, l_2, the PDF is

$$\int_{w_1}^{w_2} \int_{l_1}^{l_2} f(\boldsymbol{x}, t) \, \mathrm{d}x_2 \, \mathrm{d}x_1$$

where $\boldsymbol{x} = (x_1, x_2)$. The PDF for the radial case is found using the same polar transformation as in the previous section.

When time is also segmented into blocks, each detection recorded consists of the spatial and temporal cell the individual was sighted in, c_i. The sufficient statistics for these data are then the counts of the number of individuals seen in each spatio-temporal cell $m_1, \ldots, m_C$ where C is the total number of cells. These counts have a multinomial distribution where the probability of an individual being included in cell c that spans spatial locations in the

set S and the time interval a, b is

$$\int_S \int_a^b f(\boldsymbol{x}, t) \, \mathrm{d}t \, \mathrm{d}\boldsymbol{x}$$

3.4.3 Observer Effort

In the ideal model presented, observers survey transects continuously during some time interval $[0, T]$ for a total survey time T. In reality, there are periods within this interval where the observer ceases to survey. The reason for this could be for the observer to rest, for one observer to replace another, or for the observer to further investigate a recent sighting. The total time the observer spends surveying within the time interval is termed effort. For CDS, this poses no complication because only the time spent on effort need be considered. For MDS, the time the observer spends off effort must be known because during this time animals continue to move.

Time spent off effort is equivalent to assuming that the encounter intensity during this time is zero. Let $e(t)$ be one when the observer is on effort at time t and zero otherwise. Thus, for a given encounter intensity λ, the intensity for the survey can written as $e(t)\lambda(\boldsymbol{x}, t)$. When averaging over all an individual's path, movement during off effort time is accounted for while detection probability during that time is zero.

This has a practical implication. Effort is routinely recorded in distance sampling surveys; however, in some instances only the total time spent surveying is recorded and not the time spent off effort. In these cases, MDS can be used but the probability of detecting an individual may be underestimated. For example, if an observer goes off effort temporarily and then upon resuming the survey immediately detects an individual very close, it is most likely this individual moved to this location whilst the observer was off-effort; however, if this period of off-effort is ignored, the only possibility is that the observer failed to detect this individual as it moved closer, implying that detection probability is low.

3.4.4 Measurement Error

One assumption of CDS and MDS is that measurements are made without error. In distance sampling surveys, exact locations and exact detection times may contain error due to the devices used to record this information or due to human involvement. When error is present, one possible remedy is to group data as described in Section 3.4.2. Alternatively, Marques (2004) developed measurement error models to account for this. These can be readily included in MDS models. In Section 3.2, the MDS likelihood presumed that the exact

location was recorded during an encounter. Suppose instead that the recorded location $\boldsymbol{y} \sim \mathrm{N}(\boldsymbol{x}, \epsilon^2 \boldsymbol{I})$ has a bivariate normal distribution with mean equal to the true location and standard deviation, representing the error, ϵ in each direction ($\boldsymbol{I}$ is the 2×2 identity matrix). The PDF of the observed data is then given by

$$[\boldsymbol{y}, t] = \int [\boldsymbol{y} \mid r(\boldsymbol{x}, t)][r(\boldsymbol{x}, t) \mid \vec{\boldsymbol{x}}][\vec{\boldsymbol{x}}] \, \mathrm{d}\vec{\boldsymbol{x}}$$

where $\vec{\boldsymbol{x}}(t)$ is the location of the animal at time t and $r(\boldsymbol{x}, t)$ is the event that the animal is detected whilst in occupying location $\boldsymbol{x}$ at time t.

Furthermore the error ϵ can depend on the true location. In many cases, this is likely to be the case since the recorded locations further from the observer are likely to have greater error than those close by.

3.4.5 Clusters

For many populations, individuals move around in groups or clusters. In CDS, each group is treated as a single object and the above analysis conducted treating clusters as individuals. The resultant density estimate is then the density of clusters, $\hat{D}_c$. To estimate population density, $\hat{D}$, the density of clusters is multiplied by the estimated mean cluster size, $\hat{\mathbb{E}}(s)$. Mean cluster size is estimated from the recorded sizes of each detected cluster. As larger clusters are more likely to be detected, this can lead to an overestimated mean cluster size, thus a regression based estimate is used, regressing cluster size against distance (Buckland et al., 2001). Given this, the population density is estimated $\hat{D} = \hat{\mathbb{E}}(s)\hat{D}_c$. This approach rests upon the snapshot assumption: clusters cannot fuse or break apart in a snapshot of time.

For MDS, the snapshot assumption is abandoned and so cannot be used to justify using clusters as single, indivisible units. To do so, the assumption must be made that clusters do not break apart or fuse during a survey. In reality, this assumption has the same disadvantage as when it is made in CDS analyses. Explicit, continuous-time movement models for animal groups are rare in the literature (Langrock et al., 2014; Scharf, Hooten, Johnson, & Durban, 2018) and often require additional latent modelling to be used with the existing models for latent individual paths. For some species, clusters can consist of hundreds of individuals; simultaneous models for the movement of all such individuals in the group is not possible given the current computational methods. Future research to incorporate group movement with MDS is discussed in Chapter 6. In this chapter, clusters will be treated as individuals.

3.5 Case Study: ETP Dolphins

Distance sampling methods are routinely applied in shipboard surveys of cetaceans. The effect of non-responsive animal movement on the inference obtained from these surveys is not well understood. A rule of thumb is used in line transect surveys: if the individuals move at a speed less than half the speed of the observer, then bias from animal movement is negligible. This heuristic is based upon a limited simulation study (Hiby, 1982) which considered a detection process with one set of parameters. The bias caused by movement is a function not only of the relative speed of the animal but also of the detection process and the size of the survey plots. These effects are non-linear and difficult to assess. MDS provides the possibility of incorporating knowledge of how the target species moves and investigating whether this reality has a practically significant effect on density estimation. Further to that, MDS provides the alternative required if movement is discovered to be a substantial source of bias. This investigation is critical because the inference gained from distance sampling analyses is used to inform decisions on the policy and practice of organisations that manage and conserve animal populations.

In this section, MDS analysis is applied to a line transect survey of spotted dolphins in the Eastern Tropical Pacific. Both the detection and movement models are fit together, using a joint likelihood. The inference from MDS is compared to that obtained using CDS to determine if dolphin movement may affect density estimation.

3.5.1 Survey

The survey was conducted in 2006 in the eastern tropical Pacific as part of a series of surveys across many years, see Gerrodette and Forcada (2005) for more details on the entire survey protocol. Here, the analysis concentrates on estimating the number of dolphin schools within the core area (defined by Gerrodette and Forcada (2005)). As discussed in Section 3.4.5, dolphin schools are treated as individual units of detection and group size is estimated separately. During the survey, for each dolphin school encountered, the centroid of the school and the exact time of the encounter were recorded. For this analysis, only sightings in Beaufort state two or less were retained as the detection process varies under highly heterogeneous visibility. In high Beaufort sea states, in particular, dolphins can move close to the observer and evade detection; such detections have a high influence on the detection probability estimator and such sensitivity can cause overestimation in density. An important future extension of this analysis would incorporate these environmental covariates into the MDS model, allowing for all sightings and effort to be used to estimate abundance.

Overall, there were 40 sightings made of spotted dolphins within a perpendicular distance of 5 kilometres from the transect line. The truncation distance of 5 kilometres was choosen as the 95%)$quantile of the observed perpendicular distances$.

During the survey, after each encounter, the ship entered an off-effort phase termed "closing mode". This was used to validate the recorded location of each cluster, the recorded cluster size, and species ID. The times whenever the observer went on and off effort were recorded. This was accounted for as described in Section 3.4.3.

Dolphin school movement was described by Brownian motion. Independent tag data on nine spotted dolphins was used to estimate the movement parameters. The location of each dolphin was recorded approximately every fifteen minutes over a period of 1–2 days. Full details of the tracking study are given by Scott and Chivers (2009). Average dolphin speed was 7.4 km/h. The speed of the observing ship was on average 17 km/h and so, by the commonly used heuristic, bias from animal movement is expected to be negligible.

Finally, dolphin school size was estimated using the regression method (Buckland et al., 2001). The mean school size was estimated to be 164 from the 104 sightings made during the survey.

3.5.2 Results

The estimated abundance in the core area differed substantially between CDS and MDS. Incorporating animal movement reduced the density estimate by 22% (33292 animals) (Table 3.1).

	Estimate	CV(%)	LCL	UCL
CDS density	910	15.5	634	1186
MDS density	707	12.5	533	880

Table 3.1 *Maximum likelihood estimates of spotted dolphin school density (per* $10^6 km^2$*) with coefficient of variation (CV) and lower and upper* 95% *confidence interval bounds for conventional distance sampling (CDS) and distance sampling with movement (MDS)*

The large reduction in the abundance estimate indicates that even though the dolphins move relatively slowly compared to the ship, bias can be substantial, because, whilst being surveyed, they can move a large distance compared to the width of the transect. This highlights the danger of assessing whether movement is a problem based solely on relative animal speed; MDS can account for the interdependent effects of animal speed, transect width, and detection function shape.

The goodness-of-fit of both CDS and MDS can be assessed by considering the estimated

probability density of the observed perpendicular distances (Figure 3.4). MDS had a similar goodness of fit as CDS to the data (chi-squared test gives p-value of 0.31 for CDS and 0.35 for MDS); however, the estimated detection function differs considerably between the two methods (Figure 3.5). The CDS estimated detection function has a narrower shoulder and smaller detection scale indicating that animal movement has caused negative bias in the estimation of detection probability. If the survey had indeed taken place in a snapshot of time, CDS estimates the probability of an animal being detected, given it is inside the transect, to be 0.47; MDS estimates this to be 0.62. Note, this deficiency does not result in a marked difference in goodness-of-fit to the observed data, but has an important effect on the final abundance estimate.

The goodness-of-fit of the movement model can also be assessed. Under Brownian motion, the movements between two recorded locations are assumed to be independent Gaussian random variables with mean zero and standard deviation $\hat{\sigma}\sqrt{\Delta t}$ where $\hat{\sigma}$ is the estimated movement rate and Δt is time between the two records. For each tagged individual, a residual was computed for every recorded movement by computing the cumulative probability of that movement under the estimated movement model. Under the probability integral transform, these residuals ought to be uniformly distributed. For ease of interpretation, these residuals were converted to Gaussian quantiles, and so these transformed residuals ought to be normally distributed. For the nineteen individuals, the Shapiro-Wilks test failed to reject the hypothesis the residuals were normally distributed for fourteen individuals. Appendix B, Figure S2 contains quantile-quantile plots for each individual. By visual inspection, the movement model has adequate fit with evidence that residuals from individuals $6 - -10$ have unusual distributions. All these individuals were tagged in 1993, compared to the other individuals which were tagged in other years. This may suggest a systematic rounding of recorded movements for these individuals compared to others. Furthermore, a runs test (REF) was performed on each individual's series of residuals to assess whether residuals were independent; the hypothesis was independence was rejected for all but six individuals, suggesting that movement is correlated through time. This is unsurprising, but highlights an important extension for future work.

Thus, when knowledge of how animals move is available, MDS can be used to assess whether CDS inference is reliable. This example shows that the simple heuristic used is not infallible and applying MDS to other current applications to distance sampling may also lead to CDS estimates being questioned.

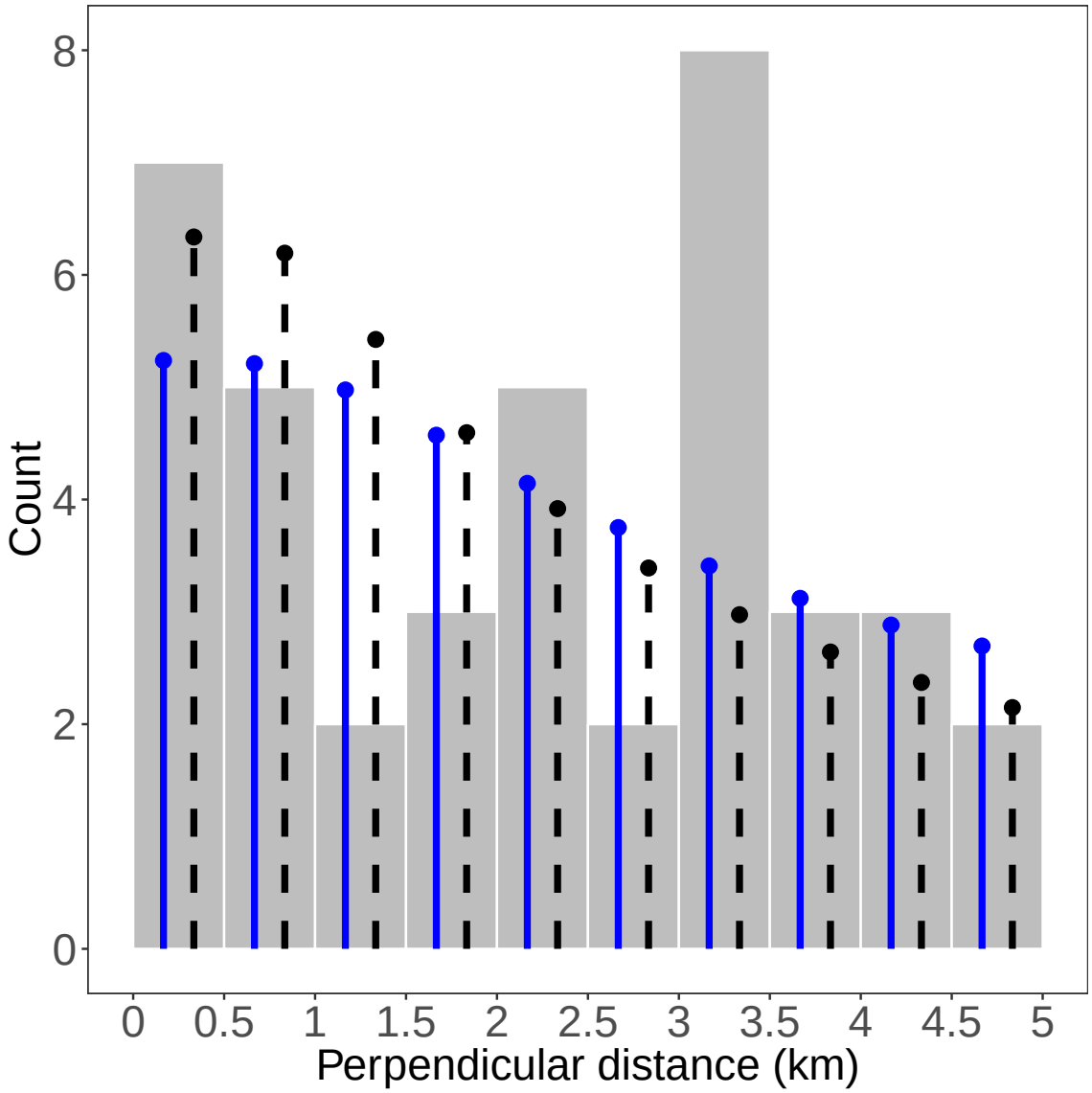


Figure 3.4 *Observed number of spotted dolphin schools sighted in each* 0.5 *km perpendicular distance from the transect line (shaded bars) with expected number of sightings from conventional distance sampling model (dashed lines) and distance sampling with movement (solid lines)*

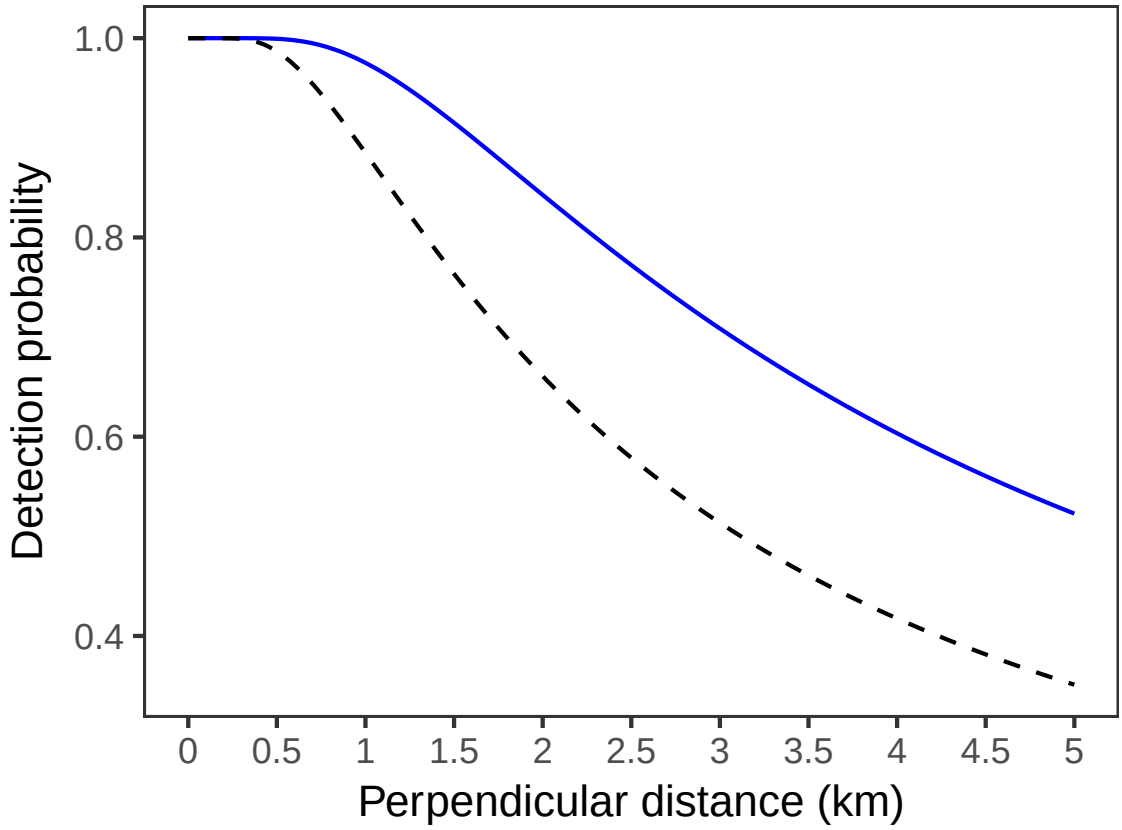


Figure 3.5 *Estimated detection function for a hypothetically immobile spotted dolphin population for conventional distance sampling (dashed line) and distance sampling with movement (solid line)*

3.6 State-Switching Distance Sampling

During a survey, individuals can switch behaviours, changing their movement and detectability. A common example of this is found when distance sampling is conducted by ship on cetaceans. Cetaceans dive and yet can only be sighted when they are at or near the sea surface. Thus, at any one moment in time, only part of the population within a transect are available for detection. Hence, the resultant density estimate will not be an estimate of the total population density but an estimate of those available, who make up an unknown proportion of the population. This is termed availability bias.

In CDS, it is reasoned that as distance sampling surveys are snapshots of the population, one must only determine the proportion of individuals available for detection in this snapshot. Scientific observation and telemetry can be used to estimate the proportion of time an animal is available and so quantify the proportion of the population that was available during any single snapshot, $\hat{p}_a$. The density of the entire population, D, is then found using this estimate and the estimate of the available population density D_a:

$$\hat{D} = \frac{\hat{D}_a}{\hat{p}_a}.$$

The quantity $\hat{p}_a$ is termed a multiplier in distance sampling literature. When applied, the uncertainty in $\hat{p}_a$ is often ignored or incorporated with the delta or bootstrap method.

In MDS, the multiplier approach cannot be justified by the snapshot assumption. Instead, the availability process can be modelled explicitly. Borchers et al. (2013) developed an explicit hidden Markov model for the availability process within CDS, acknowledging that surveys take a non-zero time to conduct and so the estimate $\hat{p}_a$ of the instantaneous probability of being available is likely an underestimate of the chance an individual is available at some time during the survey. Abstractly, this model describes state-switching distance sampling where animals can switch between different states (behaviours) that can affect their detectability. For availability models, the states are available and unavailable where in the latter an animal is undetectable. Borchers et al. (2013) use tags to estimate state-switching parameters of the availability process.

In Section 2.4.4, state-switching encounter models were developed. While individuals move along their unobserved path, they can switch between states, affecting how they move and the intensity of their encounters with detectors. Just as encounter models can be used to generalise CDS to MDS, the same approach can be used to generalise state-switching distance sampling (SS-DS) to state-switching distance sampling with movement (SS-MDS).

For an availability process, in particular, both the movement and detection processes can change depending on whether the animal was available at time t, $a_t = 1$, or unavailable $a_t = 0$. The movement model can switch between two sets of parameters or indeed switches can be made between two different movement models depending on the state of the individual. Furthermore, the encounter intensity at time t can be specified as $a_t\lambda(\boldsymbol{x}, t)$, that is, the animal cannot be detected when unavailable. This is equivalent to the observer effort adjustment in Section 3.4.3. The difference is that the availability history of each individual is unknown, just as its movement path is unknown. Hence, not only must one average over all movement paths, but also one must average over all possible state-switching histories.

The detection parameters are estimated from the observed locations and times of the encounters. To estimate the movement parameters auxiliary information on how individuals move is required. Similarly, to estimate the state-switching process, auxiliary information on this process is required. For cetaceans, availability is linked to their depth in the water column. A simple approach would be to use telemetry data. As described by Borchers et al. (2013), the depth profile of each tag can be split into periods of "availability" and "unavailability" based on a specified depth threshold. Another example is songbirds, where birds can switch between perched singing and silent movement, affecting their movement and detectability; acoustic tags or human observation can be used to record data on how birds switch between these two behaviours.

Given this, the state-switching process is observed and a hidden Markov model can be used to estimate the switching parameters. Behaviour is highly heterogeneous and tagging studies often have small sample sizes, thus variation between individuals is likely to be high and bias in state-switching estimates substantial, in some cases. Thus, extending inference from these small sample observations to the entire population must be done with caution. Let $\mathcal{L}_s(\boldsymbol{\theta}_s)$ be the hidden Markov model likelihood of the state-switching process with parameters $\boldsymbol{\theta}_s$.

As when incorporating information on individual movement, one can assume that the observed series of state-switches are made on individuals that are independent of those observed in the distance sampling survey. The joint likelihood is then the multiplication of the likelihoods for each component:

$$\mathcal{L}_{e,s,m \mid n} = \mathcal{L}_{e \mid n}\mathcal{L}_m\mathcal{L}_s.$$

Maximising this conditional likelihood would provide estimates of detection, movement, and state-switching parameters simultaneously.

3.6.1 Availability Simulation Study

A simulation study was conducted to assess the performance of SS-MDS for an availability process. A shipboard cetacean line transect survey is simulated with 100 transects. The survey area is 100 square kilometres and the true population size is 100 cetaceans. The same detection process is used as in the simulation study in Section 3.3 and cetaceans are again assumed to move according to a diffusion model. Here, animals can switch between available and unavailable states. The average speed of the animal is denoted ν. Animals were not detectable when unavailable.

The average time (in minutes) spent by a cetacean in the available state is denoted τ_a and the time spent unavailable, τ_u. Each simulation scenario consisted of specifying values of ν, τ_a, τ_u. The selected values were chosen to emulate the behaviour of Bowhead whales (Laidre, Heide-Jørgensen, & Nielsen, 2007) and Beaked whales (Borchers et al., 2013). Laidre et al. (2007) found Bowhead whales had dive durations that ranges from 3 to 18 minutes with surface intervals between 4 and 30 minutes from 44 whales fitted with satellite or time-depth recorders. Borchers et al. (2013) used reported mean dive and duration parameters to specify dive and surface durations for Beaked whales of 2 and 26 minutes respectively. All other variables in the simulation where held constant. Each scenario was simulated 500 times.

For each simulation, tag data on ten individuals was simulated to inform the movement model. The availability history of each animal was simulated as a Markov chain such that on average an animal spent τ_a time available and τ_u time unavailable. Results are given in Table 3.2.

A similar simulation study was conducted by Borchers et al. (2013) to demonstrate the efficacy of modelling the availability process with a HMM. This simulation study is intended to show that the performance is similar once movement is incorporated.

SS-MDS was negligibly biased across the scenarios considered and confidence interval coverage was close to nominal. Notably, relative bias increased slightly for the long diving scenarios (scenarios 10–12). As longer divers are more likely to remain undetected, sample size for these scenarios was lower, resulting in higher bias. Overall, however, the simulation study proves the concept of SS-MDS and indicates that the method can perform well under a realistic scenario. For cetaceans, where movement and availability are both issues that need to be addressed, SS-MDS can be used to incorporate knowledge about movement, diving, and detectability to produce an improved density estimator and to propagate the uncertainty that arises from each of these independent data sources.

Scenario	ν	τ_u	τ_a	b	CI
1	0.5	5	5	-0.9	95
2	0.8	5	5	-0.6	98
3	1.0	5	5	-0.1	94
4	0.5	5	10	-0.3	96
5	0.8	5	10	0.7	93
6	1.0	5	10	0.5	97
7	0.5	10	5	-0.6	97
8	0.8	10	5	0.1	95
9	1.0	10	5	-0.2	96
10	0.5	30	2	1.4	96
11	0.8	30	2	2.0	94
12	1.0	30	2	2.2	94

Table 3.2 *Percentage Bias (b) in density estimation and estimated* 95% *confidence interval coverage (CI) for a simulated cetecean shipboard line transect survey where cetaceans move at speed* ν *relative to the ship and are unavailable for detection whilst in a dive. Dives have mean duration* τ_u *(time unavailable in minutes) and mean surface time is* τ_a *(time available in minutes). Each simulation consisted of* 100 *transects surveyed for* 8 *hours each.*

3.7 Case Study: Seabirds

State-switching MDS can further be used to extend the application of distance sampling to fast-moving species. There has long been a need for statistical methods to analyse distance sampling ship surveys of seabirds in flight. Ronconi and Burger (2009) in a review of seabird abundance estimation stated that "Application of distance sampling to birds in flight is a major obstacle that must be addressed." Yet, the speed of seabird movement relative to the ship has prohibited the use of distance sampling methods as the snapshot assumption is grossly violated. One approach is to only record seabirds that pass abeam; doing so is equivalent to a snapshot method. This approach has not been widely adopted. Typically, strip transect methods are used, despite the fact that these methods are similarly based upon a snapshot assumption and so can be prone to overestimating density also. For plot sampling, various *ad hoc* methods are used to minimise any bias resulting from seabird movement and mathematical methods have been used to account for movement by using the heading and speed of each sighted bird (Spear, Ainley, Hardesty, Howell, & Webb, 2004).

CDS has been used to estimate the population density of seabirds on the sea surface. When resting on the surface, the seabirds move slowly and so the snapshot assumption is met. As in the availability case, this results in an estimate of the density of resting seabirds and so an independent estimate of the time each bird spends on the sea surface is required. The

multiplier approach is then taken to estimate the total population size. This approach has two disadvantages. First, it can only use records of encounters with birds on the sea surface, birds in flight must be ignored. Secondly, CDS does not account for the reality that some birds are detected in flight and then only recorded once they land on the sea surface; hence, bias from animal movement is still present and may be further pronounced since birds may be recorded closer to the observer than they would have been if recorded when first sighted in flight.

State-switching MDS provides the statistical methods to estimate seabird population density using encounters from birds both in flight and on the sea surface. Independent tag data can be used to estimate bird movement and the rate at which birds switch between two states: flying and resting on the sea surface. It should be noted, as discussed in Section 3.2.4, that independent tag data can also be used to adjust plot counts, thereby generalising the popular strip transect methods used at sea.

3.7.1 Balearic Shearwaters

A shipboard survey of Balearic shearwaters (*Puffinus mauretanicus*) was conducted in the Bay of Biscay for four years (2013–2016). One hundred and forty-five line transects were surveyed across the Bay by ship at an average speed of 15 km/h (Figure 3.6). Sixty-nine shearwaters were detected, though only two were seen on the sea surface. Shearwaters were recorded in flocks with mean cluster size 3.01. For each encounter, the exact sighting time and location were recorded.

The aim is to fit a behaviour-switching MDS model to these data where detection and movement parameters are estimated jointly.

Movement Analysis

Independent tag data on seven shearwaters in the Mediterranean were used to estimate seabird movement in flight and when resting. Exploratory analysis shows the two possible states that govern the velocity process (Figure 3.7). A state-switching diffusion model was assumed to describe seabird movement with two behavioural states presumed to describe flight and rest. Velocity in the resting state is denoted ν_0 and denoted ν_1 in flight. Over a very long time period, the proportion of time spent in the resting state is denoted τ_0 and τ_1 is the proportion of time spent in flight. As discussed in Section 2.3, there are several approaches to fitting this movement model (Hooten, Johnson, et al., 2017). They can be classed as either velocity-based or location-based. The simpler approach is velocity-

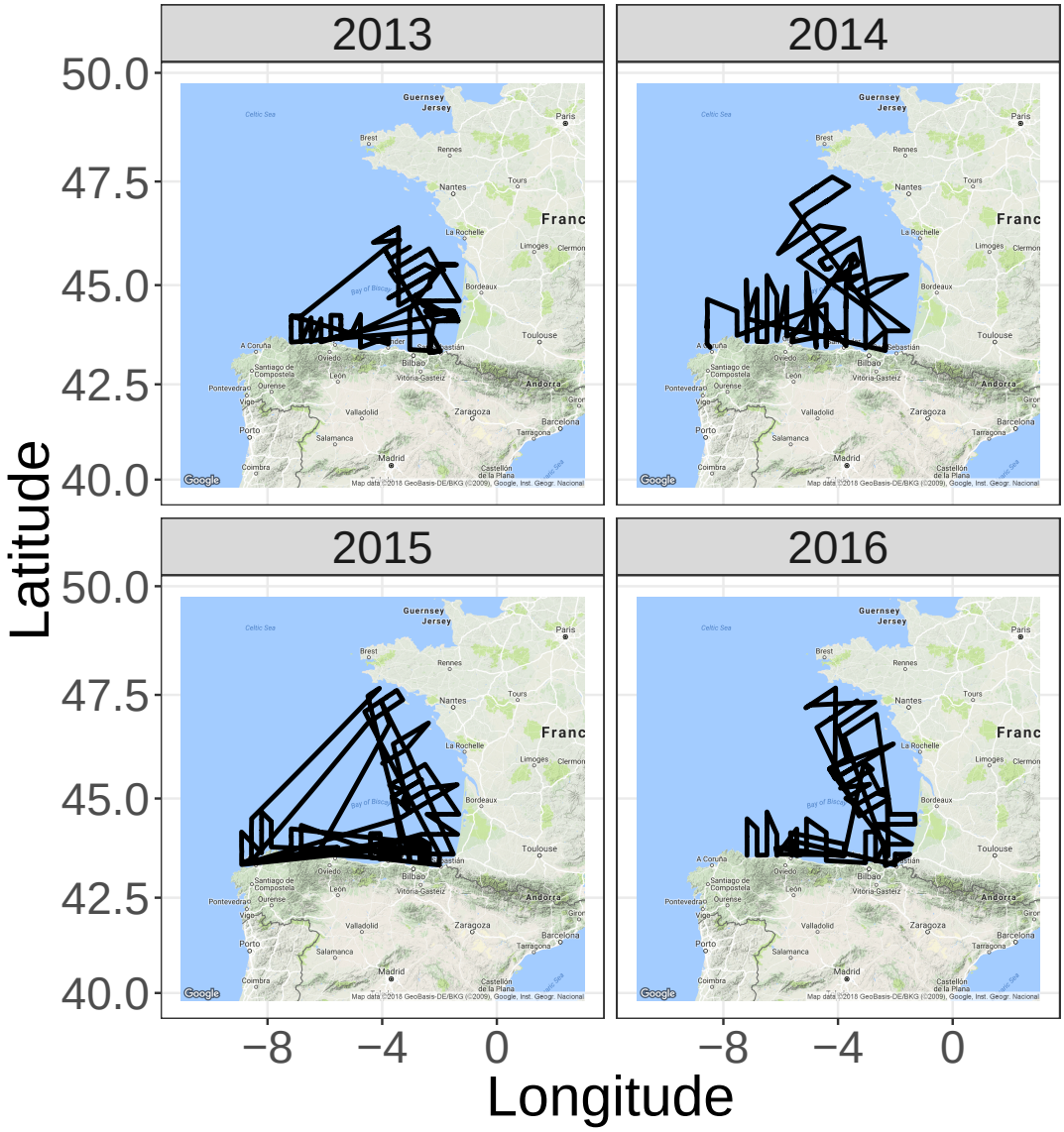


Figure 3.6 *Transect lines surveyed (solid lines) in Balearic shearwater shipboard line transect survey in each year.*

based where the average velocity is computed between each recorded tag location and the parameters of the movement model estimated from these differences (Table 3.3).

The second approach, considerably more computationally expensive, is a location-based algorithm where the recorded locations themselves are used and the path taken by each shearwater in between location fixes is averaged over (Table 3.4).

	Estimate	LCL	UCL
ν_0	1.18	1.12	1.25
ν_1	38.24	34.16	43.91
τ_0	0.62	0.61	0.63
τ_1	0.38	0.37	0.39

Table 3.3 *Maximum likelihood estimates of velocity-based shearwater movement model fit to* 7 *tracks with lower and upper* 95% *confidence interval bounds for mean speed* ν_b *and mean proportion of time* τ_b *in each state* b

	Estimate	LCL	UCL
ν_0	0.81	0.38	1.52
ν_1	36.92	35.71	38.41
τ_0	0.65	0.61	0.70
τ_1	0.35	0.30	0.39

Table 3.4 *Maximum likelihood estimates of location-based shearwater movement model fit to* 7 *tracks with lower and upper* 95% *confidence interval bounds for mean speed* ν_b *and mean proportion of time* τ_b *in each state* b

Fitting each model to the tag data provided similar results, though uncertainty in the location-based estimates was higher, indicating that velocity-based uncertainty is underestimated as it does not recognise the dependence between velocities other than the dependence induced by the behaviour process.

The velocity-based approach was used in the MDS analysis. The location-based model is substantially more computationally intensive and its use within the MDS model severely limits the spatial and temporal resolutions that can be used to approximate the path integration. This ultimately leads to poorer inference. More computationally efficient behaviour-switching continuous-time movement models have not been developed in a likelihood-based framework (Pedersen et al., 2011). Future development of such models would allow for location-based movement analyses to be used with MDS models.

Many seabird movement analyses do not use explicit movement models. Instead, a threshold velocity below which the shearwaters are assumed to be resting, and above which assumed to be in flight, is used. A common threshold used is two kilometres per hour. For the

shearwater tag data, this would lead to empirical parameters similar to the model-based estimates: $\hat{\nu}_0 = 1.43$, $\hat{\nu}_1 = 31.72$, $\hat{\tau}_0 = 0.71$, $\hat{\tau}_1 = 0.29$.

The goodness-of-fit of the velocity-based movement model was assessed using the Gaussian pseudo-residuals defined by (Zucchini et al., 2016). This is similar to the goodness-of-fit analysis by REF. These residuals were computed for each individual. The Shapiro-Wilks test rejected the hypothesis these residuals were normally distributed for every individual. Quantile-quantile plots (Figure S4) indicate that the distribution of residuals has heavier tails than the normal distribution.

MDS analysis

Due to the low number of sightings of birds on the sea surface, a model where detectability changed with behavioural state was not considered. A hazard of the form $\alpha r^{-\beta}$ where r is the radial distance between the observer and the animal was used.

Population density in the Bay of Biscay was estimated for each year, assuming common detectability across years (Table 3.5). It is difficult to compare densities among years due to differing coverage of the study area. Higher densities of shearwaters have been observed in the north of the Bay (Jones et al., 2014) and this is surveyed less in the first two years compared to the last. This may explain the large difference in estimated densities. The average abundance across all years was 0.06 individuals per square kilometre.

	Density	CV	LCL	UCL
2013	0.0032	0.7200	0.0009	0.0112
2014	0.0107	0.6700	0.0033	0.0350
2015	0.0767	0.7100	0.0221	0.2664
2016	0.1300	0.9800	0.0262	0.6486
Pooled	0.0580	0.5400	0.0214	0.1569

Table 3.5 *Maximum likelihood estimates of shearwater density (to the nearest integer) per* 100 *square kilometres for years* 2013–2016 *and pooled across years with coefficient of variation (CV), lower and upper* 95% *confidence interval bounds*

Goodness of fit of the model was tested by comparing the estimated PDF of observed perpendicular distances to the observed distribution (Figure 3.8). The Kolmogorov-Smirnov test failed to reject the fit of the model at the 5% significance level.

This example demonstrates how state-switching MDS can be used to apply distance sampling in a situation to which it was hitherto inapplicable. It is anticipated that this method will provide an alternative to the current approach taken by seabird surveys for both dis-

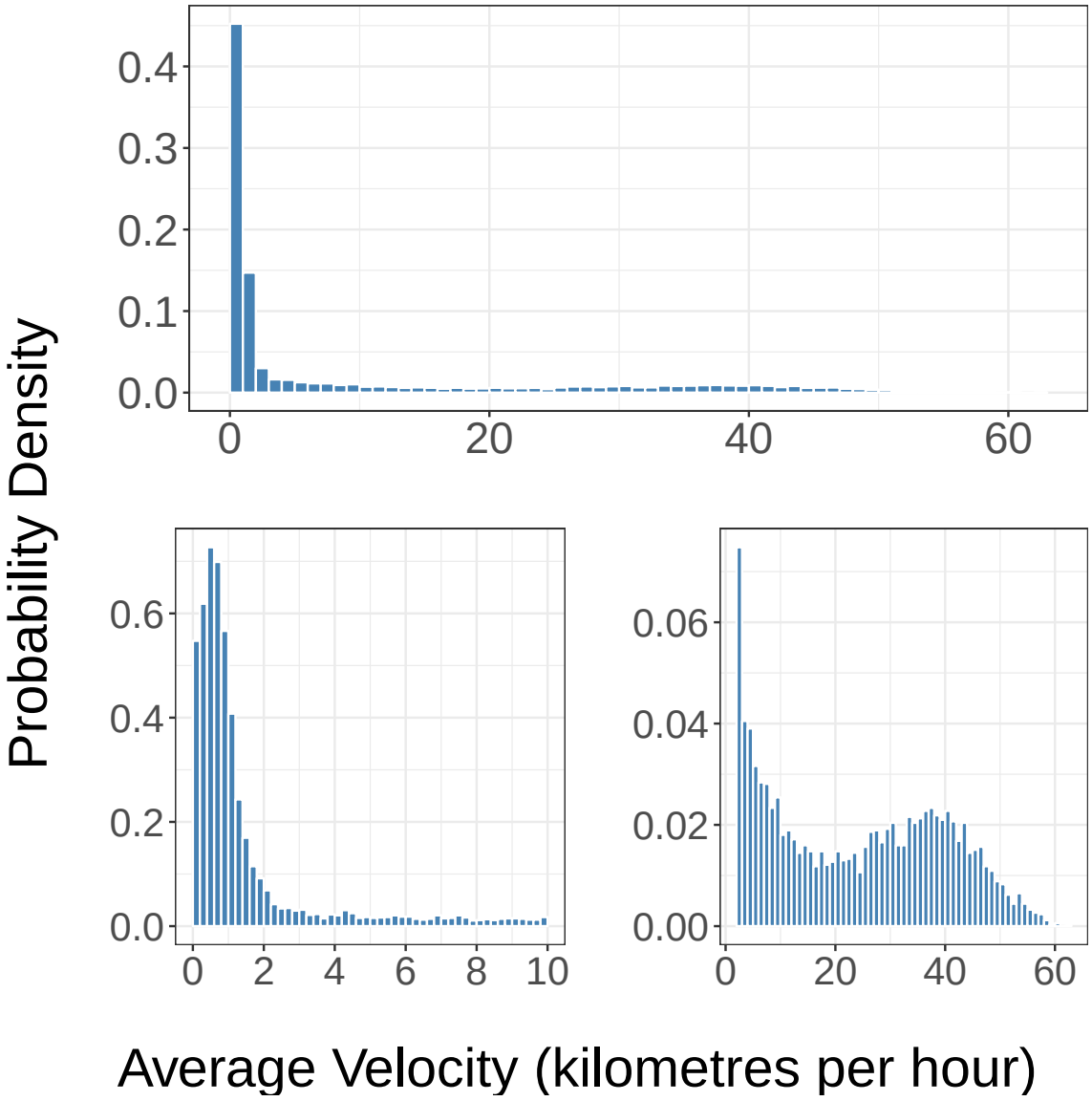


Figure 3.7 *Empirical probability density of average velocity (kilometres per hour) between recorded tag locations for seven tagged Balearic Shearwaters across all velocities (top), for all velocities below* 10 *kilometres per hour (bottom-left) and for those above* 10 *kilometres per hour (bottom-right).*

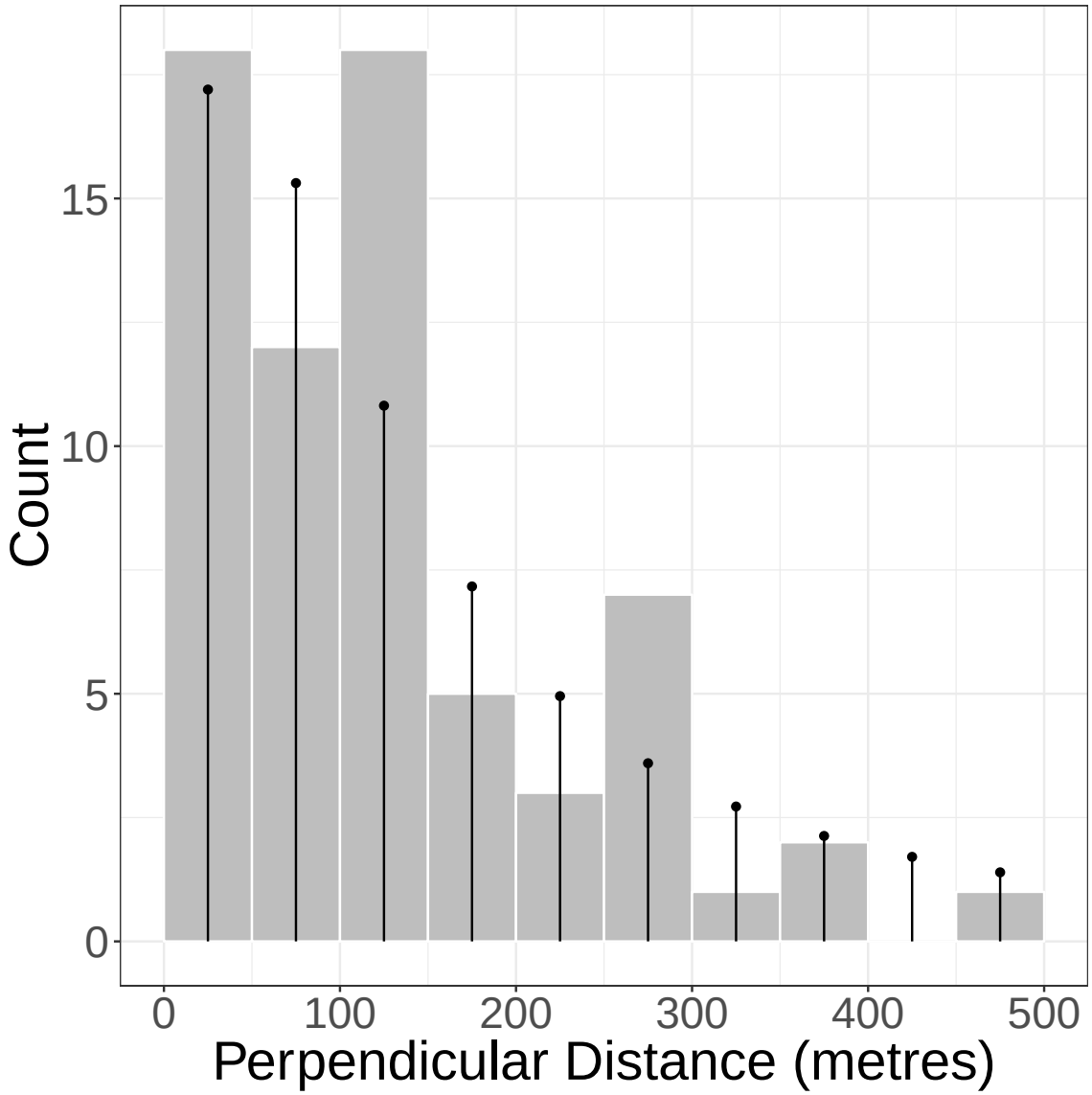


Figure 3.8 *Observed number of shearwater flocks sighted in each* 50 *m perpendicular distance from the transect line (shaded bars) with expected number of sightings from distance sampling with movement and two behavioural states (solid lines)*

tance sampling and strip sampling. There are widely conflicting abundance estimates of Balearic Shearwaters in the Bay (Authier et al., 2017; Pettex et al., 2016) and worldwide, partly due to their shifting distribution, but also due to the relative biases that occur in the different methods used to collect and analyse data. The Balearic shearwater is registered as an endangered species while its population size is highly uncertain. Thus, it is important to develop methods that can provide density estimates whilst accounting for significant bias.

SS-MDS is available to account for the most severe causes of bias: bird movement and uncertain detectability. Further work to develop distance sampling for seabirds in flight is key to making the best use of shipboard surveys, in particular SS-MDS can be extended to allow detectability and movement to change with the time of day and season when surveys take place.

3.8 Covariates in Distance Sampling

Section 2.4.6 described how covariates can be included in encounter models to account for spatial and temporal variation. Multiple covariate distance sampling (Marques & Buckland, 2003) is the name given to models where the detection parameters in CDS change with respect to observed covariates. These covariates are included using a link-function formulation. For example, for a hazard with scale parameter σ, covariates $z_1, \ldots, z_K$ can influence variation in σ through a log-link function:

$$\log(\sigma) = \beta_0 + \beta_1 z_1 + \ldots + \beta_K z_K$$

where $\beta_0, \ldots, \beta_K$ are parameters to be estimated. The same formulation can be used in MDS.

Additional to this, covariates can influence the movement. For example, the ideal free gas model can have a mean speed that depends on temporal or spatial covariates using a log-link function. The analytical formulations used in plot sampling cannot be used in this case; however, simple numerical integration can be used to compute the probability of detection across the spatio-temporal conditions that occurred during the survey. For diffusive movement, a similar approach could be taken for plot or distance sampling. This addition may be particularly important for long-term surveys where animals are likely to exhibit changing behaviour that affects detectability. For example, in Howe, Buckland, Després-Einspenner, and Kühl (2017), distance sampling applied to camera trap data was shown to be affected by the heterogeneous behaviour of animals between night and day; a

simple temporal covariate could be used in an MDS model with ideal free gas or diffusive movement to account for this.

Multiple covariate distance sampling methods focus on including covariates into the detection model; the analogous extension to MDS models involves covariates in the detection and movement models. Density surface modelling (DSM) (Miller et al., 2013) describes inclusion of covariates into the model for density. DSM extends the use of distance sampling to not only quantify population density but to infer population distribution. The idea is that the density surface is an estimate of the average population distribution during a snapshot in time. In reality, distance sampling surveys are conducted over days, weeks, months, or even years and so the meaning of the estimated surface is unclear. Practically, it is interpreted as the average density at each location over an infinitely long time period, that is, the stationary distribution of the population over the space (Yuan et al., 2016). The same two approaches can be used with MDS where the adjustment to the density estimate due to movement is made.

An open question is whether the use of MDS will only affect the estimation of absolute abundance or whether it will also affect the estimated distribution. When covariates are only included in the detection and density components of the model, there appears no reason to propose that the qualitative inference on animal distribution would change due to the use of MDS. Yet, if spatial covariates are included in the movement model, then there may be cause to suspect that MDS would propose substantially different spatial inference. Suppose individuals move more quickly in some areas compared to others within the survey region; this may be due to prey availability, habitat suitability or resources. For example, in a corridor connecting two important habitats, individuals may move quickly so as to minimise time spent in unsuitable terrain where they may be more exposed to predation. Consider a distance sampling survey conducted on such a population. It may be known *a priori* that animals prefer certain habitats, but rather than stratify the analysis, an alternative approach is to use DSM (Hedley & Buckland, 2004). The idea being that including a habitat covariate in the detection and density models would account for animal preferences (Miller et al., 2013). During the survey, animals move more slowly in their preferred habitat compared to the movement corridor between suitable habitats. Thus, the positive bias caused by animal movement is greater in the corridor compared to the suitable habitat and so even though fewer animals are sighted in this corridor, the density in this space is inflated and so the effect of habitat on animal density is considered to be smaller, or perhaps insignificant. In short, animal movement could wash out patterns in animal distribution. In other cases, where animals move more quickly when in their preferred

habitat, the opposite effect could occur, the pattern in distribution is exaggerated. Overall, it is possible that when the relationship between space and movement dynamics is not accounted for, this relationship can bias inference from DSM on animal distribution.

A similar application of covariates in distance sampling where MDS may have a profound effect is in distance sampling experiments. Conservation and agricultural managers require inference on whether a given action or treatment is effective in either reducing or increasing the density of a population. Thus, the treatment and a control are applied to different populations or different sections of the same population. Distance sampling is then used to estimate the density for each and a covariate (treatment or control) is used to estimate the effect of the treatment. The same issue arises as for DSM, in short, if the treatment affects animal movement, then without accounting for this, the estimated treatment effect will be biased and the inference obtained, and any action taken on the weight of this inference, inaccurate. It will depend on the situation whether the effect of animal movement is substantial enough to affect the final conclusions made from an experiment. Further research should focus on applying MDS to the estimation of spatial distribution and distance sampling experiments, ensuring that the effect of any covariates on animal movement is considered.

3.9 Mark-Recapture Distance Sampling

A common assumption in one-dimensional CDS is that the probability of detection on the line or at the central point is one, i.e, $g(0) = 1$. Theoretically, this assumption is not required for two-dimensional CDS or MDS. Borchers and Cox (2017) show, however, that for small sample size, estimation of $g(0)$ can be poor even when 2D data are used.

Mark-recapture distance sampling (MRDS) (Borchers et al., 1998) is a popular extension of CDS that allows for the case where $g(0) < 1$. To achieve this, two observers are required. Chapter 2 presents encounter models for more than one detector during a survey and these methods can be used to incorporate movement into existing MRDS models. There are two possible survey methods that can be used in MRDS: trial-based or independent-searching.

3.9.1 Independent Searching

Suppose each observer searches for individuals independently. Let λ_j be the encounter intensity associated with observer $j = 1, 2$. In Chapter 2, observers are assumed to be independent given the path each individual has travelled: their detections are correlated

only due to the movement of the individual. For each detected individual, it is recorded which observer saw the individual, when it was seen, and in what location. For conventional MRDS, the snapshot assumption is used and so, in theory, individuals can be uniquely identified by their static location. In practice, often, a third party receives detection records from both observers independently and matches detections that are believed to come from the same individual. Discrepancies in recorded location, due to error and animal movement, are common; often, locations are averaged or the location recorded by the first observer to sight the individual used in subsequent analyses.

In MDS, both locations and detection times from each observer can be used when both observers detect an individual. Furthermore, sightings made by only one observer provides information about where an individual was likely to move and thereby evade detection by the other observer. The PDF of the observed encounters is given by Theorem 2.1:

$$\mathcal{L}_{e \mid n} = \prod_{i=1}^{n} \prod_{j=1}^{2} \int \lambda_j(\vec{\boldsymbol{x}}_i, t_{i,j})^{\omega_{i,j}} S_j(\vec{\boldsymbol{x}}_i, t_{i,j})[\vec{\boldsymbol{x}}_i] \, \mathrm{d}\vec{\boldsymbol{x}}_i$$

where S_j is the survival function for observer j, $\omega_{i,j} = 1$ if observer j detected individual i (and zero otherwise), and $t_{i,j}$ is the time of the encounter between observer j and individual i if it occurred and is equal to the total time spent surveying the transect otherwise.

Similarly to Section 3.2, this conditional likelihood can be maximised to estimate the detection parameters for each observer and so estimate the probability of detection over the survey. Both a design-based and model-based approach to density estimation is possible.

Despite observers searching independently, it is common for there to be dependence between observers that is not explained by correlation induced by an individual's movement path (Buckland, Laake, & Borchers, 2010). Individuals can display conspicuous behaviour or traits that induces correlation in detections between observers. Assuming detections between observers are independent is termed a full independence double observer model (Buckland et al., 2010).

An alternative approach for MDS models similar to that presented by (Buckland et al., 2010) is to allow for dependence between the observers by modelling the discrepancy between the joint and marginal survival functions. In particular, let $S_{\cdot}(\boldsymbol{x}, t)$ be the probability an individual travels path $\boldsymbol{x}$ and survives detections up to time t. Define the dependence between the observers to be the function:

$$\Delta(\boldsymbol{x}, t) = \frac{S_{\cdot}(\boldsymbol{x}, t)}{S_1(\boldsymbol{x}, t) S_2(\boldsymbol{x}, t)}.$$

This is a measure of the distance between the joint and marginal distribution. Independent observers would require $\Delta = 1$, while positive dependence would imply $\Delta > 1$.

Let δ_{12} be a function such that

$$\Delta(\boldsymbol{x}, t) = \exp\left(-\int_0^t \delta_{12}(\boldsymbol{x}(s), s); \mathrm{d}s\right).$$

The function δ_{12} is defined similarly to a hazard function, but it is not a hazard since it can be negative when dependence between observers is positive. Despite this, the mathematical form is convenient. Let $S_{j_1 \mid j_2}$ be the probability of surviving detection from observer j_1 given survival from observer j_2. It follows that

$$\Delta(\boldsymbol{x}, t) = \frac{S_{1 \mid 2}(\boldsymbol{x}, t)}{S_1(\boldsymbol{x}, t)}.$$

Thus, if $\lambda_{1 \mid 2}$ is the associated hazard of $S_{1 \mid 2}$, we have that $\lambda_{1 \mid 2} = \lambda_1 + \delta_{12}$. Similarly, $\lambda_{2 \mid 1} = \lambda_2 + \delta_{12}$. In short, δ_{12} is an adjustment made to the hazard of one observer when it is known that the other observer has failed to detected the individual. To fit this model, all that is required is a functional form for δ_{12}.

3.9.2 Trial-Based

An alternative survey method is where one observer, the tracker, searches in advance of the other, the searcher. When the tracker sights an individual, they continue to track the movement of that individual and then record whether or not the searcher detected the individual. Thus, each individual detected by the tracker constitutes a trial.

The tracker can record the observed movement path of the individual while it is at risk of detection by the searcher. This means that the movement path of the individual is observed and no longer latent. Thus, path integration can be restricted to only those paths that intersect with the observed locations of the individual. This is easily achieved using MDS models.

An alternative method to set up trials is to survey a transect where one or more tagged individuals may move. The movement paths of tagged individuals is known and so again, this can be incorporated into MDS to estimate the chance of missing an individual at zero distance. Effectively, the tags act as multiple additional observers whose detection probability is certain.

3.10 Responsive Movement

This chapter focussed on movement that is not in response to the observer. Responsive movement is a common problem in distance sampling surveys. For point transects, it is somewhat mitigated when observers wait at a point until individuals resume non-responsive behaviour and then begin to survey. For line transects, searching effort is focussed ahead of the observer, along the line, so that individuals are sighted before they are motivated to respond to the observer's oncoming presence. MRDS, double-observer, methods can be used to account for responsive movement that occurs between surveying by the first and second observers. Thus, if the first observer searches in advance of the second, detections of responding individuals by the second observer can be accounted for (Borchers & Cox, 2017; Conn & Alisauskas, 2017). Response that occurs before either observer detects the individual is not accounted for.

MDS provides the statistical framework to incorporate responsive movement. In particular, individuals can move according to an advection-diffusion model where they are either attracted or repelled by the observer. If $\vec{\boldsymbol{x}}(t)$ is the position of an individual at time t and $\vec{\boldsymbol{z}}(t)$ is the position of the observer at time t, a response movement model could take the form:

$$\mathrm{d}\vec{\boldsymbol{x}}(t) = \alpha(\vec{\boldsymbol{z}}(t) - \vec{\boldsymbol{x}}(t))\ \mathrm{d}t + \sigma^2 \mathrm{d}\vec{\boldsymbol{w}}(t)$$

where σ^2 is the diffusion parameter, $\vec{\boldsymbol{w}}$ is a 2D Brownian motion process, and α is the response parameter. When $\alpha > 0$, individuals are attracted to the observer, when $\alpha < 0$, they are repelled and $\alpha = 0$ corresponds to non-responsive movement. This movement model can be used within MDS. There is little additional computational burden because the position of the observer is known and the same for all individuals, making the adjustment convenient.

As with non-responsive MDS, additional information on individual movement is required. Diffusion σ can be estimated from telemetry data as before. The parameter α may only be estimated from repeated recorded locations of individuals who are moving in the vicinity of the observer. This could be collected in two ways. In a double observer survey, individuals may be recorded by both observers, providing two recorded locations for those individuals. These locations can be used to estimate α. Alternatively, the observer can move through an area where there is one or more tagged individuals. The response can then be estimated from the tag locations and the known observer path.

3.11 Conclusion

This chapter shows that in distance sampling surveys of wild animal populations, the encounter is the fundamental unit that is sampled, not distances or static points. The statistical theory from Chapter 2 provides the necessary conceptual and computational developments to incorporate movement into the sophisticated detection models used in distance sampling. Doing so demonstrates three conclusions:

1. distance sampling inference can be substantially biased and incorporating movement can mitigate this. Further, it provides the methods required to investigate the joint relationship between covariates, density, and movement that has been neglected in density surface and point process modelling approaches. Thus, the work furthers the goal of ensuring the inference that influences population wildlife management is sound and robust;

2. MDS is a direct generalisation of distance sampling and inherits many of the major extensions of distance sampling such as availability modelling, MCDS, and MRDS. Furthermore, it embeds these extensions within the more abstract framework of encounter models.

3. MDS removes the snapshot assumption. Thus, MDS can be applied to populations, such as fast-moving birds, where such an assumption has prohibited the use of distance sampling. Furthermore, MDS provides the basis to apply distance sampling to data that is collected over very long time periods. The advent of affordable technology, such as cameras, automated rovers, and acoustic devices, means transects are surveyed over times where movement is undeniable. Methods to account for this so far require artificial snapshots to be made (Howe et al., 2017; Kyhn et al., 2012), requiring a subjective balance between wasting data and violating the snapshot assumption. MDS provides a fruitful alternative for future research.

Chapter 4

Continuous-time Spatial Capture-Recapture

4.1 Introduction

The ostensible objective of spatial capture-recapture (SCR) is to estimate the density of a wild animal population (Borchers & Efford, 2008; Efford, 2004; Royle, Chandler, Sollmann, & Gardner, 2013). It relies on the assumption that each member of the population has an identifying mark. Two or more detectors are then placed in the survey and record each encounter they have with each individual. The recaptures of individuals over time and space provides the necessary information to estimate the probability of detection and thus population density. It has become apparent, however, that the data collected in SCR surveys provide more information about the population than merely its density (for example, Royle et al. (2017); Sutherland et al. (2015)). In particular, the spatial component of the model provides inference on where animals are and how they use the space in which they reside. Increasingly, SCR is being used to make inference about the relationship animals have with their habitat, an objective normally pursued using animal telemetry data and models from the field of movement ecology.

Continuous-time spatial capture-recapture (Borchers et al., 2014) is an extension of conventional SCR where the exact detection times of animals are recorded. The methods of continuous-time SCR are the focus of this chapter and, henceforth, the acronym SCR will refer to continuous-time methods. Current SCR methods are discussed in Section 4.2 and shown to not account explicitly for animal movement. Subsequently, the theory of encounter models (Chapter 2) is shown to generalise SCR methods by incorporating an explicit an-

imal movement model (Section 4.3). The new method is applied to a case study of male jaguars (*Panthera onca*) in the Cockscomb Wildlife Sanctuary Basin in Belize (Section 4.4). Finally, the contributions made, limitations of, and potential extensions of the method are discussed (Section 4.5).

4.2 Continuous-time spatial capture-recapture

Consider a population with density $D(\boldsymbol{x})$ at location $\boldsymbol{x}$ over a survey area $\mathcal{A}$. A series of $J > 2$ detectors are placed in this survey area. Whenever a detector encounters an animal, the exact time of this detection and the unique identity of the animal is recorded. This information is the minimum required for continuous-time spatial capture-recapture modelling. This information is recorded whenever an animal is detected, and so each animal in the population has an associated encounter history (or capture history) that lists each time the animal was seen and by which detector. Only the encounter histories of those animals seen at least once are observed; the number of animals never seen is unknown and to be estimated.

To do this, a point in space, $\boldsymbol{x}_i$, is associated with each animal in the population, termed the animal's activity centre. These activity centres are spread across the survey area according to a point process model, the simplest being the Poisson process with rate D.

4.2.1 Detection

In any instant of time, the hazard of an animal being detected by detector j is determined by the distance between the animal's activity centre and the location of the detector. Animals whose centre of activity is further from the detector are less likely to be detected than those closer. This is described mathematically by the encounter rate function $\lambda(r, t)$ where r is the radial distance between the detector and activity centre and t is the time. In SCR, animals are assumed to spend their time around their activity centre according to a given distribution. The most common choice is a bivariate Gaussian distribution. A common approach to deriving an encounter rate function is to determine a period of time, τ, over which the Gaussian distribution describes the animal's range. The bivariate Gaussian distribution can then be used to describe the probability an individual would be detected at least once during this time: $g(r) = g_0 \exp\left(-\frac{r^2}{2s^2}\right)$ with two parameters s, the spatial range of the animal, and g_0 the probability of an animal whose activity centre coincides with the detector's location ($r = 0$) being seen in time τ. The encounter intensity function can then

be derived as $\lambda(r,t) = -\frac{1}{\tau}\log(1 - g_0\exp(-\frac{r^2}{2s^2}))$ (Borchers et al., 2015). Of course, a range of functional forms for this intensity could be chosen (Efford, 2017) and the parameters can depend on time, but the above is preferred due to its ease of interpretation. Intuitively animals roam around their activity centre and spend time in each location around this centre according to a Gaussian distribution and so the parameter s describes the animal's range. This interpretation is the first step toward making inference about animal movement.

Suppose n unique animals were seen at some time during the survey and that animal i has $n_{i,j}$ encounters with detector j at times $\boldsymbol{t}_{i,j} = (t_{i,j,1}, \ldots, t_{i,j,n_{i,j}})$. Let $S_t(r)$ be the probability an animal with an activity centre a distance r from a single detector eluding detection for a time interval of length t. The encounter rate is defined such that

$$\frac{\mathrm{d}S_t}{\mathrm{d}t} = -\lambda(r,t)S_t(r).$$

In other words, at every instant of time, $\lambda(r,t)$ of those animals with an activity centre r from a detector that have eluded detection up to that instant are detected and have no longer eluded detection. This implies that $S_t(r) = \exp\left(-\int_0^t \lambda(r,s)\,\mathrm{d}s\right)$.

The probability density function for the detection times $\boldsymbol{t}_{i,j}$ is given by

$$[\boldsymbol{t}_{i,j} \mid \boldsymbol{x}_i] = \left(\prod_{e=1}^{n_{i,j}} \lambda(r_j(\boldsymbol{x}_i), t_{i,j,e})\right) S_T(r_{i,j})$$

where $r_j(\boldsymbol{x})$ is the distance between location $\boldsymbol{x}$ and detector j, and T is the total time spent surveying.

Encounters with different detectors are assumed to be independent, thus the PDF of all encounters between animal i and every detector $\boldsymbol{t}_i = (\boldsymbol{t}_{i,1}, , \ldots, \boldsymbol{t}_{i,J})$ is given by:

$$[\boldsymbol{t}_i \mid \boldsymbol{x}_i] = \prod_{j=1}^{J} [\boldsymbol{t}_{i,j} \mid \boldsymbol{x}_i].$$

4.2.2 Likelihood

Animals are also assumed to be independent. Yet, only the encounter histories of those animals seen at least once are observed. Thus, the PDF of the observed data is conditional on being encountered at least once. The probability of an animal with activity centre $\boldsymbol{x}$ being encountered at least once is given by $p(\boldsymbol{x}) = 1 - \prod_{j=1}^{J} S_T(r_j(\boldsymbol{x}))$ (Borchers et al., 2015). Imagine that all activity centres are spread over the survey area and then

those activity centres belonging to animals that were never seen are removed. An activity centre at $\boldsymbol{x}$ is thus retained with probability $p(\boldsymbol{x})$. A key property of the Poisson point process with rate $D(\boldsymbol{x})$ is that when thinned in this way, the resultant point pattern is also a realisation from a Poisson process with rate $D(\boldsymbol{x})p(\boldsymbol{x})$ (Borchers & Efford, 2008). Thus, the number of detected animals n has a Poisson distribution with mean D_{enc} where $D_{enc} = \int D(\boldsymbol{x})p(\boldsymbol{x})\ \mathrm{d}\boldsymbol{x}$. This mean increases as detection probability increases. The PDF of activity centres for observed animals is thus

$$[\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n] = \frac{D_{enc}^n \exp(-D_{enc})}{n!} \prod_{i=1}^{n} D(\boldsymbol{x}_i)p(\boldsymbol{x}_i).$$

The full likelihood of the encounter rate parameters and the density parameter is given by (Borchers et al., 2014)

$$\begin{aligned}\mathcal{L} &= \frac{(D_{enc})^n \exp(-D_{enc})}{n!} \prod_{i=1}^{n} \int D(\boldsymbol{x}_i)p(\boldsymbol{x}_i)\frac{[\boldsymbol{t}_i \mid \boldsymbol{x}_i]}{p(\boldsymbol{x}_i)}\ \mathrm{d}\boldsymbol{x}_i \\ &= \frac{(D_{enc})^n \exp(-D_{enc})}{n!} \prod_{i=1}^{n} \int D(\boldsymbol{x}_i)[\boldsymbol{t}_i \mid \boldsymbol{x}_i]\ \mathrm{d}\boldsymbol{x}_i\end{aligned}$$

where division by $p(\boldsymbol{x}_i)$ is required as one must condition on the probability of the animal being seen at some time during the survey.

4.2.3 Estimation

This likelihood can be computed using numerical integration over all possible activity centres for each animal. The likelihood reduces to a product of one-dimensional integrals. Maximum likelihood methods can then be used to produce point and interval estimates of the parameters. Density can be assumed to be constant across space $D(\boldsymbol{x}) = D$ for all $\boldsymbol{x}$ or can be described by a spatial smooth such as a cubic spline. The primary inference from the fitted model is the population density.

An alternative approach is Bayesian estimation. Rather than numerical integration over all activity centres and rather than base inference on the likelihood conditional on an animal being seen, Royle and Young (2008) adopt a data augmentation approach. A Markov Chain Monte Carlo (MCMC) algorithm is used where in each iteration, the activity centre of every animal in the population, including those never seen, is simulated and so the complete-data

likelihood can be computed:

$$\prod_{i=1}^{N}[\boldsymbol{t}_i, \boldsymbol{x}_i] = [\boldsymbol{t}_i \mid \boldsymbol{x}_i][\boldsymbol{x}_i]$$

where N is the number of animals within the study region. A simulant of N is generated within each iteration of the MCMC algorithm and so the complete likelihood is calculable. To do this, an augmented population of size M is created where $M - n$ of the individuals have empty encounter histories. The probability of an individual in this super-population being included in the realised population of size N is then estimated as a parameter. The size of the super-population M is chosen by the analyst and must exceed any reasonable true population size. This approach is slower than numerical integration, but is more flexible when modelling is more complex, for example, when activity centres move over time.

4.2.4 Model selection and fit

For maximum likelihood methods, model selection is performed using Akaike information criterion (AIC) or its finite sample size correction (AICc). Model selection with the Bayesian approach has not been researched, but a reversible-jump MCMC or posterior predictive measure could be used to choose between differing encounter intensity or spatial models.

Goodness-of-fit for SCR models is still under development and there is no standard approach. Monte Carlo tests can be performed in a fashion similar to the posterior predictive checking described by Gelman et al. (2014). New data are simulated from the fitted model and a test statistic computed from these data. The observed test statistic's value for the original data is then compared to this empirical distribution and a simple hypothesis test performed. The test statistic can reflect a single aspect of the data, for example, how many encounters with animals occur on average per day, or can be based on the parameter estimates obtained from each simulated data set, for example, by computing the deviance of the model.

4.2.5 Covariates

The density and encounter rate parameters can be influenced by spatial, temporal, or individual-level covariates. Spatial and temporal covariates can be included using a link-function formulation.

Relating density to spatio-temporal covariates can provide inference on how density varies in different animal habitats and how animal populations are changing over time with respect to observed processes. For encounter rate parameters, including spatio-temporal covariates

allows detection probability to vary according to environmental conditions. Acknowledging temporally varying encounter rates is the primary use of continuous-time SCR, since when encounter rates and density are constant, no information is gained from using exact detection times compared to the counts of how many times each animal is recorded by each detector (Borchers et al., 2014).

Further to that, however, the encounter rate provides information on how animals use space and so relating this to spatio-temporal covariates provides a picture of habitat use and habitat connectivity. This inference is crucial to wildlife and environmental management and the potential of SCR to provide this inference on a population-level scale has led to greater interest in this line of inquiry.

4.2.6 Movement

At the most rudimentary level, the SCR model can be used to infer where each observed animal's activity centre is likely to be (Borchers & Efford, 2008). Bayes Theorem can be used:

$$[\boldsymbol{x}_i \mid \boldsymbol{t}_i] = \frac{[\boldsymbol{t}_i \mid \boldsymbol{x}_i][\boldsymbol{x}_i]}{[\boldsymbol{t}_i]}.$$

Further to this, the encounter rate parameters have been used to make inference on animal movement. As already noted, the range parameter, s, provides information about animal activity range. This is also termed "home range" in many contexts. Yet, home range has many meanings within the ecological literature and is prone to mis-interpretation. In this chapter, activity range will be used. This can easily be used to estimate activity area, the area within which an individual spends 95% of its time. This estimate is derived from the assumption that each individual's space use is described by a bivariate Gaussian. This can further be extended, by allowing dependence on covariates, to allow activity range to change with spatial or temporal processes. This gives inference similar to resource selection functions where animal space use is related to what resources are available within their range.

Despite this, inference about animal movement from SCR models can be biased. SCR estimation of population density is robust to violations of assumptions; however, encounter rate and movement parameter estimation is not (Efford, 2014; Efford, Borchers, & Byrom, 2009). Commonly, for small sample sizes, encounter rate and movement parameter estimators are negatively correlated. The underestimation of one is counterbalanced by the overestimation of the other; this leads to unbiased density estimation. Thus, when used for its primary purpose, SCR is effective and reliable. Yet, if interest shifts to inference

about animal movement, one must be cautious. In many cases, such inference is simply not warranted as the bias in movement parameter estimation can be severe. In effect, the disentanglement of the detection and movement processes is poorly achieved when re-captures are low.

This shortcoming can be mitigated by incorporating telemetry data. Even a small sample size of telemetry can be effective when separating the detection and movement inferences. Given telemetry records $\boldsymbol{r}_i$ for individuals $i = 1, \ldots, m$ where m_i records are made for individual i: $\boldsymbol{r}_i = (\boldsymbol{r}_{i,1}, \ldots, \boldsymbol{r}_{i,m_i})$; the activity range can be estimated as given by Calhoun and Casby (1958):

$$\hat{s} = \frac{\sum_{i=1}^{m} \sum_{j=1}^{m_i} \|\boldsymbol{r}_{i,j} - \bar{\boldsymbol{r}}_i\|}{\sum_{i=1}^{m} m_i - m - 1}.$$

where $\bar{\boldsymbol{r}}_i$ is the mean location of individual i over its period of telemetry tracking. This can also be estimated by maximum likelihood when assuming the locations of each individual are an independent realisation from a bivariate Gaussian. A joint likelihood with both the telemetry and SCR data can be maximised to share information about s between the two (Bird, Lyon, Nicol, McCarthy, & Barker, 2014; Linden et al., 2017). This prevents the substantial bias in s. This approach relies on the assumption that the tracked animals move similarly to the animals in the SCR survey.

Note that in all current approaches to integrating telemetry and SCR data, locations of an individual recorded over time are treated as an independent sample from a distribution. This is in agreement with the assumption in SCR that detections are independent. Animals are idealised as points that teleport around their activity areas without moving through the intervening space. This means correlations in detections are not accounted for. Furthermore, it discards the information encounters give about where animals are likely to be in between their sightings. Replacing the implicit movement model within SCR with an explicit one removes these two limitations. That is the focus of this chapter.

4.3 Continuous SCR with Movement

The statistical theory of spatial capture-recapture is entirely formulated around the concept of an activity centre. Around this centre, an animal pops up randomly from one position to the next and encounters detectors at a rate proportional to the distance separating each detector and the centre. In reality, animals move in continuous paths across space. They encounter detectors according to the distance between their current location and the location of the detector. In short, continuous-time spatial capture-recapture is an encounter

process as described in Chapter 2. In this section, spatial capture-recapture is shown to be a type of encounter survey and the appropriate encounter model is presented.

4.3.1 Encounter Model

An encounter model is built from two parts: a detection model and a movement model.

The detection model is similar to that used in Section 4.2.1 with the difference that encounter rate varies with an animal's location on its continuous path, rather than being a function of a single point in space, the activity centre. Let animal i travel a path $\vec{\boldsymbol{x}}_i$ during the survey such that $\vec{\boldsymbol{x}}_i(t)$ is the location of animal i at time t. Precisely as before, the encounter rate $\lambda(r,t)$ has a chosen functional form, such as a half-Gaussian function. The difference is that now the PDF of encounter times $\boldsymbol{t}_{i,j}$ between animal i and detector j depends on the entire contiguous path the animal travelled:

$$[\boldsymbol{t}_{i,j} \mid \vec{\boldsymbol{x}}_i] = S_T(\vec{\boldsymbol{x}}_i) \prod_{e=1}^{n_{i,j}} \lambda(r_j(\vec{\boldsymbol{x}}_i(t)), t_{i,j,e})$$

where

$$S_t(\vec{\boldsymbol{x}}_i) = \exp\left(- \int_0^t \lambda(r_j(\vec{\boldsymbol{x}}_i(s)), s) \, \mathrm{d}s \right).$$

The likelihood of all detection times for an individual $[\boldsymbol{t}_i \mid \vec{\boldsymbol{x}}_i]$ is then derived as above as encounters with different detectors are assumed to be conditionally independent, conditional on the path the animal travelled. This is a weaker assumption than made in conventional SCR as it allows for the reality that encounters with nearby detectors are correlated due to the correlation in the animal's movement.

The only difference between the encounter survey discussed in Chapter 2 and a spatial capture-recapture survey is that only the encounter histories of those animals that were seen at least once is observed. To account for this, as discussed in Section 4.2.2, the PDF of the observed data is conditional on being observed at least once.

The key difference between the current method and an encounter model approach to continuous-time spatial capture-recapture surveys is that encounter modelling includes an explicit model for animal movement. As discussed more fully in Chapter 2, once a movement model is selected, this provides a probability measure with respect to all the paths an animal could have traversed during the survey. Let Φ be this probability measure. Thus, if $p(\vec{\boldsymbol{x}}) = 1 - S_T(\vec{\boldsymbol{x}})$ is the probability of being detected at some point during the survey for an animal that travels the path $\vec{\boldsymbol{x}}$, then the probability of being detected at some point

during the survey unconditional on path travelled is given by

$$p_{enc} = \int p(\vec{\boldsymbol{x}})\ \mathrm{d}\Phi(\vec{\boldsymbol{x}}).$$

It is the complement to the probability of eluding detection, averaged over every possible path an animal could have travelled, weighting each by the likelihood of that path occurring according to the chosen movement model. Such an integral is termed a path integral and Section 2.5 describes how these integrals can be approximated numerically for encounter models.

Specifying the encounter rate function and the movement model is sufficient to construct an encounter model. For a spatial capture-recapture survey, however, the total population size is unknown and must also be estimated. In conventional SCR, activity centres are static and the model for population density is a point process model. When animals are permitted to move, it is unclear what such a point process would correspond to. Here, this process is assumed to describe the initial position of each animal at the beginning of the survey; the point from which their path begins. Hence, the point pattern of each individual's initial position is described by an inhomogeneous Poisson process, as above, with rate $D(\boldsymbol{x})$.

The likelihood of the parameters for the detection, movement, and density processes, look very similar to the likelihood for the conventional SCR model despite the difference in the approach:

$$\mathcal{L} = \frac{D_{enc}^n \exp(-D_{enc})}{n!} \prod_{i=1}^{n} \int D(\vec{\boldsymbol{x}}_i(0))[\boldsymbol{t}_i \mid \vec{\boldsymbol{x}}_i]\ \mathrm{d}\Phi(\vec{\boldsymbol{x}})$$

where $D_{enc} = \int D(\vec{\boldsymbol{x}}(0))p(\vec{\boldsymbol{x}})\ \mathrm{d}\Phi(\vec{\boldsymbol{x}})$.

4.3.2 Movement

The full class of advection-diffusion movement models is described in Section 2.3. The simplest movement model is a diffusion model where animals move according to Brownian motion with a mean speed parameter ν. The path integration for this model can be computed numerically and the encounter model fit by maximum likelihood. Such a model, however, is not a direct generalisation of the conventional spatial capture-recapture. Animals were assumed to roam around an activity centre within an activity range. Under a diffusion model, animals are not constrained to remain within any range and given enough time will explore the entire survey region. For highly migratory species, this transitory behaviour may be a good description; however, for many species surveyed by spatial

capture-recapture, individuals tend to remain in an activity range that is bounded either due to competitive terroritory, prey availability, or habitat suitability. In these cases, the diffusive model will not capture the affinity of individuals to their activity areas which may lead to biased inference on encounter rate and spatio-temporal distribution.

An Ornstein-Uhlenbeck (OU) movement model (Uhlenbeck & Ornstein, 1930) requires each animal to have a single point in space to which they are attracted, an activity centre. For conventional SCR, animals distribute their time around this centre according to a Gaussian distribution. Given enough time, under an OU process, animals would do the same. In other words, the OU process describes the implicit movement model assumed within conventional SCR. The standard SCR approximation, rather than accounting for every continuous path around the activity centre, uses the average proportion of time an animal spends in each location around its activity centre.

The equation describing OU movement is a stochastic differential equation:

$$\mathrm{d}\vec{\boldsymbol{x}}(t) = \alpha(\boldsymbol{\mu} - \vec{\boldsymbol{x}}(t))\,\mathrm{d}t + \sigma\,\mathrm{d}\vec{\boldsymbol{w}}(t)$$

where $\vec{\boldsymbol{x}}(t)$ is the animal's location at time t, $\boldsymbol{\mu}$ is the animal's activity centre, $\alpha > 0$ is a parameter quantifying the animal's attraction to its activity centre, σ is the diffusion parameter, and $\vec{\boldsymbol{w}}$ is the two dimensional standard Wiener process.

4.3.3 Estimation

In Chapter 2, computational methods were developed to compute the path integral for movement models where an animal's activity centre was known. In particular, the numerical methods can be used to compute $[\boldsymbol{t}_i \mid \boldsymbol{x}_i, \boldsymbol{\mu}_i]$ where $\boldsymbol{\mu}_i$ is the activity centre of animal i. As above, activity centres can be described as a Poisson point process and numerical integration be used to average over all possible activity centres for each animal. In practice, this double-integration approach (first, integrating over all paths, and then integrating over all activity centres) is computationally intensive and is, at present, infeasible.

An alternative is a Bayesian approach, using Markov chain Monte Carlo (MCMC) to integrate over all activity centres. To do this, data augmentation is used. A super-population of M animals is simulated where M is assumed to be larger than the unknown size of the study population. Prior distributions are specified for each encounter rate parameter and diffusion parameter. Further, for each animal i in the super-population, a prior is specified for $\boldsymbol{\mu}_i$, the animal's activity centre. Thus, within an MCMC algorithm, the observed data are augmented with the activity centre of every animal. Hence, within each MCMC iter-

ation, the numerical methods at hand can be used to compute a marginal likelihood: the likelihood given the observed data and activity centres, marginalised over all possible paths each animal in the population could have travelled around their proposed activity centre. Activity centres and other parameters can then be sampled from the posterior using an MCMC algorithm such as a Metropolis-Hastings (Hastings, 1970) or Hamiltonian Monte Carlo (Hoffman & Gelman, 2014). This approach is less computationally intensive than double-integration and uses the combined strength of data augmentation and numerical integration.

4.3.4 Model selection and fit

As with conventional SCR, when fit by maximum likelihood, models can be compared by AIC. For a Bayesian approach, the deviance information criterion, Bayes factors, or a reversible-jump MCMC algorithm could be used to select between different movement or detection models. This is not routinely done in the literature due to the computational burden and is not explored here.

Goodness-of-fit of these models can be assessed in the same way as described for existing continuous-time SCR models. Further to this, assessing the goodness-of-fit of encounter models is discussed in Section 2.7. In particular, the concept of an encounter residual is introduced. Encounter residuals can be computed for continuous-time SCR surveys to assess goodness of fit and identify failings in the models used. The deviation of the encounter residuals points toward the model's failing. Positively-skewed distributions imply that encounters are more clustered in time than to be expected, suggesting something is lacking in the movement model. A common reason for this may be the phenomenon known as "trap-happiness" where animals seek out detectors. The reverse, where animals avoid detectors ("trap-shyness") would lead to negatively-skewed residuals since the duration between encounters would be unusually long compared to what one would expect given the fitted model. Outlying residuals for a single encounter or for many encounters with a single animal can also inspire further investigation and may lead to adopting a different model or to grounds for the removal of a data point or entire individual from the analysis.

4.3.5 Covariates

As before, parameters and covariates can be related using a link function formulation. Section 4.2.5 discussed how encounter rate parameters can vary with space and time. The difference is that when an explicit model for animal movement is incorporated, the inter-

pretation of encounter rate parameters changes, in particular, the detection range has no relation to the animal's movement, only to the range at which a detector can record an encounter. In short, the encounter rate parameters describe only the efficacy of a detector to detect and so covariates that influence a detector's ability to record an animal ought to be considered, for example, weather, lighting, or technological capability (some detectors may use older technology compared to others).

Environmental and temporal covariates can also be related to movement processes through the diffusion and advective movement parameters. Diffusion can change depending on space or time, allowing the model to identify areas where animals move more quickly, for example in movement corridors, or identify times where animals are more active, day compared to night. For the Ornstein-Uhlenbeck model, the attraction of each animal to its activity centre can vary according to time and space; for example, an animal may be less tightly bound to its activity centre during breeding periods, when searching for a mate. Similarly, the diffusion component of the OU process can change over space; this means that animals not only diffuse around their activity centre, but can do so heterogeneously, depending on the surrounding environment. In particular, animals may have preferred routes or habitats through which they are more likely to travel. These questions of home ranging and landscape connectivity are a central focus of movement ecology; this framework allows these questions to be addressed using continuous-time spatial capture-recapture data.

4.3.6 Movement Inference

Incorporating an explicit animal movement model with continuous-time spatial capture-recapture allows for inferences to be made on the spatio-temporal distribution of the individuals sighted during the survey. In this section, three related forms of inference are considered: activity range estimation, spatio-temporal distribution, and animal-animal interactions.

Activity range

For an individual i that moves according to a OU process, the distribution of locations over an infinite time, termed here the stationary distribution, is bi-variate Gaussian with mean $\boldsymbol{\mu}_i$, the animal's activity centre, and variance $\frac{\sigma^2}{2\alpha}$ (Gillespie, 1996). This implies that the estimated activity range, denoted s, is this standard deviation:

$$\hat{s} = \frac{\hat{\sigma}}{\sqrt{2\hat{\alpha}}}.$$

This quantity has the same interpretation as in conventional SCR, hence the use of the same symbol, s.

This assumes that the parameters do not vary with space or time. If either σ or α vary with covariates, the activity range of each individual will be irregular in shape and change according to the covariate values around each individual's activity centre. The stationary density in location $\boldsymbol{x}$ at time t for individual i is proportional to a bivariate Gaussian with mean $\boldsymbol{s}_i$ and variance $\frac{\sigma(\boldsymbol{x},t)^2}{2\alpha(\boldsymbol{x},t)}$. Each activity range could be summarised by the radii of the maximal inscribing and minimal circumscribing circles. Alternatively, the mean activity range can be estimated at each direction from the activity centre, providing a picture of ranging in each direction. Finally, averaging over all directions would summarise activity range in a single number as in the case with no covariates.

Given an estimate of activity range, $\hat{s}$, the stationary activity area can be estimated as the area within which an animal spends 95% of its time in the long-run. This is one possible definition of home-range. Nevertheless, it is termed the stationary activity area here for clarity. Home-range estimation was arguably the first aim of movement ecology. It summarises information on animal movement and how animals relate to their habitat. SCR models that allow for the activity range of each animal to depend on covariates exist (Sutherland et al., 2015) but do not directly incorporate movement.

Spatio-temporal distribution

Activity range estimation does not fully describe the temporal distribution of an individual over the survey time. Encounters give information on where an individual spends their time. In particular, the posterior density of each individual's location can be used to present meaningful inference on animal movement and distribution in space and time. The posterior distribution, given all encounters that occurred in the survey $\boldsymbol{r}$, of an individual at time t, denoted $\vec{\boldsymbol{x}}(t)$, is found by marginalisation over the posterior of the parameters, $\boldsymbol{\theta}$, and activity centre, $\boldsymbol{\mu}$:

$$[\vec{\boldsymbol{x}}(t) \mid \boldsymbol{r}] = \int [\vec{\boldsymbol{x}}(t) \mid \boldsymbol{r}, \boldsymbol{\theta}, \boldsymbol{\mu}][\boldsymbol{\theta}, \boldsymbol{\mu} \mid \boldsymbol{r}] \, \mathrm{d}\boldsymbol{\theta}\mathrm{d}\boldsymbol{\mu}.$$

The predictive distribution $[\vec{\boldsymbol{x}}(t) \mid \boldsymbol{r}, \boldsymbol{\theta}, \boldsymbol{\mu}]$ is computed by Bayes Theorem and path integration:

$$[\vec{\boldsymbol{x}}(t) \mid \boldsymbol{r}, \boldsymbol{\theta}, \boldsymbol{\mu}] = \int \frac{[\boldsymbol{r} \mid \vec{\boldsymbol{x}}(t), \boldsymbol{\theta}, \boldsymbol{\mu}][\vec{\boldsymbol{x}}(t) \mid \vec{\boldsymbol{x}}, \boldsymbol{\theta}, \boldsymbol{\mu}]}{[\boldsymbol{r} \mid \boldsymbol{\theta}, \boldsymbol{\mu}]} \, \mathrm{d}\boldsymbol{\Phi}(\vec{\boldsymbol{x}}).$$

This integration can be approximated by the Viterbi algorithm (Zucchini et al., 2016) using

the hidden Markov model approximation in Chapter 2.

Thus, the spatio-temporal distribution of each individual can be inferred from its encounters with detectors. Clearly, individuals that are encountered most will provide more insightful inference. This distribution could be graphed over time to show how the location of each individual varies over time and space. Useful quantities that can be derived from the posterior predictive distribution are ranging behaviour and a time-spent distribution.

Home-range estimation typically quantifies an individual's use of space over a single period. Given the posterior predictive distribution of animal location, the mean distance travelled by an individual over different temporal scales can be estimated. Furthermore, even though covariates may not be explicitly included in the movement model, the relationship of predicted location with temporal or spatial covariates can be investigated to discover what can explain animal movement.

A further quantity of interest is the time-spent distribution. This was introduced by Pedersen et al. (2011), termed there as the residency distribution. This is an estimate of the average time spent during the survey at each location $\boldsymbol{x}$, denoted $\mathcal{T}_i(\boldsymbol{x})$ for individual i. The time-spent distribution is given by

$$\mathcal{T}_i(\boldsymbol{x}) = \int_0^T [\vec{\boldsymbol{x}}_i(t) = \boldsymbol{x} \mid \boldsymbol{r}] \, \mathrm{d}t$$

where T is the duration of the SCR survey. Notice that $\int \mathcal{T}_i(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} = T$, as required. Again, this distribution can be compared with spatial and temporal covariates to investigate what determines where animals go and how long they spend there. An alternative definition of home-range, here termed the realised activity range, is the 95% level set of $\mathcal{T}$. This will be less than the stationary activity range as it is a measure of the animal's range over the survey rather than over an infinite amount of time.

Animal-Animal interaction

Inference on group animal movement (e.g. Hooten, Scharf, Hefley, Pearse, and Weegman (2018); Langrock et al. (2014); Russell, Hanks, and Haran (2016)) commonly centres on the individual and focusses on modelling relationships between tracked or tagged individuals, ignoring others. SCR surveys are designed to capture and re-capture a proportion of the study population. In the previous sections, it has been discussed how this can be used as a form of telemetry information to infer individual movement. An additional exciting opportunity is to compare the movement of animals and to investigate animal-animal

interactions.

The simplest description of animal-animal interaction would be the area of the spatial overlap in the estimated stationary or realised activity ranges. This describes the overlap of space use and territory demarcation.

For many investigations, spatial overlap is the only inference to investigate territory; however, it ignores any possible temporal separation of individuals. Individuals may use the same location the same proportion of time, but do so at different times. The posterior predictive distribution of the path of each individual can be used to investigate this. One way to summarise this would be to estimate a "hard-core" distance that separates each pair of individuals, intuitively quantifying their personal space. Here, the separation distance for a pair of individuals is defined as the minimum distance such that the pair of individuals are at least that far apart 95% of the time.

One could go further and compute the mean time two individuals i, j spent a distance r or less from each other during the survey, denoted $\mathcal{S}_{i,j}(r)$:

$$\mathcal{S}_{i,j}(r) = \int_0^T \int\int_{a(\boldsymbol{x},r)} [\vec{\boldsymbol{x}}_i(t) = \boldsymbol{x} \mid \boldsymbol{r}][\vec{\boldsymbol{x}}_j(t) = \boldsymbol{y} \mid \boldsymbol{r}] \, \mathrm{d}\boldsymbol{y} \, \mathrm{d}\boldsymbol{x} \, \mathrm{d}\boldsymbol{t} \tag{4.1}$$

where $a(\boldsymbol{x}, r)$ is the set of all locations at distance r or less from location $\boldsymbol{x}$. This separation function depicts the estimated interaction of two individuals. It is similar to Ripley's K-function, used to quantify the correlation in point process patterns. If two individuals have home-ranges that overlap in space, it is possible that their separation function can show that they maintain a separation in time; conversely, it may show that they do share space contemporaneously. Either way, this provides important inference allowing conclusions to be made not only about habitat use but also on social interaction within the study population.

4.4 Case Study: Jaguars

Jaguars (*Panthera onca*) are classified by the IUCN as a "near threatened species" (Quigley & Harmsen, 2017). Furthermore the IUCN highlight the lack of knowledge there is on jaguar populations meaning that the threats they face may be underestimated. In particular, there has been some evidence that jaguar habitat is being lost due to deforestation; thus inference on individual range in particular is of critical importance. Jaguar populations tend to have low density and each jaguar ranges widely; both effects make population assessment more difficult. A better understanding of jaguar populations is a necessary first step toward improved conservation. In this section, three key properties of a jaguar population are

considered: density, individual activity ranges, and social interaction. Camera trapping and telemetry are used together to provide statistical inference on each of these aspects. To do this, both movement and detection processes must be considered and, thus, the encounter methods above are used.

4.4.1 Survey

This case study focuses on a male jaguar population in the Cockscomb Basin Wildlife Sanctuary, Belize. The sanctuary contains approximately 500 square kilometres of tropical rainforest with the Maya mountains to the west and the Caribbean sea to the east. The sanctuary consists mostly of a flat basin for two major rivers, as well as many smaller rivers. Furthermore, the sanctuary includes an established trail system (approximately 65 kilometres in length) that is used extensively by male jaguars. For further details of the survey area, see Harmsen, Foster, Silver, Ostro, and Doncaster (2010).

Twenty-one cameras were placed along trails within the sanctuary, separated on average by two kilometres (Figure 4.1). Cameras take photographs of each passing jaguar with an eight second delay between pictures. From these pictures, jaguars can be uniquely identified. Furthermore, the exact time of each encounter was recorded by the camera. Thus, the camera traps produced a continuous-time spatial capture-recapture data set. Encounters during the dry season April–July 2013 were used in the analysis. Only encounters with male jaguars are considered here. The movement of male and female jaguars is known to differ substantially (Harmsen et al., 2010), and, as a result, female jaguars are detected far fewer times. The lack of information on the female population is a current research problem in the assessment of the jaguar populations.

Additionally to the camera traps, telemetry data on two male jaguars (spanning approximately one year each) within the sanctuary was also collected in 2015–2016 with radio-collared tags.

The aim of this study is to analyse jointly the camera trap and telemetry data; making inference on both the population and individual level. The camera trap survey provides information on jaguar density and, seen as a thinned form of telemetry, on the activity range of each captured individual; the telemetry data provides more detailed information on the ranging and movement of two individuals. Encounter model methods can bring the strengths of both data sets together.

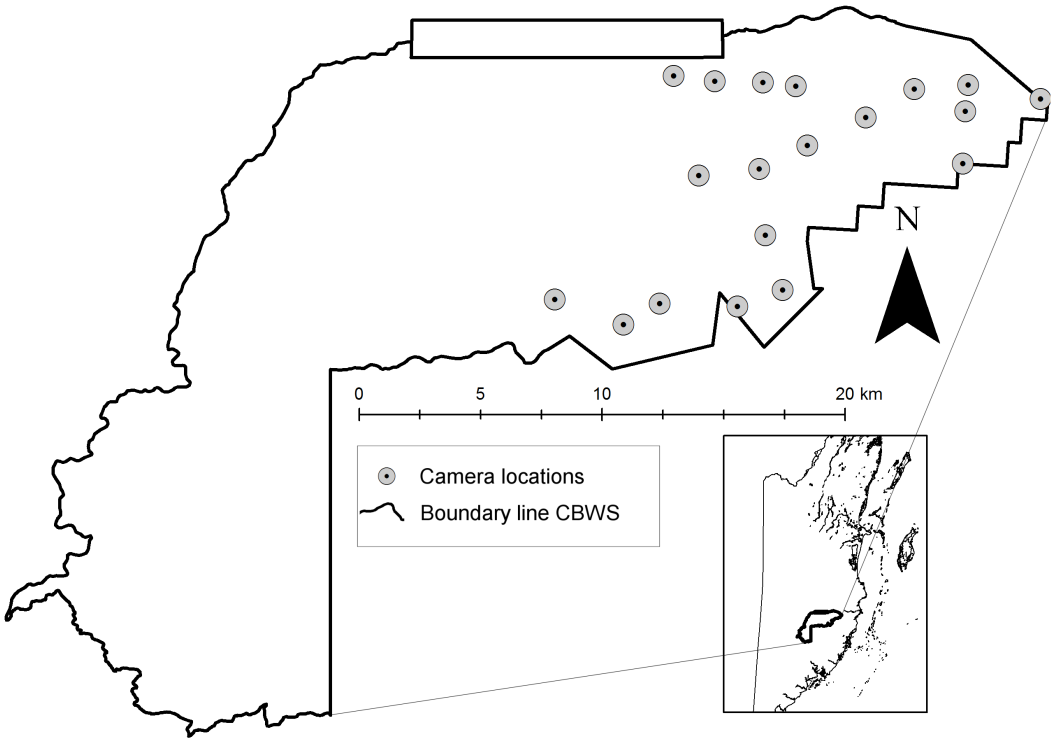


Figure 4.1 *Diagram of Cockscomb Wildlife Sanctuary Basin, Belize with cameras (represented by points) placed along an existing trail network*

4.4.2 Methods

Both the continuous-time SCR (CTSCR) and encounter models (ENC) are considered, as described above. The telemetry data are included in both models. For both models, a half-normal hazard function is used.

For the encounter model, jaguars are assumed to move according to an Ornstein-Uhlenbeck process with the addition that jaguars can diffuse at a different rate along rivers or trails. Jaguar movement was constrained to be within 25 kilometres of the park boundary and to be within the coastline. Therefore, there is the assumption that jaguars detected in the survey do not move outside of this range; it is, however, possible that they may as suitable habitat exists to the west.

This spatial covariate was included as jaguar movement is known to be biased along rivers and trails. Thus, the movement process has three parameters: α, the strength of attraction of each jaguar to his activity centre, σ_0, diffusion rate outside trails and rivers, and σ_1, diffusion rate on trails and rivers. The activity centre of each jaguar is also estimated due to the use of data augmentation.

The augmented population has size 100 individuals and a grid size of 500 metres is used in the path integration. Allowing for jaguar movement with an irregular boundary is easily accommodated with encounter models; the HMM approximation can account for boundaries by inhibiting transitions to grid cells outside the irregular survey area.

A Metropolis-Hastings random walk MCMC algorithm was used to approximate the posterior of all parameters and activity centres. The algorithm was based on the likelihood marginalised over all individual paths using HMM quadrature. Comparing independent chains, convergence appeared to occur after 1000 iterations by diagnostic trace plot inspection and the Gelman-Rubin convergence statistic (Gelman et al., 2014). A conservative burn-in of 5000 iterations was used. The posterior is approximated by a further 20000 iterations.

4.4.3 Abundance

Over the survey, 14 unique individuals were encountered. Each individual was detected 13 times on average. Estimated abundance was similar for CTSCR and the encounter model (Table 4.1). The daily encounter rate for an individual with an activity centre that coincides with a camera location was higher for the encounter model: 0.047 (SE = 0.012) for CT-SCR and 0.065 (SD = 0.012) for the encounter model. This can reflect that individuals

do not explore their entire activity area and so being detected in one place can reduce the probability of being detected at detectors on the other side of their activity range. CTSCR models cannot account for this through a movement model and so do so by estimation of lower encounter rate overall.

	Abundance	SD	LCL	UCL
CT	37.2	10.0	22.0	63.1
ENC	38.4	9.4	22.0	59.0

Table 4.1 *Maximum likelihood estimates of mean abundance (number of individuals) with standard error (SD), lower and upper* 95% *confidence intervals (LCL, UCL) for continuous-time spatial capture-recapture (CT) compared with posterior means, posterior standard deviations (SD), and* 95% *credible intervals for abundance from encounter model (ENC) with Ornstein-Uhlenbeck animal movement*

Goodness of fit of the two models was assessed by computing the encounter residuals from the detection times. Residuals from the CTSCR appear slightly positively skewed, indicating that encounters were clustered in time (Figure 4.2). This skew was not present in the residuals for the encounter model. Nevertheless, a Kolomogorov-Smirnov test found no evidence of poor fit in either model's residuals and the quantile-quantile plots for the two models are similar. This was somewhat to be expected for the jaguar survey as the correlation between detections has been found to be low in previous analyses (Borchers et al., 2014) and thus bias from correlation to be less of a problem.

Nevertheless, the residuals suggest a marginal deficiency in the CTSCR model that is removed by considering continuous animal movement paths. In this case, this deficiency has minimal effect on density estimation, but this cannot be known *a priori*. Simulation studies by Borchers et al. (2014) have shown that the bias from CTSCR depends on the detector spacing and the encounter intensity: higher intensities and closer detectors lead to more correlated encounters, causing bias. The encounter model provides the alternative methodology when these issues arise.

4.4.4 Movement

Incorporating an explicit movement model into SCR brings with it the ability to make inference on animal movement. Here, telemetry data on two male jaguars were incorporated. Thus, movement parameters are inferred from detailed information from the telemetry data and adjusted for the heterogeneity of encountered individuals. A common issue in movement ecology is the step where inference made on a small sample size of individuals is used to make conclusions about a population as a whole (Hooten, Buderman, Brost, Hanks, & Ivan,

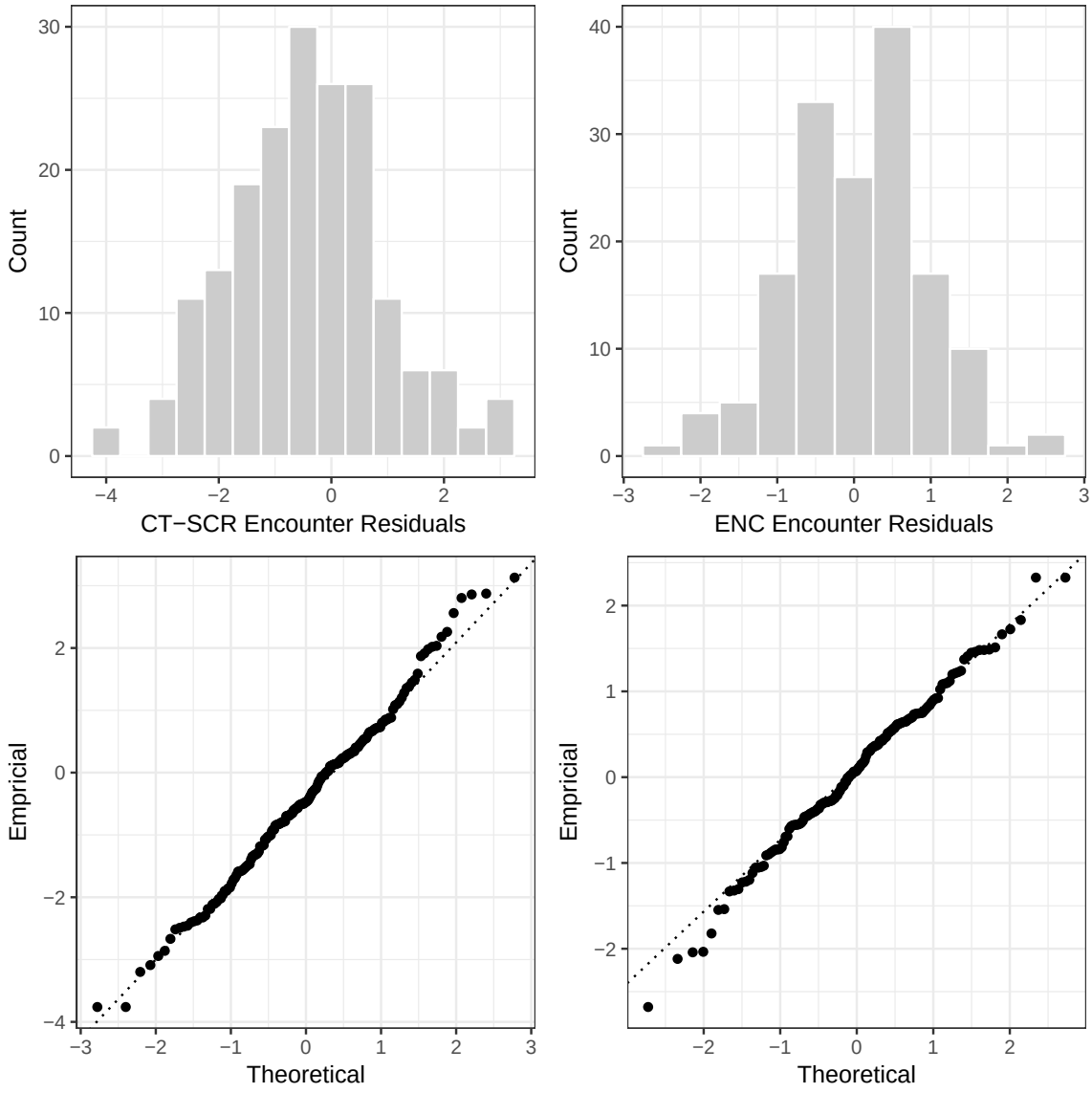


Figure 4.2 *Histogram (top) of encounter residuals with QQ-plot (bottom) comparing residuals to theoretical normal distribution for continuous-time SCR (left) and encounter model with movement (right)*

2016). Here, the camera trapping survey data can be used to extend inference to a larger proportion of the population.

	Estimate	SD	LCL	UCL
σ_0	0.058	0.002	0.055	0.061
σ_1	0.068	0.001	0.065	0.071
ENC H	175.1	11.7	154.6	199.4
CTSCR H	174.2	15.4	146.1	205.7

Table 4.2 *Posterior mean estimates standard deviations (SD) and lower (LCL), upper (UCL)* 95% *credible interval bounds for stationary activity area, H, in square kilometres from continuous-time SCR (CT-SCR) and encounter model (ENC) with Ornstein-Uhlenbeck movement model with posterior mean diffusion parameter on, σ_1, and off, σ_0, trails.*

Jaguars were estimated to diffuse significantly further along trails compared to when moving off-trail (Table 4.2). Estimated stationary activity area, the smallest area within which the animal spends at least 95% of its time, was similar between the methods. Estimated activity area size from a fitted Ornstein-Uhlenbeck model (using the `ctmm` package, see Calabrese, Fleming, and Gurarie (2016)) for the two tagged jaguars was 186 and 153 square kilometres, respectively. Goodness-of-fit was assessed by computing pseudo-residuals from the preditive distribution of each movement conditional on the individual's current location. Appendix C, Section C3 shows the quantile-quantile plots of the residuals for the two tagged individuals. The movement model appears to fit well for the first individual, but fails to capture movements of the second individual in the extremes. This may be due to difference in their activity patterns: the first individual's recorded locations are clustered around a single point, while, for the second individual, there is an indication of multiple clusters, meaning the individual may have multiple centres of attraction. An important extension for future analysis would be to consider movement models that allow for multiple activity centres; this may be especially important for surveys that are conducted over longer time periods.

The mean activity area in Table 4.2 is an estimate of the area of the long-term space use for a jaguar in an average environment. During the survey, encounters with each jaguar give information on their realised range: the area within which they spent 95% of the survey time. The time-spent distribution shows the estimated mean time a jaguar spent in each location (Figure 4.3). Notice, that the inclusion of the spatial covariate means each jaguar's activity range is estimated to be stretched along these networks.

Additionally, the estimated realised activity area can be computed (Table 4.3). The realised range is smaller than the stationary range because animals do not explore their entire

long-term activity area during the survey. The realised areas also take account for the environmental features within each individual jaguar's range; here, the availability of rivers and trails will affect the ranging of each jaguar.

These activity ranges are much larger than those reported by Rabinowitz and Jr (1986) for the Belize population; however, Rabinowitz and Jr (1986) estimated home range using a minimum area method that is known to underestimate range by not accounting for the correlation in locations over time (Fleming et al., 2015). Two other popular methods (Worton, 1987) to quantify activity range, minimum convex polygon and kernel density estimation, also underestimate range for this reason. Current estimates of jaguar ranges are made with these existing methods (Quigley & Harmsen, 2017). Thus, an important future step would be use of models that account for temporal autocorrelation either through movement modelling or autocorrelated kernel density estimation (Fleming et al., 2015). Jaguar ranges may be much larger than previously thought, and since a leading threat to jaguars is the shrinking of their individual ranges, better assessment of this quantity is important.

ID	$\hat{H}$	LCL	UCL
1	149	141	162
2	161	146	176
3	153	142	169
4	143	132	156
5	153	142	167
6	174	157	193
7	174	157	191
8	157	146	170
9	155	141	169
10	166	150	183
11	123	106	140
12	148	138	161
13	172	158	189
14	154	143	165

Table 4.3 *Posterior mean estimates and lower (LCL), upper (UCL) 95% credible interval bounds for realised activity area, H, in square kilometres for each individual (ID) from encounter model with Ornstein-Uhlenbeck movement model.*

4.4.5 Interaction

The estimated spatial overlap in the realised activity areas of male jaguars can be substantially high (Table 4.4). The average estimated percentage overlap of a jaguar's activity

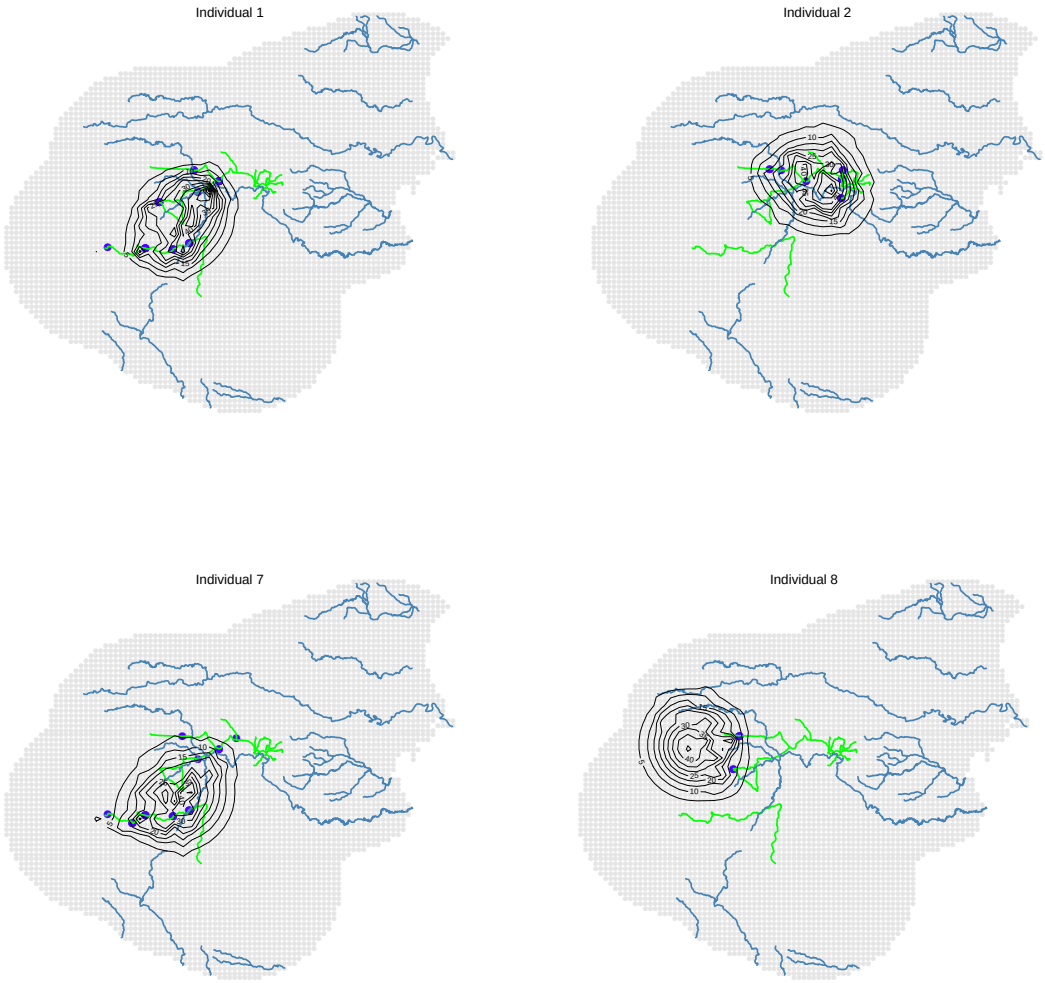


Figure 4.3 *Contours (black) of the posterior mean time spent distributions for individuals 1, 2, 7, 8 over the survey time with the locations of cameras each individual was detected at (blue points). Rivers (blue lines) and trails (green lines) are included.*

range with its closest neighbour was 74%. This corresponds with similar findings (Harmsen et al., 2010; Rabinowitz & Jr, 1986). The question remains whether jaguars do move in close proximity to each other or whether they timeshare their spatial overlap. To investigate, the mean posterior distance that each jaguar spends more than 95% of time from its closest neighbour was estimated. This distance is a quantification of an imagined hard core radius around each jaguar which its neighbours are unlikely to invade. This was estimated to be 1.9 kilometres. Equivalently, this corresponds to a circular hard-core area around each jaguar of approximately 10 square kilometres. This is small compared to the size of the mean jaguar activity range, and small compared to the mean spatial overlap in the activity ranges of neighbours. Altogether, this provides evidence that jaguars share space and interact at a close social scale.

This investigation of interaction is not in-depth. Interaction will change over time both by season and by the biological chronology of the population, for example, when breeding occurs. Importantly, however, these methods provide the modelling framework to make progress in this direction.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1		40	56	19	17	44	96	24	40	14	1	24	21	17
2	37		11	0	63	62	35	6	86	65	0	23	69	53
3	54	11		44	0	24	65	44	12	0	19	25	0	1
4	20	0	47		0	0	31	9	0	0	56	0	0	0
5	17	66	0	0		33	17	0	57	75	0	2	91	42
6	37	57	21	0	29		35	25	65	40	0	54	31	56
7	82	33	57	25	15	35		22	32	12	5	18	18	13
8	22	6	43	8	0	27	25		7	0	1	53	0	1
9	38	90	12	0	56	73	36	7		66	0	28	59	65
10	13	63	0	0	69	42	12	0	62		0	8	67	60
11	2	0	24	65	0	0	7	2	0	0		0	0	0
12	24	25	26	0	2	64	22	56	30	9	0		4	30
13	19	64	0	0	81	31	19	0	53	64	0	3		37
14	17	56	1	0	42	64	15	1	66	64	0	29	42	

Table 4.4 *Posterior mean estimates of percentage overlap of each pair of jaguars encountered in the SCR survey where the $(i, j)^{th}$ entry is the spatial overlap in activity areas of individual i and j expressed as a percentage of the activity area of individual i*

4.5 Discussion

There are two primary contributions made when continuous SCR is recognised as an encounter model and an explicit animal movement model incorporated.

4.5.1 Contribution

1. **This model can account for the correlations in detection times that are caused by contiguous animal movement**. For the jaguar case study this improvement was only marginal because the correlation is low in this survey. There are two properties that determine this correlation. First, the detector configuration, detectors that are more closely knit will induce a higher correlation in the recaptures because there is shorter time between being seen at one detector and the next. The second is the probability of being seen at a detector. The lower this probability, the lower the correlation. In effect, captures at detectors can be seen as telemetry that is thinned by the detection process, the lower this thinning, the higher the correlation. Thus, studies with more closely placed detectors or have detectors with a greater detection range will have greater correlation between detection times and accounting for this may have a more significant effect on inference obtained.

2. **Explicit animal movement models allow the relationship between animals and their habitat to be investigated more deeply.** With an explicit model for animal movement, we can easily incorporate telemetry data. Telemetry data have already been used in SCR models to estimate the activity area but have not been used to incorporate knowledge of animal movement in more detail, for example by quantifying each animal's affinity to their activity centre. SCR data provide information on a larger proportion of the population than typically surveyed in telemetry, while telemetry provides more detailed data on how animals move than an SCR survey. Incorporating both into a single modelling framework allows inference on many animals to be made based on our best knowledge of both the population and animal movement. Here, two possible inference tools have been discussed, the time-spent distribution which estimates the time spent by an animal in each location. This is of interest in itself, but it could also be related to environmental covariates to determine the long-term resource selection or utilisation distribution for each animal. A direct equivalence, however, between this time-spent distribution and the resource selection function has not been proven theoretically, but parallels could be drawn.

The second inferential tool is on interaction. Even though individuals are assumed to be independent, the inference obtained on each of them can be examined to determine if any interaction occurs between them. This was done first by computing spatial overlap in their activity areas during the survey. Spatial overlap can be a useful tool to consider how individuals relate given their characteristics such as age or social status, for example. The second measure of interaction was separation distance, to quantify not only spatial overlap, but spatio-temporal overlap. For it is possible that individuals can have a substantial spatial overlap in their activity areas but share the space such that they never occupy it at the same time. For the jaguars, spatio-temporal overlap was high suggesting greater social cohesion than once thought. Future work could consider the distribution of separation distances under the hypothesis the individuals do not interact; from this, a test of the interaction hypothesis could be conducted formally.

Again, separation distance can be used to assess how individuals of differing social status or gender relate. For the jaguar case study, only male jaguars were considered. It would be interesting to see how these interaction measures could be used to investigate breeding relations between male and females. It has been observed that the density of males and their range change in the breeding season.

4.5.2 Limitations

There are two major limitations of the encounter model approach to be highlighted. The most obvious limitation is the computationally intensive methods required to fit the model. The path integration requires the intensive algorithm as discussed in Chapter 2. On top of this, however, to have individuals obey a more biologically reasonable movement model, the Ornstein-Uhlenbeck, a further level of latency is added, each animal's activity centre. Numerical integration over all possible activity centres would be an ideal approach, but coupled with the path integration this becomes, at present, computationally infeasible. It is discussed in Chapter 6, how progress could be made toward this goal. Here, the alternative approach of data augmentation was used. The Bayesian MCMC algorithm effectively performs the integration over all activity centres for each individual. For the jaguar study, containing fourteen individuals, this was achievable. SCR surveys often have small samples sizes and so this method may be feasible for many of them; however, for studies containing hundreds of animals the time required to fit the model may become impractical.

The second limitation is the ability to make sound inference about animal movement. One must be cautious when making inference about animal movement and habitat use from the encounter model. Individuals are only observed at detectors and so the detector locations will influence the inferences made. For example, if detectors were placed so far apart that few re-captures where made, information about each individual's activity range is low and so most of this is informed by any telemetry data used. Effectively, for every individual there is a weighted mean of the range as described by the telemetry data and that implied by their re-captures. For one individual in the jaguar study, only one detector captured this individual. Inference about this individual's movement is still legitimate, but all that is being done is that all inference about that animal is based on the telemetry data and the environmental covariates within that individual's range. It is thus recommended that inference on movement be considered only with those individuals that have been recaptured sufficiently.

It is an open question what an optimal detector configuration is. This will most certainly vary according to the species and what inference is to be obtained. For the jaguars, there is a trade-off. Detectors are placed primarily on trails and thus estimating jaguar movement off-trail is difficult and so dominated by the telemetry data. Hence, placing detectors off-trail would provide better inference on how individuals behave off-trails. Yet, cameras are placed on trails not only out of practical convenience but because males use these trails and so there are a sufficient number of re-captures to estimate density by SCR. So, if cameras are placed off-trail then one would obtain fewer re-captures of jaguars, leading to poorer information about their density and range. Using more cameras would always be a preferable approach, but economically is seldom a useful suggestion.

Finally, considering detector placement, detectors that are placed tightly together will provide better information on animal movement as the time between re-captures will be small. This is precisely the opposite of what one would desire if using current continuous-time SCR methods as longer times between detections are more likely to ensure detections are independent, as is assumed in those methods. Whilst for encounter models, more recaptures over a short time provide something similar to telemetry data. Saying that, shrinking the detector spacing often leads to a smaller proportion of the population being detected, therefore losing the primary benefit of using the SCR survey data: it gives a larger cross section of the population than a telemetry study would. These are still open questions and a rigorous investigation of study design for SCR surveys is still to be conducted.

4.5.3 Extensions

There are a number of possible extensions to these methods.

1. **Time varying / environmentally varying encounter parameters**. Jaguars are known to be crepuscular and so most captures occur near dawn and dusk. Thus, animal movement rate should vary according to time of day, this was not investigated in this case study. Furthermore, allowing the movement of the animal to change according to space gives a similar inferential framework as step-selection functions, that is, given a certain location what is the probability an animal would move in each possible direction. For the jaguar case study, trails and rivers were preferentially moved along by male jaguars and so if placed in a random location near a river or trail, jaguars would prefer to move in this direction. Hence, similar inference to a step-selection function can be obtained such as given a set of spatial and temporal covariates what would be the probability of an individual moving to one location compared to another.

   Overall, encounter models can provide inference on short-term movement of the individual and, through the time-spent distributions, on the long-term movement of the individual, and both are directly related.

2. **Behaviour**. For the simple case of crepuscular animals, a temporal covariate can account for them moving around more at dawn or dusk. However, animals have a range of different behaviours: they forage, they search for a mate, they rest. And these behaviours need not occur during certain, known times of the day. A latent behaviour-switching process would allow for more detailed inference. Furthermore, the time-spent distributions can be extended to infer not only how long an animal spent in each location, but how long it spent in each location with each behaviour, or how long it exhibited each behaviour. Unlike activity centres, the latent behaviour process can be included without further computational difficulty. This is similar to the inference made by Pedersen et al. (2011) when using telemetry data.

3. **Dispersal**. In the case study, activity centres were static. Yet, within the MCMC algorithm, an individual's activity centre could drift over time, termed dispersal. For jaguars, there is evidence that dispersal occurs, particularly for those males that only occupy the study area during breeding season. Incorporating this would not require more computational difficulty, but it would increase the time until the MCMC chain converges due to the dramatic increase in the number of parameters. A further

embellishment would be to allow activity centres of different individuals to be related, such as to repel each other, to reflect territorial behaviour.

4. **Model-based interaction**. Not only interaction between activity centres, but interaction in their continuous-time movement. This is a more challenging development because the assumption of independent movement of animals allows for the computations to be achieved independently, if dependence between individuals is incorporated it is not clear how to overcome this computationally. The difficulty is computer storage. For individuals to interact, one must store probability densities of every individual on the spatial grid, for every time, and evolve these densities over time simultaneously. This is not possible given current techniques presented in Chapter 2. The partial differential approach to the movement modelling used in Chapter 2, however, does provide a fruitful framework for future development where interactions between individuals is described by mutually interacting advection terms.

5. **Interaction with "others"**. An exciting prospect is that due to data augmentation, not only are those animals that are encountered considered but also those who were not encountered. Therefore, rather than simply investigate the interaction between encountered animals, we can account for the fact that some individuals may be responding to interaction with unseen members of the population. Similar to how a planet's location can be inferred from its gravitational effect on neighbouring bodies, the existence of an animal in a population, beyond detectability, can be inferred from any interactive behaviour within the model of animals that are encountered. This approach is fundamentally different to group dynamic models in movement ecology that are limited to the modelling of tagged animals only and not untagged group members.

In conclusion, the fusion of continuous-time survey data and telemetry data with the encounter model brings new possibilities to the investigation of animal populations in terms of their density, movement, spatio-temporal habitat use, and interaction.

Chapter 5

Discrete-time Spatial Capture-Recapture

5.1 Introduction

Historically, capture-recapture surveys were carried out by scientific observers who visit a study population on two or more occasions (Pollock, Nichols, Brownie, & Hines, 1990). These observers would capture some sample of the population, uniquely mark those individuals, and then release them back into the population again. On subsequent occasions, they would do the same whilst also recording any individuals in their sample that have been previously captured. Thus, each marked individual has a capture history that is recorded over time. This information is used to estimate the number of individuals that were never captured, and thus, estimate the total population size.

Spatial capture-recapture (SCR) (Borchers & Efford, 2008; Efford, 2004) is an extension that recognises that the probability an individual is captured depends on the position in space of its activity centre relative to where the captures occur. Incorporating space has two advantages: it accounts for a large source of individual heterogeneity in capture probability, individuals that are active closer to the detectors are more likely to be captured, and including space provides a rigorous estimate of the effective area sampled by the detectors, thus allowing estimates of abundance to be converted to estimates of density.

In the previous chapter, continuous-time SCR was presented where the exact time of a capture was known, such as when using cameras with the ability to time-stamp each recorded photograph. For physical detectors or DNA hair traps, exact capture times are unknown, and so the survey is divided into occasions where each occasion begins when the detectors

are replenished (e.g. physical traps emptied or hair traps replaced). For cameras or microphones, where time can be recorded, occasions can still be created artificially by dividing the captures into days, weeks, or months. Doing so, despite the ease of continuous-time methods, is a routine approach in the application of SCR (Borchers et al., 2014).

An aim of this thesis is to explore how animal movement can be incorporated into existing population abundance methods. For discrete-time SCR, this leads to an investigation of whether the activity centre of each individual can be permitted to move over time. At present, standard SCR methods make the assumption that activity centres are static and that individuals explore the area around their activity centre according to a Gaussian distribution within each occasion. Estimates of population size using SCR are known to be robust to violations of these assumptions (Efford, 2014). Although, such violations can lead to underestimated uncertainty in density estimation and can cause substantial bias in the estimation of other parameters such as activity area (Royle et al., 2016). Furthermore, extensions of SCR that allow for open populations, where individuals may be born or die during the survey, can lead to biased inference when the movement of individuals over these longer surveys is not taken into account.

A common approach when movement is undeniable is to use a robust design (Hines, Kendall, & Nichols, 2003). In a robust design, the survey time is split into primary occasions and then each of these is split into secondary occasions. The assumption is made that activity centre movement does not occur within primary occasions; thus, standard SCR methods can be applied to each primary occasion separately; furthermore, if primary occasions are treated as independent, parameters can be shared. This is the standard approach within the maximum likelihood SCR framework (e.g. Karanth, Nichols, Kumar, and Hines (2006)). Captures of the same individual in two different primary occasions are treated as captures of two distinct individuals, thus movement of the individual over time does not affect estimates of activity range. For open population models, where individuals may die or spontaneously arrive during the survey, the robust design ignores these individual-level behaviours, providing estimates of population size for each primary occasion. This has the advantage that information about detectability can be shared across primary occasions. Ultimately, however, this avoids the need to break the assumption of immobile activity centres, rather than attempting to remove it.

Methods to remove this assumption have been developed in a Bayesian framework (Gardner et al., 2010). The Bayesian approach, conceptually, simulates each individual, detected or not, in the survey area and thus allowing activity centres to change over primary occasions is simply a matter of adding a random effect for every individual, for every primary occasion.

It is the aim of this chapter to fit this model by maximum likelihood. This is to explore whether likelihood methods would be faster, more amenable to analyses involving large data sets, or useful in the improvement of existing Bayesian methods. The necessary statistical theory and computational algorithms to do this are those that have been developed in Chapter 2.

The encounter model presented in Chapter 2 is formulated for surveys conducted in continuous-time. In this Chapter, it is shown that a discrete-time encounter model can be developed (Section 5.3) and benefit from the computational methods in Chapter 2. This method can allow individual activity centres to move over time according to a Gaussian process. Allowing activity centres to move over time is the theoretical contribution. The inferential and ecological contribution depends on the context of the survey and the motivation for allowing activity centres to move. This is somewhat similar to mixed modelling where random effects can describe variability but their interpretation depends on the underlying cause.

In this chapter, three possible motivations for allowing activity centres to move are considered:

1. individuals do not explore their entire activity range within each occasion (Section 5.4);
2. some individuals in the population are resident with immobile activity centres and some transient with moving activity centres (Section 5.5);
3. individuals disperse over long-term surveys (Section 5.6).

These three violations of the fundamental assumptions of SCR can be accommodated by incorporating movement. To show this, the theory of discrete-time SCR is briefly described (Section 5.2) before the computational methods required to allow activity centre movement are introduced. The method is then applied to real and simulated data to demonstrate its use in each of the three situations mentioned above.

5.2 SCR

The discrete-time spatial capture-recapture model is similar to the continuous-time model discussed in the previous chapter. Conceptually, every individual i has assigned to it an activity centre, $\boldsymbol{x}_i$. Activity centres are assumed to arise from a spatial point process, most commonly an inhomogeneous Poisson point process with rate $D(\boldsymbol{x})$.

Within each occasion, the detection process can be described using the theory from Section 2.2. In particular, there is an encounter intensity or hazard function that quantifies the rate an individual encounters each detector, commonly depending on the distance between the detector and the individual's activity centre. The difference between the theory as presented in Section 2.2 and the application here is that instead of considering an individual's continuous location over time, the method is applied to their activity centres over time. Let $\lambda_j(\boldsymbol{x})$ denote the encounter intensity between detector j and an individual with activity centre $\boldsymbol{x}$. Examples of functional forms for encounter intensities and their associated hazards are given in Section 2.2.

Following classical capture-recapture, the probability of an individual being captured in a given occasion can be modelled directly. This probability is given by Theorem 2.2 to be $p_j(\boldsymbol{x}) = 1 - \exp(-\lambda_j(\boldsymbol{x}))$. A common functional form for p_j is the half Gaussian:

$$p_j(\boldsymbol{x}_i) = g_0 \exp\left(-\frac{\|\boldsymbol{x}_i - \boldsymbol{z}_j\|^2}{2s^2}\right)$$

where

- $\boldsymbol{z}_j$ is the location of detector j;
- g_0 is the probability an individual with activity centre coincident with the location of detector j is captured in an occasion; here, this is termed the detection function intercept;
- s is the activity range of an individual within a single occasion.

The associated encounter intensity for any specified functional form for p_j is easily derived: $\lambda_j(\boldsymbol{x}) = -\log(1 - p_j(\boldsymbol{x}))$.

Ultimately, the detection function quantifies the amount of time an individual spends in each location around its activity centre and thins this by the capture probability. Implicitly, this assumes that an individual explores its entire range within each occasion and does not move around its range piecemeal. In other words, the range of the individual over the entire survey is assumed to be the same as its range within each occasion. This can be a reasonable assumption when occasions span a long interval of time; however, for short occasions, it is possible that individuals may only explore their activity range partially. When this occurs, the activity range within each occasion is overestimated; thus, use of this model makes the incorrect assumption that individuals were failed to be captured rather than the correct assumption that they failed to explore their entire range. It is this shortcoming of the

standard approach that is the focus of this Chapter.

Given a model for the detection process, the likelihood for the capture history of individual i, denoted $\boldsymbol{\omega}_i$, can be derived for each type of detector that may be used:

- *Multi-catch traps*: within an occasion, each individual is either not captured or captured by a single trap; the response is binary and the capture record for individual i at trap j on occasion k is denoted $\omega_{i,j,k}$. This random variable has a multinomial distribution with $J+1$ outcomes where J is the number of traps. This model is easier to formulate using the hazard function for each trap, λ_j. Conceptually, traps are competing for each individual, and so a competing risks model can be used (Borchers & Efford, 2008). The hazard of an individual being captured by any trap is the sum of the individual hazards $\Lambda(\boldsymbol{x}_i) = \sum_{j=1}^{J} \lambda_j(\boldsymbol{x}_i)$.

  The first outcome of the multinomial represents the event an individual is not captured by any trap. This has probability $q_i = \exp(-\Lambda(\boldsymbol{x}_i))$ for individual i. The remaining J outcomes represent the events where an individual is captured and has total probability $p_i = 1 - q_i$. The probability of an individual being captured by trap j given it is captured by at least one trap is simply the proportional hazard: $\dfrac{\lambda_j(\boldsymbol{x}_i)}{\Lambda(\boldsymbol{x}_i)}$.

  The probability density of the observed capture history is thus:

$$[\boldsymbol{\omega}_i \mid \boldsymbol{x}_i] = \prod_{k=1}^{K} q_i^{\omega_{i,\cdot,k}} p_i^{1-\omega_{i,\cdot,k}} \prod_{j=1}^{J} \left(\frac{\lambda_j(\boldsymbol{x}_i)}{\Lambda(\boldsymbol{x}_i)} \right)^{\omega_{i,j,k}}$$

  where K is the total number of occasions and $\omega_{i,\cdot,k} = 1$ if individual i is captured in occasion k and zero otherwise.

- *Proximity detectors*: For proximity detectors, individuals can be detected by multiple detectors multiple times within a single occasion. The derivation of the associated probability density is similar to Theorem 1.1. For count proximity detectors, $\omega_{i,j,k}$ is the number of times individual i is detected by detector j in occasion k. By Theorem 1.1, this can be assumed to have a Poisson distribution with mean $\lambda_j(\boldsymbol{x})$. For binary proximity detectors, $\omega_{i,j,k} = 1$ if the individual i was detected by detector j at least once during occasion k and zero otherwise; this random variable has a Bernoulli distribution with probability parameter $1 - \exp(-\lambda_j(\boldsymbol{x}))$.

The likelihood given all observed capture histories $\boldsymbol{\Omega}$ can then be written as:

$$[\boldsymbol{\Omega}] = \int \dots \int [\boldsymbol{x}_1, \dots, \boldsymbol{x}_n] \prod_{i=1}^{n} \frac{[\boldsymbol{\omega}_i \mid \boldsymbol{x}_i]}{p_\cdot(\boldsymbol{x})} \, \mathrm{d}\boldsymbol{x}_1 \dots \boldsymbol{x}_n$$

where one integrates over all possible activity centres each of the n captured individuals may have had. The probability density of the observed activity centres is given by an inhomogeneous Poisson process with rate $D(\boldsymbol{x})p_\cdot(\boldsymbol{x})$ where $p_\cdot(\boldsymbol{x})$ is the probability an individual with activity centre $\boldsymbol{x}$ is detected at least once during the survey: $p_\cdot(\boldsymbol{x}) = 1 - \exp(-K\Lambda(\boldsymbol{x}))$. This is equivalent to a thinned Poisson point process where the detection probability is the thinning probability.

There are two approaches to parameter estimation. Maximum likelihood estimation (MLE) uses numerical integration to compute the marginal likelihood, which is maximised to obtain parameter estimates (Borchers & Efford, 2008). The Bayesian approach uses data augmentation to estimate population size and detection parameters (Royle & Young, 2008). A meta-population of M individuals is created, augmenting the observed capture histories with $M - n$ empty capture histories. An inclusion parameter is then used to estimate the proportion of this meta-population that constitutes the true population of size N. Inference can then be based on the full likelihood:

$$[\boldsymbol{x}_1, \dots, \boldsymbol{x}_N][\boldsymbol{\omega}_1, \dots, \boldsymbol{\omega}_N \mid \boldsymbol{x}_1, \dots, \boldsymbol{x}_N]$$

Since the inception of SCR, both the Bayesian and MLE approaches have been developed in tandem. The common theme is that MLE is more computationally efficient, simpler to test by simulation, and lends itself more easily to spatio-temporal modelling. The Bayesian method is simpler to implement and is more flexible, especially when modelling individual level variation such as uncertain ID or open population models where individuals may be born or die during the survey.

5.3 Discrete-time encounter model

The marginalised likelihood used in MLE methods is computationally tractable because it can be reduced to a two-dimensional integration over space for each individual. When the activity centre of an individual changes by occasion and these centres are dependent over time, the integration is $2K$-dimensional where K is the number of occasions. Due to the curse of dimensionality, numerical integration techniques are either inaccurate or

require a high computational burden to produce an accurate result. The alternative, Monte Carlo simulation, is effectively equivalent to taking a Bayesian approach. This obstacle is the same as that found when integrating over all possible paths within a continuous space in discrete time as discussed in Section 1.5. In that case, HMM quadrature is presented to integrate over all simple paths (Definition 1.13) which is equivalent to a discrete-time encounter process. Thus, the efficient algorithm for this task can be applied here to compute the marginal likelihood.

Let $\boldsymbol{x}_{i,k}$ be the activity centre for individual i on occasion k. The likelihood contribution from a single individual i is thus

$$[\boldsymbol{\omega}_i] = \int \dots \int [\boldsymbol{x}_{i,1}, \dots, \boldsymbol{x}_{i,k}][\omega_i \mid \boldsymbol{x}_{i,1}, \dots, \boldsymbol{x}_{i,K}] \, \mathrm{d}\boldsymbol{x}_{i,1} \dots \boldsymbol{x}_{i,K}$$

where $[\omega_i \mid \boldsymbol{x}_{i,1}, \dots, \boldsymbol{x}_{i,k}]$ is the appropriate extension for the given detector type where the activity centre is changed by occasion. The exact form of this function will depend on the model considered. The only assumption made is that this function must be Markov-separable (Definition 1.14); this is equivalent to the assumption that occasions are independent given an individual's activity centres.

The probability density of the activity centres for individual i is given once a model has been specified for the initial distribution of activity centres at the beginning of the survey and how activity centres change over time. As before, an inhomogeneous Poisson process can describe the position of activity centres at the beginning of the survey. At the end of each occasion, the activity centres are assumed to move according to a bivariate Gaussian distribution, that is, $\boldsymbol{x}_{i,k+1}$ given $\boldsymbol{x}_{i,k}$ is a bivariate Gaussian distributed variable with mean $\boldsymbol{x}_{i,k}$ and variance $\sigma^2 \boldsymbol{I}$ for identity matrix $\boldsymbol{I}$.

Theorem 2.5 shows that the marginalised likelihood contribution $[\boldsymbol{\omega}_i]$ can be approximated by the likelihood for a hidden Markov model. It is only an approximation as space is discretised in G cells for some large G. The approximation is computed using three components:

1. *initial distribution*: a row vector of length G with g^{th} entry equal to the probability that $\boldsymbol{x}_{i,0}$ lies inside grid cell g. This probability is deduced from the spatial point process model.

2. *observation matrix*: a $G \times G$ diagonal matrix $\boldsymbol{F}_k$ with g^{th} diagonal equal to the probability of observing the capture history recorded for occasion k given the individual has an activity centre that lies inside cell g during occasion k;

3. *transition matrix*: a $G \times G$ matrix $\boldsymbol{\Gamma}_k$ with $(g,h)^{\text{th}}$ entry the probability of an activity

centre moving from cell g to cell h at the end of occasion k

Given these three components, the integral can be approximated by the formulae in Theorem 2.5. However, counterintuitively, the computational gains of the methods in Chapter 2 cannot be realised once a continuous-time movement model is replaced with a discrete-time model. The reason for this lies in how the computational gains are made: direct calculation of $\boldsymbol{\Gamma}_k$ is avoided by computing the transition rate matrix $\boldsymbol{R}_k$ and then using a Krylov approximation to compute the vector-matrix products involving $\boldsymbol{\Gamma}_k = \exp(\boldsymbol{R}_k \Delta t)$ for some time-step Δt. The transition rate matrix describes the rate an individual moves from one grid cell to the next over a very short time; thus, this matrix is typically very sparse as individuals can only move small distances in small times. Conversely, over a long time period, such as over the period separating two capture occasions, an individual's activity centre may have moved a substantial distance and $\boldsymbol{\Gamma}_k$ is typically a dense matrix. It is this computational obstacle that has prevented wider use of spatial hidden Markov models, HMMs with a spatial latent process.

The solution proposed here is to assume a continuous-time movement model and thus obtain a transition rate matrix $\boldsymbol{R}_k$ as shown in Section 2.5.2. Matrix-vector products involving $\boldsymbol{\Gamma}_k$ can then be computed by time-stepping. Effectively, every required matrix-vector product $\boldsymbol{v}\boldsymbol{\Gamma}_k$ is computed as

$$\boldsymbol{v}\boldsymbol{\Gamma}_k = \boldsymbol{v}\exp(\boldsymbol{R}_k) = \boldsymbol{v}\exp\left(\sum_{t=0}^{T} \boldsymbol{R}_k \Delta t\right) = \boldsymbol{v}\prod_{t=0}^{T} \exp(\boldsymbol{R}_k \Delta t)$$

where $\Delta t = \dfrac{1}{T}$ and T is chosen to ensure the approximation has the required accuracy. The matrix-vector product $\boldsymbol{v}\exp(\boldsymbol{R}_k \Delta t)$ can be computed more efficiently, due to the sparsity of $\boldsymbol{R}_k$, compared to Γ_k. Choice of T is discussed by Saad (1992a) and there are approximate upper bounds for the error in this approximation that are used to select T (Sidje, 1998).

This approach allows the computational savings discussed in Chapter 2 to take effect and the integration over all activity centre paths to be approximated efficiently. Furthermore, the approach is convenient as it allows for the time between capture occasions to be accounted for in the model. For example, if Δt_k is the time between the end of occasion k and the start of occasion $k+1$, then $\boldsymbol{\Gamma}_k = \exp(\boldsymbol{R}_k \Delta t_k)$ can be computed. This is an important adjustment for surveys where occasions are irregular since activity centre locations are likely to be less correlated between occasions separated by longer intervals of time.

Given this method for computing the high-dimensional integral, the marginal likelihood can be computed and parameters of the movement and detection processes estimated by

maximum likelihood. In the remainder of this chapter, this algorithm is used to compute the marginal likelihood for different formulations of $[\boldsymbol{\omega}_i \mid \boldsymbol{x}_{i,1}, \ldots, \boldsymbol{x}_{i,K}]$. Nevertheless, it is the same computational algorithm that makes the application feasible.

5.4 Uneven Space Use

Over a survey, individuals explore an activity area. Standard SCR models assume that individuals explore the same area within each occasion. In many cases, it is possible that individuals use their activity area partially. Conceptually, one can imagine the activity area over the entire survey being composed of a superposition of smaller activity areas that are explored within each occasion. The centres of these activity areas shift from one occasion to the next and their average is the activity centre over the entire survey. Allowing activity centre to change by occasion can account for uneven space use during the survey.

Such models have been developed in the Bayesian framework by Royle et al. (2016) and applied to a data set of black bears. The MLE method proposed above is applied to the same case study for comparison.

5.4.1 Black Bears

The survey was conducted in Fort Drum, New York. DNA hair traps were deployed over an eight week period and replenished every week. The objective was to investigate whether individuals used their activity area unevenly, leading to an overestimation of their weekly activity range.

The Bayesian approach, as applied by Royle et al. (2016), allowed activity centres to move according to a Gaussian distribution and the detection probability was assumed to be Gaussian also. The same formulation was used in the MLE approach. Table 5.1 shows the posterior means from the Bayesian model and the estimates from the MLE model for all parameters.

Estimates and uncertainty from both methods appear similar. This indicates that the discretisation of space, to approximate the integration, can lead to inference that is as accurate as produced by continuous space methods such as MCMC. As a side note, on a desktop computer with 32 gigabytes of memory and a 3.2GHz processor, the Bayesian model took four hours to produce 40000 iterations, while the MLE model was fit in ten minutes. It is difficult to compare the methods fairly, but it is clear that the MLE method provides inference in less time. The gain is minimal in practical terms, but it provides evidence that

	Bayes	MLE	Basic SCR
g_0	0.15 (0.10, 0.21)	0.13 (0.09, 0.18)	0.11 (0.09, 0.14)
s	1.69 (1.44, 1.98)	1.71 (1.44, 2.02)	1.98 (1.75, 2.26)
σ	1.07 (0.61, 1.61)	0.98 (0.57, 1.67)	-
D	0.17 (0.14, 0.20)	0.17 (0.12, 0.22)	0.17 (0.12, 0.23)

Table 5.1 *Bayesian posterior means (Bayes) and maximum likelihood estimates (MLE) for SCR with activity centre movement and maximum likelihood estimates for standard SCR model (basic SCR) with associated* 95% *credible and confidence intervals in brackets for detection probability intercept g_0, weekly activity range s in kilometres, standard deviation of weekly activity centre movement σ in kilometres, and population density per square kilometre, D.*

the MLE method should be considered when application of the Bayesian method to larger sample sizes or larger, data-augmented populations seems infeasible. One disadvantage of the maximum likelihood approach is that it is more difficult to estimate uncertainty in derived quantities of interest compared to the Bayesian approach where the posterior for such derived quantities is easily obtained.

A standard SCR model was fit to the data by maximum likelihood (Table 5.1). The standard SCR model estimated weekly activity range to be larger than when movement is incorporated. Comparison by AIC (SCR model has 1179, while model with movement has AIC 1173) or by a likelihood ratio test (p-value < 0.005) provided evidence that activity centres moved over time. Given the short period of the study, one can infer this is most likely due to individuals using their activity area unevenly over the survey time, rather than due to shifts in their activity area.

It is likely that in many applications detections of individuals are clustered within their own activity area. A model with mobile activity centres can account for this and can be used to test for whether such an effect exists. Viewing each activity centre as a two-dimensional random effect for each individual, a model with moving activity centres simply extends this to allow a series of correlated random effects over time. The MLE method provides inference efficiently as an alternative to the Bayesian approach without compromising on the quality of inference.

5.5 Transience

Allowing each individual's activity centre to move from one occasion to the next can be termed transience. In the previous section, mobile activity centres were used to account for uneven activity area use. The term transience, however, is used to describe a different

underlying cause: it describes those individuals whose activity range shifts over time, and thus a model for such movement is not merely a device to account for uneven space use, but an attempt to describe a biological process.

The case where every individual's activity centre moves is called complete transience, and has the same likelihood as applied in the previous section. This could be relevant for studies on populations that are highly mobile, moving with prey/resource availability or predator pursuit. This is contrary to the population idealised by conventional SCR methods: individuals are all resident with stationary activity centres. Other populations may comprise a mixture of these two extremes: some individuals will be resident and best described by a stationary activity area, while others are transient and move through the study area over time. This is termed partial transience. Transient individuals are common in studies and can lead to overestimation of activity range, underestimation of encounter rate, and bias in estimation of uncertainty. Here, a model for complete and partial transient populations is presented.

5.5.1 Model

Models for complete transience and complete residence exist within the Bayesian framework (Royle et al., 2016), but none have yet been implemented for partial transience. In this section, this model is presented and fit by maximum likelihood methods. Again, this is made possible by the ability to compute the path integral.

The population is divided into two classes: residents and transients. Residents are assumed to have stationary activity centres, while a transient's activity centre drifts by occasion according to a Gaussian distribution. Let ψ be the probability an individual is a transient within the population. For individual i, let $z_i = 1$ if the individual is a transient and zero otherwise. It follows that $[\boldsymbol{\omega}_i \mid z_i = 0]$ is the likelihood contribution of individual i under the conventional SCR model as described , while $[\boldsymbol{\omega}_i \mid z_i = 1]$ is the likelihood contribution from a complete transient model. In other words, the partial transience model is a mixture model:

$$[\boldsymbol{\omega}_i] = [\boldsymbol{\omega}_i \mid z_i = 0](1 - \psi) + [\boldsymbol{\omega} \mid z_i = 1]\psi$$

with mixture probability equal to the probability of being a transient. Intuitively, the observed capture histories are clustered by the model into those coherent with stationary activity centres and those not. Computation of this likelihood is achieved by numerical integration for the standard SCR component and by the algorithm for path integration for the transient component.

5.5.2 Simulation Study

The method was tested by the same simulation setup used by Royle and Young (2008) to assess the performance of standard SCR models when applied to populations with transience. In the central ninth of a 31 square kilometres study area, a 10 by 10 grid of proximity detectors was placed. A constant density of 1.2 individuals per square kilometre was simulated for each scenario. A Gaussian detection function was used with fixed parameters $g_0 = 0.2$ and $s = 0.7$ kilometres. This led to approximately 150 unique individuals captured per occasion.

Four populations were considered: lowly transient ($\psi = 0.25$), half transient ($\psi = 0.5$), highly transient ($\psi = 0.75$), and completely transient ($\psi = 1$). For each population, two levels of movement were considered: movement equal to the activity range and smaller than the detector spacing $\sigma = 0.7$, and movement that exceeded both $\sigma = 1.5$. Each scenario was simulated 500 times. The estimated median bias and 95% confidence interval coverage were calculated. The median bias was used as the asymptotic distribution of the maximum likelihood estimator of ψ was poorly approximated by a Gaussian distribution. This is a common phenomenon when considering the asymptotic properties of estimated probabilities and odd ratios (Lyles, Guo, & Greenland, 2012). The estimated parameter is not ψ but $\beta = \log\left(\frac{\psi}{1-\psi}\right)$. The parameter β is approximately Gaussian distributed and has negligible mean bias; however, ψ can have substantial positive bias for small sample sizes. Despite this, ψ remains median unbiased. For the purposes of checking the performance of the method, the median was thus preferred. As a side note, the mean bias observed in ψ across all scenarios was $< 5\%$ in all cases.

ψ	σ	$\hat{g}_0$	$\hat{s}$	$\hat{\sigma}$	$\hat{D}$	$\hat{\psi}$
1.00	0.7	0.00 (0.93)	-0.92 (0.91)	-0.54 (0.91)	0.26 (0.98)	-
1.00	1.5	1.10 (0.96)	-0.89 (0.94)	0.84 (0.97)	-0.57 (0.95)	-
0.75	0.7	0.58 (0.96)	-0.12 (0.96)	-0.41 (0.95)	1.46 (0.96)	0.15 (0.99)
0.75	1.5	1.27 (0.95)	-0.19 (0.95)	-0.81 (0.93)	-2.42 (0.95)	-0.55 (0.90)
0.5	0.7	1.36 (0.98)	-1.01 (0.95)	-0.94 (0.93)	-0.26 (0.98)	-0.05 (0.98)
0.5	1.5	0.64 (0.96)	-0.34 (0.96)	1.32 (0.96)	-0.14 (0.94)	0.34 (0.94)
0.25	0.7	-0.34 (0.94)	0.18 (0.98)	1.71 (0.93)	-0.30 (0.96)	0.74 (0.94)
0.25	1.5	0.60 (0.96)	-0.21 (0.99)	0.22 (0.96)	-0.35 (0.96)	3.52 (0.97)

Table 5.2 *Median bias and, in brackets, 95% confidence interval coverage for estimated detection probability intercept $\hat{g}_0$, weekly activity range $\hat{s}$, standard deviation activity centre movement $\hat{\sigma}$, population density $\hat{D}$, and probability of being a transient $\hat{\psi}$ for different levels of true transience probability ψ and activity centre movement σ with fixed detection parameters and density*

Overall, the estimated bias and confidence interval coverage for all parameters across all scenarios is reasonable (Table 5.2). In particular, the substantial overestimation of activity range found by Royle and Young (2008) to be present in standard SCR models has been reduced and is now negligible. Furthermore, inference on transience is obtained accurately, providing further detail on the behaviour of captured individuals. Bias in the estimated transience range σ and probability ψ is greater for the lower transient population $\psi = 0.25$ as fewer transients are observed in each survey.

The complete transience model was developed by Royle and Young (2008) and implemented in a Bayesian framework. Royle and Young (2008) found similar bias in each parameter; credible interval coverage for the detection and transience parameters was also similar. However, estimated 95% credible interval coverage for density was underestimated ($< 90\%$). Here, confidence interval coverage for density is close to nominal. The reason for this disparity is not due to a difference between the maximum likelihood and Bayesian approaches, but due to a conceptual error in the implementation of the Bayesian method.

In the standard implementation of Bayesian SCR models, a buffer region around the detectors is specified subjectively and the assumption made that no individual detected in the survey has an activity centre outside this area. This is a form of restrictive prior on the activity centres. The meta-population of individuals created by the data augmentation approach is then simulated to have activity centres within this buffer area. Density is thus the number of individuals in the meta-population that is estimated to be in the realised population divided by the area of this buffer. When allowing for transience, Royle and Young (2008) adopted the same buffer and so the same restrictive prior on the initial activity centre of each individual. For subsequent occasions, the activity centre of each individual could be outside the buffer. In their simulation, however, individuals captured in the survey could have activity centres located outside this buffer at the beginning of the survey. Thus, under the Bayesian model as implemented, the number of captures and so the density is considered to be less variable than as constructed in the simulation. This led to underestimation of variability and thus less than nominal credible interval coverage. The same effect did not occur in the maximum likelihood method as the size of buffer was chosen so as to ensure it includes the initial activity centre of any individual that could be seen during the survey. If the size of the buffer region were increased in the Bayesian implementation, the credible interval coverage would improve.

Another issue discovered when performing the simulation study concerned the model fitting. Estimation of ψ and σ is only meaningful when the observed captures indicate that transience may have occurred. In the simulation scenario where $\psi = 0.25$, in around 5% of

cases, no transient individuals were detected in the survey; in these cases, the model, fit by numerical optimization, failed to converge, as $\hat{\sigma} = 0$ lay on the boundary of the parameter space. These cases were excluded on the grounds that in such a situation, as the captures contain no indication of transience, a transience model would not be considered appropriate. It should also be noted that for low ψ, the likelihood is more flat due to the greater uncertainty in this parameter; gradient-based numerical optimization methods, such as the Gauss-Newton algorithm, performed poorly in these cases compared to the more robust Nelder-Mead simplex algorithm.

In conclusion, the simulation study demonstrates that the method is applicable to transient populations. This can serve to improve uncertainty estimation in density and the validity of biological interpretations of activity range. Viewed as a mixture model, the method can be extended to include multiple mixtures describing sub-populations of residents, mildly transient, and highly transient individuals. It is likely the numerical optimization required to fit such models would demand clear evidence in the observed data for such models, similar to classical Gaussian mixture models, where substantial overlap of two mixtures can lead to non-identifiability.

5.6 Dispersal

The final violation of the immobility assumption considered here is termed dispersal. Dispersal is the process of an individual's activity centre drifting over a long time period; it is distinct from transience only in biological interpretation and focus: transience describes short-term, nomadic behaviour in the population, while dispersal describes drift over the temporal scale at which the population's dynamics are at work.

A central assumption of many capture-recapture methods is that the surveyed population is closed. During the survey time, no individuals can leave or enter the survey: none can emigrate from the study, die, be born, or immigrate. When a population is surveyed over a time span where these processes of migration and survival are irrefutably at work, the population is said to be open. There are two popular capture-recapture models used to study open populations: the Cormack-Jolly-Seber (CJS) (Cormack, 1964) and the Jolly-Seber (JS) (Jolly, 1965; Seber, 1965). The CJS model is applied to surveys where a known number of individuals are alive in the population and their capture histories recorded after this time. From this, estimates of capture probability and survival probability are obtained. CJS methods do not estimate density as they are based entirely around the known number of individuals tracked from the beginning of the survey. This is equivalent to conditioning

on the first capture of an individual. The Jolly-Seber model is an extension of CJS that allows for individuals to enter the population, either by birth or immigration. Entering the population is known as recruitment. JS models can be used to estimate survival, recruitment, and density from capture-recapture data. Bayesian and maximum likelihood methods exist to fit CJS and JS models within a spatial capture-recapture likelihood (Gardner et al., 2010; Glennie, Borchers, Murchie, Harmsen, & Foster, 2017).

The maximum likelihood approach models each individual's life history as a Markov chain. For the CJS model, individuals can exist in two life states: available and unavailable. These states are often called "alive" and "dead." However, CJS models cannot distinguish between the event of an individual's death and the event the individual has permanently emigrated from the study region. Both events lead to the same observation: no subsequent captures. For this reason, estimates of survival probability are termed estimates of apparent survival. Permanent emigration from the study region leads to an underestimation of survival, as all those individuals that emigrate are assumed dead (Hines et al., 2003). JS models are similar where individuals can exist in three states: awaiting to be recruited, available, and unavailable. The switching of individuals from one state to the next is a Markov process governed by the recruitment and survival probabilities.

There has long been a desire to separate the survival and migration processes to obtain unbiased estimation of survival probabilities. To do so, one must account for the movement of individuals from one occasion to the next, and, crucially, allow individuals to leave the area around the detectors permanently. A Bayesian approach to this was developed by Ergon and Gardner (2014). The maximum likelihood alternative is developed in this section. The computational algorithm required to compute the integration over all activity centre movements requires the integrand to be Markov-separable; for the CJS and JS models, formulated as hidden Markov models as in Glennie et al. (2017), this property is met.

5.6.1 Model

Here, the CJS model with dispersal is described; the extension to the JS model is equivalent to that described in Glennie et al. (2017).

Let $l_{i,k} = 1$ when individual i is available (alive) during occasion k and zero otherwise. For CJS models, it is assumed that $l_{i,1} = 1$ for all individuals i. For brevity, it is assumed that all individuals are tracked from the same occasion onward; in practice, each individual's capture history may span a different set of occasions; if so, this is easily accommodated as shown by Gardner et al. (2010).

The SCR extension of the CJS is formulated as a hidden Markov model with two hidden states. Let $\boldsymbol{v}_{i,k}$ be a 2×1 row vector with entry r equal to the probability $[l_{i,k} = r - 1]$. By assumption, $\boldsymbol{v}_{i,1} = (0, 1)$. The transition probability matrix is the 2×2 matrix:

$$\boldsymbol{\Gamma}_l = \begin{pmatrix} 1 & 0 \\ 1 - \phi & \phi \end{pmatrix}$$

where ϕ is the survival probability.

Again, let $\boldsymbol{x}_{i,k}$ be the activity centre of individual i during occasion k. The probability density of the observed capture history for individual i during occasion k depends on the relevant activity centre and the life state of the individual: $[\boldsymbol{\omega}_{i,k} \mid \boldsymbol{x}_{i,k}, l_{i,k}]$ is given by the form appropriate to the detectors used when $l_{i,k} = 1$, is unity when $l_{i,k} = 0$ and the capture history $\boldsymbol{\omega}_{i,k}$ represents no captures, and finally is zero when $l_{i,k} = 0$ and the capture history indicates captures occurred in occasion k for individual i. Let $\boldsymbol{P}_{i,k}(\boldsymbol{x}_{i,k})$ be a 2×2 diagonal matrix with r^{th} diagonal equal to $[\boldsymbol{\omega}_{i,k} \mid \boldsymbol{x}_{i,k}, l_{i,k} = r - 1]$. It follows from the theory of hidden Markov models (Zucchini et al., 2016) that

$$[\boldsymbol{\omega}_i \mid \boldsymbol{x}_{i,1}, \ldots, \boldsymbol{x}_{i,K}] = \boldsymbol{v}_{i,1} \prod_{k=1}^{K} P_{i,k}(\boldsymbol{x}_{i,k}) \Gamma_l.$$

This is the integrand necessary for the integral over all activity centre movements. The likelihood can then be computed and numerically optimized to obtain maximum likelihood estimates of survival and detection probability.

5.6.2 Voles

Ergon and Gardner (2014) used the Bayesian approach to the CJS-SCR model with dispersal to investigate the effect of accounting for dispersal when estimating survival. They applied the Bayesian method to a study of voles over a small trapping grid within a single season.

The population was surveyed on four primary occasions approximately 22 days apart. Within each primary occasion, 192 traps were placed in a regular grid at 7 metre intervals. The traps were physical multi-catch traps that were checked every six hours. Each primary occasion consisted of between three and five secondary occasions, each secondary occasion ending after six hours. Over the entire survey, 158 unique individuals were captured (75% being female). As noted by Ergon and Gardner (2014), re-captures occurred more frequently in secondary occasions that spanned the morning/afternoon hours compared to those that spanned the evening.

The aim is to fit a CJS model to this data set, conditioning on the initial capture of each of the 158 encountered individuals. For this model, a Gaussian detection function with intercept g_0 and range s is assumed. Between two primary occasions that occur Δt apart, voles are assumed to disperse according to a Gaussian distribution with mean zero and variance $\sigma^2\Delta t$; hence, the model can account for the reality that individuals may disperse further over long time periods. Between primary occasions, voles may also die with probability $1-\phi$. Separate parameters were estimated for each sex. Also, due to the apparent difference in encounter rate during the day compared to the evening, the detection parameters g_0, σ also depended on the time of day each secondary occasion took place.

The model was fit by maximum likelihood and by the existing Bayesian method for comparison (Table 5.3). The Bayesian analysis was carried out using the JAGS code supplied by Ergon and Gardner (2014) to obtain a total of 15000 posterior samples.

Inference from both methods is similar. There is evidence that males have a larger activity range within primary occasions compared to females. Also, there is weak evidence that females have a lower survival rate compared to males.

		MLE	Bayes
g_0	F	0.30 (0.26, 0.34)	0.31 (0.27, 0.34)
	M	0.26 (0.21, 0.33)	0.27 (0.22, 0.32)
s	F	4.97 (4.88, 5.07)	5.42 (5.18, 5.69)
	M	7.39 (6.91, 7.90)	7.26 (6.81, 7.76)
σ	F	3.47 (2.84, 4.21)	2.58 (1.91, 3.34)
	M	4.56 (3.31, 6.27)	4.50 (2.98, 5.97)
ϕ	F	0.75 (0.67, 0.81)	0.74 (0.67, 0.81)
	M	0.92 (0.80, 0.97)	0.89 (0.80, 0.96)

Table 5.3 *Maximum likelihood (MLE) and Bayesian posterior (Bayes) estimates for males (M) and females (F) of average daily detection function intercept g_0, daily activity range s (metres), monthly dispersal range σ (metres), and monthly survival probability ϕ with associated 95% confidence and credible intervals*

Overall, the maximum likelihood approach provides a fast alternative to the Bayesian methods. It is an effective tool for separating the commonly confounded processes of survival and emigration. This could have substantial impact on population dynamics models and integrated population models that rely on accurate survival estimates to obtain long-term projections of population stability and carrying capacity. Extending the model to allow for spatially-varying population demographics to identify areas of high mortality or migration is also possible. The maximum likelihood approach is likely to be a preferable candidate for such extensions as the computational cost will be much lower than in the Bayesian

framework.

Other than a tool for achieving true survival estimation, inference on dispersal could also be made; however, the inference obtained is likely to be closely dependent on the survey design. Only dispersal within the detector array can be observed and so any movements outside the area covered by the detectors is determined by the model chosen for dispersal. For surveys were detectors cover a larger area than individuals commonly disperse over, estimates of dispersal can reflect the true movement process; for surveys that cover a smaller area, only smaller dispersals are observed and thus dispersal will be underestimated. This will be further exacerbated if the true dispersal process is a mixture of small and large dispersal as is likely to be common in many species where individuals roam a small territory before relocating in a single bound to another in search of better prospects. There is no information in the observed captures to select one model for long-range dispersal over another unless the survey is designed so as to detect these longer dispersals. For the vole example, the detectors considered comprised only one array of a larger survey and it was observed that voles can disperse much larger distances than estimated. Hence, inference on dispersal from such models is likely only possible when the survey was designed with this purpose in mind.

5.7 Discussion

In this chapter, the theoretical and computational advances required to fit SCR models with mobile activity centres by maximum likelihood was presented. This model was then applied to three related problems that arise in the application of SCR due to animal movement. It is anticipated that these contributions will meet the need to relax the assumptions of SCR due to the realities of wild animal surveys.

5.7.1 Contribution

The contribution made by the work in this chapter is to develop analogous maximum likelihood methods to existing Bayesian models and to the extension of these models. The contributions can be described by the three contexts in which these methods would apply:

1. **Account for uneven use of activity areas**: the methods to fit by maximum likelihood a model where individuals can have a varying activity centre allows for the reality that individuals may explore a larger area across the entire survey compared to within each occasion. This method can readily be incorporated into the existing maximum likelihood software (Efford, 2017) for SCR models and provide a convenient

option for practitioners that are aware of this problem in their surveys or wish to test whether the observed captures indicate uneven space use. Obtaining activity area estimates that can be related to the movement of an individual is unimportant when SCR is solely for estimating density. Applying methods to improve the accuracy of activity range estimation is required when one wishes to relate biological concepts, such as home-range, to these parameters. SCR is increasingly being used to study not only a population's size but its territorial composition and the dependence between density and home range (Efford, Dawson, Jhala, & Qureshi, 2016). For surveys where occasions span short times compared to the entire survey time, uneven space use is likely and failing to account for it will bias any inferences made beyond population size.

2. **Account for transience within the population and estimate the proportion of transients within a population**: Bayesian methods have been developed to allow for complete transience within the population where activity centres are mobile. In this chapter, this was extended to allow for the population to be composed of both resident and transient individuals and for the proportion of transients to be estimated. Over short surveys, the presence of a small number of transients can cause disproportionate bias in the estimation of activity range and encounter rate. Over longer surveys, a large proportion of the population may be mobile and, by assuming them to be resident, lead to exaggerated activity range estimates. This is particularly relevant to the growing implementation of surveys over large national parks, reserves, and countries. Capture of individuals by detectors separated by large distances can lead to unrealistic estimates of ranges, yet the main objective of linking such large scale studies together is to obtain an idea of precisely this. Further, it is becoming clear that individual heterogeneity in transience is important. Many regions are composed of residential and transient individuals; complete intransience and complete transience models both lead to less meaningful inference. In some cases, this heterogeneity can be described by recording individual-level covariates such as sex or age. In others, the source of heterogeneity is unknown and so building models that can split the population into latent groups that behave differently is essential. The method developed in this chapter is a first step toward this.

3. **Account for dispersal in open populations**: a maximum likelihood alternative was developed to the existing Bayesian implementation of the Cormack-Jolly-Seber SCR model with mobile activity centres. The extension to the Jolly-Seber model is

also possible. It allows for population demographics to be separated into recruitment, survival, and migration without the common confounding of the three. As SCR surveys are supported over longer time periods and are repeated in consecutive years, open population modelling that can draw the most informative inference will be crucial. A drawback of existing Bayesian methods is their computational burden when applied to large data sets, particularly those that include a large number of unique individuals or a large number of occasions. Long-term SCR surveys are likely to contain both. Maximum likelihood methods also suffer from an increase in computation time and computer memory burden; yet, the relative increase in these demands is modest compared to the Bayesian approach. Additionally, it is common that an increase in data leads to a desire to fit more complex models; the maximum likelihood approach could provide a more computationally efficient avenue. Arguably, the Bayesian approach is considered more flexible when complex extensions are required. It is likely that in the future, a hybrid of the two will prove the better choice: the numerical methods to compute the marginal likelihood combined with an MCMC algorithm.

5.7.2 Limitations

Limitations of the methods to each application have been discussed. Of particular note are the identifiability and convergence issues within the partial transience model, and the dependence between inference on movement and the survey design.

A further, key limitation of the maximum likelihood approach is that the methods require a Gaussian distribution be assumed for the movement of activity centres. The computational methods that make model fitting feasible rely on this assumption. Ergon and Gardner (2014) considered models for activity centre movement including an Exponential, Gamma, and zero-inflated Gamma. They found that inference on movement can be model-dependent. This could not be explored here due to the limitation that no other movement model could be considered. Given the evidence that inference may not be robust to model choice, the ability to fit different movement models ought to be the focus of future research. The obstacle is the development of computational methods as in Section 2.5 for other forms of Lévy processes. This is discussed further in Chapter 6, in the general context of encounter models.

5.7.3 Extensions

There are many possible extensions of this work that can be driven by their biological applications. Three that are anticipated to be of general importance:

1. **Activity meta-centres**: within the Bayesian approach, instead of activity centres moving over time, they can be randomly distributed around a central point, known as an activity meta-centre. The idea is that the individual has a long term activity area with this meta-centre, but unevenly distributes itself around this area leading to within occasion activity centres that do not coincide with the meta-centre. A similar model could be fit by maximum likelihood using numerical integration.

   There has been little research into models where activity centres are correlated across occasion and distributed around a meta-centre. This is a description of a biased random walk where each movement between occasions is composed of a random step and an attraction toward the meta-centre. In continuous-time, this would be described by an Ornstein-Uhlenbeck process. As in the previous chapter, a hybrid method could be used where meta-centres are simulated within an MCMC algorithm, as activity centres are in the Bayesian approach for the standard SCR methods, then the numerical methods from this chapter can be used to integrate over all activity centre movements around this meta-centre.

   The motivation for this development is to provide richer inference on short and long term activity areas. The methods in Section 5.4 allow activity centres to move to obtain unbiased inference on the within-occasion range of each individual. Yet, over a sufficiently long time, individuals, under this model, would explore the entire survey region. In reality, their movement is constrained within their complete range.

2. **Spatial models**: as mentioned in the previous sections, the maximum likelihood approach lends itself well to the extension of spatial modelling. Uneven space use models where the movement of activity centres depends on spatial covariates is similar to the modelling of resource selection functions: individuals may use certain parts of their activity range more often than others based on what resources and choice they have within their range. For the partial transience models, the movement rate may depend on resource availability and habitat suitability such that individuals move through certain areas more quickly than others. Finally, for dispersal models, survival, recruitment, and dispersal can vary with location and the identification of hotspots or sinks is the future focus of open population models just as the classical interest has

been on temporal variation.

3. **Multi-state models**: multi-state capture-recapture models have long existed (Pledger, 2000): individual heterogeneity is known to cause bias and thus accounting for unobserved heterogeneity is a well researched problem. Unobserved heterogeneity is variation in individuals that is not explained by observed covariates. Conceptually, one can imagine the population being divided into clusters with theoretically homogeneous characteristics of individuals within clusters. The partial transience model is an example of a multi-state model with two states: resident and transient. All multi-state models are mixture models. Uneven space use, transience, or dispersal may vary between latent sub-strata of a population. Describing this can reduce uncertainty, but can also be interesting in itself. Just as detection, density, and survival parameters summarise a single property of the population, multi-state models cluster individuals according to their capture histories providing a data-driven demographic.

   As SCR surveys provide greater information on individual movement and space use, multi-state models can bring a less parametric approach to gaining insight into the members of these populations. This can be extended further to allow for individuals to switch clusters over time by a Markov process. One can imagine the life history of each individual being described by a chain of states and inferring the individual's life states from their capture history. This can bring together biological theory which would lead to construction of life history models and encounter surveys where such models can be fit to individual-level data. Such models can be used to describe a wide variety of natural processes: the dispersal of juvenile individuals from their parents, the increasing transience of males during breeding seasons, or the density-dependent competition that leads to greater mortality and migration. The methods from Chapter 2 provide the theoretical and computational framework to apply these models in future research.

Chapter 6

Discussion

In this chapter, a summary of the thesis is given and possible future research proposed. This discussion considers the encounter model approach generally. Specific discussion of the limitations and possible extensions of the method to distance sampling and spatial capture-recapture in particular are given in Chapters 3–5.

6.1 Summary

6.1.1 Encounter Model

The encounter model is the synthesis of the detection models commonly used in distance sampling and SCR with existing continuous-time animal movement models. The detection process is modelled as a Poisson point process along each animal's path where points correspond to encounters. This formulation is implicitly assumed in SCR models and is a natural deduction from the search process models in distance sampling, but here the model is made explicit and begins with the idea of an animal moving along a path. Incorporating a latent path for each individual is a novel approach to the analysis of distance sampling or SCR encounters.

The likelihood is stated using the concept of a path integral, shown to be central to the idea of marginalising over all unknown animal movement paths. A key contribution of this thesis is the algorithm presented to approximate this integral by quadrature. The integral is approximated by the likelihood of a hidden Markov model by discretising space and time. This is an approach that has been taken in previous research (Pedersen et al., 2011). The obstacle to the use of this approximation is the computational burden. Here, the sparsity of the matrices involved and the Krylov dimension-reduction technique are used

to reduce this burden; furthermore, the discrete Fourier transform is adopted for matrices of a particular symmetry (Toeplitz) to further improve performance. These adjustments were not made to make an adequate solution faster, but to make an infeasible one feasible. Without these theoretical and computational savings, application of the encounter model to distance sampling and SCR would be more difficult: in both surveys, detections occur on a small scale compared to the distances animals can travel, requiring spatial resolution to be high and the space itself to be large; thus computational burden is often high.

Hidden Markov models are a popular tool across many fields and in varied problems; it is anticipated that this work could have applications beyond the encounter model, in cases where high-dimensional hidden state-spaces are used. Furthermore, path integrals arise in many contexts from animal movement modelling, to the analysis of financial time series, and to the reaction-diffusion models found in mathematical biology. For these applications, a fully Bayesian approach has been the norm; path integration by quadrature could provide maximum likelihood alternatives or be used to improve the performance of existing MCMC algorithms by dividing the marginalisation over the model's latent structure between numerical integration and the Monte Carlo method.

6.1.2 Application to DS and SCR

This thesis focussed on distance sampling and spatial capture-recapture. They are the two most widely-used population abundance methods. Chapters 3–5 begin by briefly introducing the statistical theory behind distance sampling, continuous-time SCR, and discrete-time SCR. The encounter model approach is then described. This required some repetition, but served the purpose of showing that the abstract approach taken has advanced research in three related contexts that are often considered separately. This follows the current trend in the field to unify these approaches (Borchers & Marques, 2017). This is driven not only by theoretical aesthetics and efficiency, but by the recognition that the surveys conducted on animal populations are no longer neatly categorised as distance sampling or SCR or animal telemetry. Surveys can include static microphones or cameras, additional line transects, area searches, GPS tag data, and more. A fundamental unit common to most of these surveys is the encounter between animal and detector.

For distance sampling, incorporating animal movement has been a long standing problem (Buckland et al., 2015). The effect it has on abundance estimation is not well understood and highly sensitive to the context of each survey. A simulation study shows that bias in abundance estimation due to conventional distance sampling can be substantial when

animals move. In many surveys, it may not have as substantial an effect on abundance estimation. When it does, adopting survey protocol that mitigates this effect would be the best solution (Glennie et al., 2015). When this does not suffice, this thesis provides a model-based approach to avoid the bias caused in the estimated detection function. First, this will allow practitioners to use telemetry data to retrospectively assess whether animal movement causes practically significant bias in density estimation as shown in the spotted dolphin case study. Second, it allows practitioners to apply distance sampling to animal populations that have hitherto been unamenable to the assumption that animals do not move, as in the application to seabird surveys. Chapter 3 further shows that incorporating movement does not exclude the existing extensions of distance sampling such as multiple covariates, multiple observers, and availability processes. A single important issue that is not addressed in this thesis is how to deal with animal groups. At present, groups are treated as individuals; this is undesirable for two reasons: groups split up and fuse together over time, and groups do not move as individuals do, yet using individual telemetry to inform the movement model for groups makes this assumption. The current recommendation is to use groups as the unit that is encountered and then use mean group size to convert the estimate of group density to population density; this is a standard approach in conventional distance sampling. It is, however, more questionable when the survey is not a snapshot in time. This is an important avenue for future work.

For SCR, incorporating animal movement can reduce bias in density and activity range, but, for future development, the main contribution is that now individual animal movements can be explicitly modelled. In continuous-time SCR surveys, each captured individual's movement around their activity centre can be inferred. The importance of this is shown in the jaguar case study. An improved estimate of each jaguar's activity range, bringing together telemetry and SCR data, is a key quantity when assessing anthropogenic effects such as deforestation. These effects are heterogeneous in time and space, and difficult to assess by telemetry alone. SCR surveys cover more of the habitat and sample a larger proportion of the population than most telemetry studies do and at less cost. Telemetry studies do, however, provide a spatio-temporal description of where animals spend their time. SCR research has moved toward doing the same by allowing activity range to change over space and for encounter rate to depend on ecological distance, rather than Euclidean distance (Sutherland et al., 2015). This thesis provides the statistical tools to make inference on the spatio-temporal distribution of individuals from continuous-time SCR data directly, using a more realistic and more intuitive explicit animal movement model. This movement can depend on environmental covariates and landscape connectivity. Furthermore, the thesis

introduces how to investigate the interactions between individuals, which has not yet been considered in the SCR literature.

As microphone and camera-trap surveys gain further popularity, continuous-time SCR will become a standard statistical approach to the analysis of such data. Looking at these data, practitioners have begun to see that inference can be made not only about population density, but animal activity, movement, and interaction. These questions are routinely discussed in the movement ecology literature (e.g Hooten et al. (2016); Nathan et al. (2008); Ovaskainen et al. (2008); Patterson et al. (2017)), thus a model incorporating animal movement with continuous-time SCR is anticipated to become a natural tool for these practitioners.

The final application of the encounter model was to discrete-time SCR. The encounter model was adapted to the discrete-time setting to allow each individual's activity centre, rather than their continuous location, to move from one discrete occasion to the next. This is shown to be relevant to three related situations: surveys where individuals use their activity range unevenly, surveys where part or all of the population are transient, and long-term open population surveys where individuals disperse over space. These three problems have been considered within the SCR literature (Royle et al., 2016), but not explicitly discussed as demanding the same theoretical extension. This extension has been developed in a Bayesian framework, where the path integral is computed by Monte Carlo methods. In Chapter 5, the quadrature approach is used to fit the same models, to the same data sets, by maximum likelihood. First, this introduces the maximum likelihood alternatives to the Bayesian approaches. Second, as the inference from these approaches is qualitatively similar, it shows that the quadrature approach to path integration can be as accurate as a Monte Carlo method. Developing maximum likelihood alternatives and quadrature methods is not due to a philosophy against simulation, but to demonstrate that, in some cases, these methods can be substantially more efficient without a cost in quality. On the contrary, the hope is that the quadrature and simulation based methods can be combined to form a superior algorithm.

Overall, the thesis takes a step toward incorporating animal movement and removing the assumption of immobility central to existing methods in DS and SCR. The general approach leaves it open to many possible avenues of future work, motivated by the limitations of the work herein and by the research questions posed by ecologists.

In the remainder of this chapter, some possible ventures for future research are discussed.

6.2 Encounter Model

The encounter model is made from two parts: the detection model and the movement model.

6.2.1 Detection

Two particular future extensions of the detection model can be considered.

1. **Noisy information on location**. Animals move in paths and observations are made whenever an encounter occurs. These observations provide information on animal location. In this thesis, this observation is either the detector the animal was seen by (in SCR surveys) or the exact location measured (in DS surveys). This is a limited view: any observations that are informative about an animal's location can be used.

   In Chapter 3, the measurement error models used in distance sampling were shown to be easily included within the encounter model framework. By a similar approach, any observation $\boldsymbol{y}$ made during an encounter can arise according to a distribution whose parameters depend on the location of the individual $\boldsymbol{x}$. For the measurement error model, this was a bivariate Gaussian:

   $$\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{x}, \sigma_\epsilon^2 \boldsymbol{I})$$

   for some error σ_ϵ^2 and identity matrix $\boldsymbol{I}$. In practice, any statistical model that relates the observation to the spatial location of the individual can be incorporated. The computational algorithms developed are not affected. The detection model presented is already conceived with a spatially varying encounter intensity, $\lambda(\boldsymbol{x})$.

   A particular application of this will be in acoustic surveys. The received signal strength of a sound is informative of the source's location and an acoustic model can be used to deterministically or stochastically quantify this relationship. Furthermore, the propagation of the sound and the directionality can be incorporated. The soundscape of a study region can mean detectability is not a function of Euclidean distance but varies heterogeneously over space. As animals move around, encounters with these animals will depend on how sound propagates between the animal and the detector. At the extreme, it is possible for shadow zones, areas where no sound can propagate between the animal and detector, to exist and disrupt encounters. This may be a particular issue for acoustic DS surveys: animals initially located in a shadow

zone that move into an area of improved acoustic propagation is equivalent to animals outside a transect moving into the transect, and thus may be substantially larger than it would be if acoustic propagation were homogeneous.

The use of acoustic devices for DS is ever increasing. Slow moving surface and diving vehicles are being used to survey for cetaceans across large distances (Klinck et al., 2012). These vehicles move slowly compared to the species they search for and thus bias from animal movement is likely to be a substantial issue. For acoustic devices fixed to the sea floor, animal movement is undoubtedly an issue when snapshot methods are not used. Incorporating movement is one possible way to adapt distance sampling methods for these new technologies. Cue counting and snapshot methods (Howe et al., 2017) are existing alternatives. The former is a promising approach to acoustic surveys, however, for some species, such as cetaceans, that can produce millions of cues, practitioners may desire to return to a method that works with individuals. For the latter, practitioners may be uncomfortable with the subjective determination of the length and periodicity of snapshots.

Another use of acoustics in DS surveys is when ships tow microphone arrays. These arrays often detect the bearing, with some uncertainty, of an individual multiple times before the ship is close enough to fix the individual's exact location. At present, these encounters before the exact location are recorded are not used in any statistical analysis. This is an example of a multiple encounter survey. Extending the encounter model to incorporate the partial information that bearings give on animal location will allow for the movement of the individual to be accounted for whilst the ship determines its location, and can provide *in situ* information on how animals move during the survey. Incorporating a coefficient for responsive behaviour, as described in Section 3.10, would also be possible as these repeated encounters can be used to estimate if animals respond as the ship nears.

Methods for acoustic SCR are also a growing research interest. The detection of a sound, a cue, across multiple detectors can be used to estimate the source of the sound, the activity centre. In other words, each sound has a capture history across space rather than time. Current methods make the assumption that sounds are independent and rely upon estimates of cue production rate to convert estimates of cue density to individual density. Royle et al. (2016) and Borchers et al. (2014) show by simulation that if detections are highly correlated in space or time, then estimates of density can be biased. The required correlation for this to occur is high and so in many SCR surveys it is not a concern. For acoustic surveys of highly vocal individuals, this can be

an important issue. For cetaceans, a sustained bout of cues from a single individual is regular behaviour and thus detected cues will be spatially and temporally correlated. This leads to bias in estimated cue density, and thus estimated individual density. The encounter model, in its present form, cannot be used directly to solve this issue as it is unknown which individuals produce which cues. One research approach would be to use the trace-contrast method of Fewster, Stevenson, Borchers, et al. (2016) to associate cues that arise from the same individual.

2. **Clustered encounters with an individual**. The encounter process is conceived as an inhomogeneous Poisson point process along each animal's movement path. Encounters are independent given the intensity of this point process. Thus, including temporal and spatial covariates into the intensity can account for correlation in encounters. The encounter residuals provide a tool to assess whether this is achieved sufficiently. Furthermore, standard tools from point process modelling, such as Ripley's K-function or the pair correlation function, can be used to assess whether the observed encounters meet the pattern expected under the estimated inhomogeneous Poisson process. If encounters are clustered due to an unobserved latent process, then an extension of the current model is called for that can allow for encounters to occur according to a clustered point process such as a Cox process (Diggle, 2013).

   An alternative, and easier, solution is to consider whether clustered encounters can be described as a single encounter. This is somewhat similar to the argument of whether each recorded cue in an acoustic whale encounter is a new encounter or whether cues ought to be grouped as a single encounter. Given the assumptions of the encounter model, the latter may be a better approach.

6.2.2 Movement

Chapter 2 focussed on the class of advection-diffusion movement models. In theory, any movement model that defines a probability measure over the space of continuous paths can be used. In particular, correlated velocity continuous-time movement models are likely to give a better description of animal movement (Gurarie et al., 2017; Hooten, Johnson, et al., 2017). Animals are known to move in persistent directions. This may be particularly important for DS surveys since the bias from animal movement is known to be worse when animals move more persistently (Glennie et al., 2015).

These models, however, cannot be easily implemented due to the computational burden (Pedersen et al., 2011). The quadrature approach discretises space into grid cells and

animals are proposed to move on this grid according to a Markov process. Movement across this grid will not be a Markov process if velocity is correlated; however, the joint process of the velocity of the individual and their location is a Markov process. The quadrature approach would therefore proceed by discretising not only space but also the space of possible two-dimensional velocities the individual may have. Given the work required to make the calculation over a two-dimensional grid feasible, it is likely further work will be needed to extend this to quadrature over a four-dimensional space.

There is one compromised approach to this problem. The current framework can implement an advection when this advection is known. For the jaguar case study, advection described the biased movement of each jaguar around its activity centre. These centres were unknown. To perform this four-dimensional integration, a data augmentation scheme was used: the activity centre was simulated within a MCMC chain, once taken as known, the marginal likelihood, averaging over all the paths around this activity centre, was computed by quadrature. The same approach can be taken here. A MCMC algorithm, or particle filter, can be used to simulate the velocity of an individual over time and then, including this known velocity as an advection term, the marginal likelihood can again be computed by quadrature. This would allow for correlated velocity, but with a substantial drawback: the velocity process cannot depend on spatial covariates. The idea is that the velocity process is simulated separately from the integration over all paths given this velocity process, thus whilst simulating the velocity process, the location of the animal is unknown. This hybrid approach is, however, more computationally tenable and a good first step. The ultimate goal would be to marginalise over the four-dimensional space entirely.

Another limitation of the movement model is that the same movement parameters are assumed to describe the movement of every individual. Telemetry studies clearly show that animals of the same species, in the same area, and at the same time move differently. Even when working toward the same goal, animals move according to different strategies and have different energy budget constraints.

When animal movement is a nuisance process, it is less critical that these individual characteristics are captured by the model. Similar to the concept of pooling robustness, found in distance sampling, if the population is made up of subsets with disparate movement characteristics, this is unlikely to affect estimates of abundance. It is more important to that the sample of individuals be representative and the estimate population mean be unbiased. For distance sampling, pooling robustness arises because even though some individuals may be more or less detectable than others, on average across all individuals, the collected data can be used to estimate the population mean detectability and so estimates of abundance

remain unbiased. Similarly, if individuals move at a higher or lower rate than others, if the estimated mean movement rate is unbiased, then the estimated population-level detection probability will be unbiased and abundance estimation be robust.

Concern for individual characteristics is more important when inference on animal movement is made. For example, in Chapter 4, the encounter model is applied to a case study of jaguars. From their encounter history, the realised activity range of each jaguar is estimated as well as the spatio-temporal overlap in these ranges. This is all based on the assumption that all the jaguars move according to the same movement process with the same movement parameters. This is not likely to be the case for several reasons: jaguars will behave differently and have different priorities dependent on their life stage; also, the environmental make-up of each jaguar's range, not described by observed covariates, may influence the movement of each jaguar. Failure to account for this individual variation will lead to ranges being pulled toward the mean, shorter ranges overestimated and larger ranges underestimated. This undermines the goal of individual-based inference.

An approach to accounting for individual variation would be to include random effects (Hooten et al., 2016): allowing the movement model to adjust to the characteristics of each individual. Alternatively, a mixture model could be used where individuals may belong to different latent groups that each have different movement characteristics (e.g. McKellar, Langrock, Walters, and Kesler (2014)). This is similar to the partial transience model in Chapter 5.

Allowing for individual heterogeneity has been a central research objective for capture-recapture due to the bias it causes in abundance estimation; for encounter models, the objective is renewed due to the desire to obtain inference on an individual level.

6.2.3 Uncertain Identity

For standard SCR, individuals are always assumed to be identifiable. Despite this, there is often substantial uncertainty in individual identification, for example, in big cat camera trap surveys, it may only be possible to discern the identity of an individual when they are photographed from a particular side, as markings will differ from one flank to another. SCR methods to deal with uncertain ID have been developed (Augustine et al., 2018). These methods use known identifications to estimate activity range and then, based on this, can infer which uncertain IDs belong to which individuals. This has not been implemented in a maximum likelihood approach, to do so would involve the computation of a sum over all possible identifications. The encounter model can be extended within the Bayesian

approach: identifications are assigned within a MCMC algorithm and then, given these identifications, the likelihood of the encounter model can be computed. This is an example of the hybrid approach where the latency of the model, identifications and movement paths, is divided between numerical and simulation based methods.

For distance sampling, even though individuals need not have identifying marks, individuals are uniquely identified within each transect by their immobile position. By incorporating animal movement, this theoretical justification for not recording the same animal twice in one transect is removed. The distance sampling with animal movement model in Chapter 3 makes the assumption that the first encounter with each animal is recorded and no subsequent encounters are mistaken as first encounters with a new animal. In some surveys, this assumption will be tenable. In surveys where observers are swamped with encounters, it is unavoidable that multiple encounters with an individual may be recorded as encounters with multiple individuals. This would lead to bias in conventional distance sampling methods as well as with the encounter model approach. It is important to stress that incorporating animal movement mitigates the bias in the estimated detection function not the bias due to "double counting." The challenge is that animals are unmarked and so multiple encounters with an individual cannot be recognised as such. To account for this would again require one to sum over all possible identifications of all encounters made. No doubt this sum can be computed more efficiently by matching encounters that happen close to each other, but the computation will likely still be a burden. A simulation-based approach for multiple observer methods, where multiple sightings are classified as from the same or distinct individuals, has been proposed (Hamilton et al., 2018). A similar extension could be applied here. As distance sampling surveys are increasingly carried out by technology such as cameras or slow moving vehicles, multiple encounters with the same individual that cannot be recognised will become an important issue.

6.3 Path Integration

Path integration is a big obstacle to the application of the encounter model. In this section, the possible future development of the work begun in this thesis is discussed.

6.3.1 Markov-Separable

To make the computations feasible, one is forced to make limiting assumptions. One such assumption has already been discussed: the restriction to the class of advection-diffusion

models. Another is the restriction to integrands that are Markov-separable.

For DS and SCR, this restriction was not critical. The likelihood of both models required only the integration of Markov-separable functions. In both cases, this can intuitively be understood to derive from the assumption that the probability of being detected in any given interval of time depends only on the individual's movements within that interval of time. The detection process is memoryless.

One may question why the discrete-time SCR model developed in Chapter 5 considered activity centre movement rather than the continuous paths travelled by each animal during the survey. The latter would have been a better application of the general encounter model. Consider an animal that moves along a path $\vec{\boldsymbol{x}}$ during a discrete-occasion SCR survey. Suppose this animal is detected m times at a given detector during this occasion. By Definition 2.5, the likelihood contribution of this observation is given by a Poisson distribution with an intensity $\Lambda(\vec{\boldsymbol{x}})$ that depends on the entire path travelled by the animal:

$$[m \mid \vec{\boldsymbol{x}}] = \frac{\Lambda(\vec{\boldsymbol{x}})^m \exp(-\Lambda(\vec{\boldsymbol{x}}))}{m!}$$

This is the only observation made for the detector, the exact detection times are unknown. Consider the integral of the above quantity over all possible paths. This integrand is not Markov separable and thus the HMM approximation cannot be used.

The integrand can be written as the sum of Markov-separable functions but this sum is over all possible times the m encounters could have occurred. This is computationally intractable. It is the continuous version of the problem that arises from uncertain identity: one must consider all possible assignments of the observed encounters to the unknown number of observed individuals. Here, one must sum over all the possible times that can be assigned to the m observed encounters.

Dependence between detections in discrete-time SCR surveys was accounted for in Chapter 5 by allowing activity centres to move. To do the same by averaging over all continuous paths travelled by an animal, the Markov-separability assumption must be removed or the integrand approximated by one or a sum of Markov-separable functions.

6.3.2 Solving PDEs

Performing an integration is equivalent to solving a differential equation. HMM quadrature (Theorem 2.7) is similar to the finite difference methods commonly used to solve partial differential equations (PDEs). Theorem 2.6 shows that the probability density function

$\phi(\boldsymbol{x}, t)$ for the random variable $\vec{\boldsymbol{x}}(t)$, the position of the animal at time t, changes over time and space according to an Eulerian partial differential equation.

Finite difference PDE solutions begin by discretising the space that each variable varies over. An initial value for the function of interest is specified for each grid cell, giving an initial vector $\boldsymbol{\gamma}_0$. The aim is compute $\boldsymbol{\gamma}_t$, the discrete representation of $\phi(\boldsymbol{x}, t)$ for some later time t. To do so, the PDE is used to *update* the value of the function at each grid point over some small time-step, Δt. This is repeated until the time t is reached. In some cases, it may not be necessary to discretise with respect to all variables: one can discretise with respect to one variable and then exactly or approximately solve the subsequent PDE analytically. The advantage of this is that when updating through time, some of the variables can be updated exactly whilst those that are discretised will be updated with error. This is known as the method of lines. A comprehensive introduction to the PDE methods used in this thesis and possible extensions are given by LeVeque (2007).

In this thesis, a spatial grid is placed over the study area. A point process model or design-based assumption is used to quantify the probability an animal begun the survey in each grid cell. The advection-diffusion PDE is then used to update the animal's location over time. After each update, the probability of an animal being encountered or not can be computed from the current values associated with each grid point. The updating scheme was described in terms of a hidden Markov model. Alternatively, the updating scheme can be described as an approximate method of lines. Space is discretised, but time is not. Updates through time are approximated by a first order exponential integrator. In other words, once space is discretised, the remaining PDE can be approximated by a first-order, matrix ordinary differential equation of the form:

$$\frac{\mathrm{d}\boldsymbol{\gamma}_t}{\mathrm{d}t} = \boldsymbol{\gamma}_t \boldsymbol{R}_t.$$

The matrix $\boldsymbol{R}_t$ is the transition rate matrix as defined in Section 2.5.2. This is the same equation that arises when describing the state-switching in a continuous-time Markov chain. That is the reason why the tools used for hidden Markov modelling can be used as an updating scheme.

There is a vast literature on the numerical solution of PDEs. Forging more links between the approach taken in this thesis and alternative, or improved, methods from applied mathematics would likely be fruitful. There are three areas of research into PDEs that are of particular relevance.

- **Improved grid methods**:

In this thesis, the HMM quadrature method requires the grid to be regularly spaced and rectangular. Adaptive and multi-grid methods have been used when solving PDEs. The grid can be adapted as more information is gained about the function being approximated. This may be relevant in spatial capture-recapture surveys, in particular. For the jaguar case study, jaguars clearly had a limited activity range; therefore, it is wasteful to update the value of the PDF evenly across the whole study area when the probability of a particular individual being in some areas is effectively zero. A better scheme would have an uneven grid with poor resolution in areas where the PDF is likely to be zero and higher resolution within each jaguar's range. This could either reduce computational burden or provide better inference on a higher spatial resolution. Similar to this, multi-grid methods could be used where the PDF is updated on a coarse grid across the entire study area, to determine where animals are likely to be, and then a finer grid laid across only those areas of interest. If encounter models are applied to study areas that are substantially larger than the range of the study animals, these extensions would be a necessity.

- **Split-step method**:

  The advection-diffusion equation is an attractive class of movement models because it is intuitive; it captures two primary forces in an animal's movement: the will or motivation to move in a certain direction and the inherent randomness when doing so. In Section 2.5.5, the discrete Fourier transform (DFT) was shown to be an efficient tool in the updating step. It relied upon the matrix $\boldsymbol{R}_t$ being Toeplitz. This has the unfortunate restriction that the DFT could only be used for state-switching diffusive movement. Any advection would destroy the symmetry of the matrix and preclude the use of the DFT. Matrix sparsity and the Krylov approximation lessen the computational burden nonetheless, but the DFT could dramatically reduce computation time and computer memory requirements.

  Split-step methods are based on the idea of splitting the updating scheme into two parts: an advection part and a diffusion part. Notice, in the advection-diffusion equations, if the diffusion term is removed, the resulting PDE can be solved analytically. The proposed approach would be to update the PDF in two steps:

  1. solve the PDE analytically ignoring the diffusion component. This would be equivalent to the action of moving each animal according to its advection but not diffusion;
  2. use the DFT to update the resultant PDF as if diffusion was the only movement

process.

This approach would extend the use of the DFT to advection-diffusion rather than only diffusion movement models.

A further motivation for this research would be to extend path integration to three dimensions (or higher). If used in three dimensions, the current method is unlikely to be feasible. To improve this, the generalised Fourier transform (GFT) can be used to exploit the symmetries found in three dimensions. The DFT improves computation in two-dimensions because it can diagonalise circulant matrices. Another way of saying this is that by assuming movement is invariant to two-dimensional rotations, the DFT can be used in the updating scheme. Two-dimensional rotations are only one kind of symmetry. The GFT is the equivalent updating scheme for movement that respects any kind of specified symmetry. In particular, three-dimensional diffusion is invariant to symmetries of the cube. Åhlander and Munthe-Kaas (2005) discuss the connection between mathematical group theory, the discrete Fourier transform, and partial differential equations. This connection can lead to path integration in higher dimensions.

- **Error and stability**:

  There is a large focus in the PDE literature on the error and stability of updating schemes. The magnitude of error in the use of hidden Markov models to approximate continuous-time processes is unknown. In practice, and in this thesis, it has been shown that the approximation produces inference similar to continuous Bayesian methods. This evidence is ultimately anecdotal and not rigorous. The error and stability of the finite difference method used is well documented. The error in the numerical integration over the spatial grid at each time-step is well known. The error in the method of lines approach for the particular class of advection-diffusion equations has been quantified. Nevertheless, the error in using all three algorithms together is unknown. At present, the spatial and temporal resolution of the approximation is increased until the value of the path integral is stable, that is, does not change by a specified heuristic. It is shown in Chapter 2 that the path integral converges to a value, but the rate of convergence is not quantified. Thus, theoretically, it is possible that even though the value of the path integral is stable, the discretisation is still not sufficient. Future research into either quantifying the exact error in the HMM approximation or providing an upper bound on this error is an important direction for future research.

6.3.3 Bayesian Approach

A Monte Carlo or Bayesian approach to path integration has not been considered in this thesis. In continuous-time, Blackwell et al. (2016) provide Bayesian MCMC samplers that integrate over all the possible paths between observed locations of individuals. Furthermore, in discrete-time, state-space models (Patterson et al., 2008) involve the integration over all possible paths an animal could have travelled; the path integration in that case is performed using a particle filtering method, such as the Kalman filter. In most cases, path integration in statistical ecology arises when there is a single latent path and multiple noisy observations of this path. These observations effectively collapse the probability measure over all paths to a subspace of most probable paths. MCMC algorithms or particle filters can be efficiently applied because of this reduction. The less information on which paths are most likely, the more paths the animal could have possibly travelled. MCMC and particle filters must then cover more ground in the path space; this may lead to higher computational burden and poor mixing of MCMC chains.

The motivation for a quadrature approach to path integration is when you want to average over all paths and have no or limited information on which paths are more likely compared to others. In DS and maximum likelihood SCR, a conditional likelihood is used where one conditions on the event that an animal is seen at some point during the survey. This involves computing the probability of any animal being detected at least once by integrating over all paths across the study area. There are no data to anchor an MCMC or particle filter around; there is no reduction in the volume of paths that need be considered. A quadrature approach regularly samples the entire path space. The cost of this is discretisation error.

Future research into a Bayesian implementation of the encounter model would lead to a clearer understanding of the computational difficulties. A particle filtering approach that samples the beginning of paths according to a regular grid could be an efficient compromise.

6.3.4 Hybrid Approach

King, McClintock, Kidney, Borchers, et al. (2016) introduced the concept of a semi-complete likelihood that can be used to combine the advantages of numerical and Monte Carlo integration. Abstractly, consider a model with a vector of latent variables, $\boldsymbol{\eta}$. For encounter models, the variables would be the paths of each individual and possibly the behaviour history of each individual if a state-switching encounter model were used. The idea is to separate these latent variables into two groups $\boldsymbol{\eta} = (\boldsymbol{\alpha}, \boldsymbol{\beta})$ where marginalisation over those in the first group, $\boldsymbol{\alpha}$, is achieved by numerical integration of the likelihood over these vari-

ables. This marginal likelihood is called the semi-complete likelihood. The second group, $\boldsymbol{\beta}$, contains those variables that are to be sampled within an MCMC algorithm whose acceptance probabilities are calculated using the semi-complete likelihood. The advantage of this approach is that the numerical integration reduces the dimension of the parameter space that must be explored by the MCMC algorithm. This cuts the computational cost, may improve convergence rate, and improve mixing.

In Chapter 4, this hybrid approach was taken in the jaguar case study. The latent variables included the paths travelled by each jaguar, these formed the first group $\boldsymbol{\alpha}$, and the activity centres of each jaguar, these formed $\boldsymbol{\beta}$. In this case, the numerical path integration reduced the space of latent variables from tens of thousands to fourteen.

Given the efficient computational algorithm developed in this thesis, the hybrid approach can be a good way to increase model complexity. The flexibility of the Bayesian approach can be leveraged with the numerical path integration reducing the computational cost of this flexibility. On the other hand, the likelihood-based methods in this thesis can benefit from a semi-complete likelihood approach, as in the jaguar study. This has already been proposed above when incorporating correlated velocity movement and for methods to account for uncertain identity.

Another motivation for a hybrid approach would be to allow animal-animal interactions as in Russell et al. (2016); Scharf et al. (2018). At present, path integration is performed for each individual independently. If individual movement paths were correlated, an adapted computational algorithm, with a much greater burden, would be required. This is discussed in Section 6.5. An alternative to an explicit model for how animals interact would be to induce correlation in the paths through a latent process. In this model, animals would not move in response to each other or to a group centroid, but would move according to a latent advection field that varies in time and space. Further, there can be multiple such fields and animals can have an affinity to follow one compared to another. This would produce animal group movement since members of the same group would follow the same latent advection. To fit such a model, one can divide the latent variables: paths, for known advection fields, can be integrated over by numerical integration, using HMM quadrature; the latent fields and group memberships can be updated by MCMC using the semi-complete likelihood. This is one possible way to explicitly model animal groups in DS surveys and account for animal-animal relationships in SCR surveys.

The hybrid approach to path integration is likely to become the preferred method. It avoids the need to simulate highly correlated latent variables, but retains the flexibility often restricted by a purely quadrature-based method.

6.4 Behaviour-Switching Continuous-Time Movement

The main focus of this thesis is on improving population abundance methods and not research on movement modelling. For this reason, a novel approach within this thesis to behaviour-switching continuous-time movement has not been discussed. The connection between path integration and movement models is noted in Section 2.4.5. In Section 3.7.1, a distance sampling model is developed for seabirds where birds can switch between resting on the sea surface and flight. Thus, this DS model has a latent behaviour-switching continuous-time movement model. To estimate the movement parameters, this model is fit to telemetry data by maximum likelihood. Fitting a behaviour-switching continuous-time movement model by maximum likelihood was presented by Pedersen et al. (2011) but was encumbered by a computational burden that made the method less attractive than the alternative discrete-time behaviour-switching models.

The path integration algorithm developed in this thesis makes this alternative considerably more feasible. Continuous-time movement models, as discussed in Chapter 1, can accommodate irregularly spaced telemetry observations and can be used to jointly analyse data collected at different temporal scales. Efficient algorithms to fit behaviour-switching continuous-time movement models are a primary obstacle to the use of continuous-time models. At present, they are not competitive alternatives to the discrete-time hidden Markov models. In future work, the path integration algorithm developed in this thesis can be used to tackle this challenge.

6.5 Animal Interactions

The term "encounter" as used throughout this thesis has always referred to encounters between animals and detectors. In general, the statistical theory can be applied to any system where target particles move around a space and probabilistically encounter objects. This clearly covers a wide range of possible applications. In this section, a few applications within statistical ecology are proposed.

The study of how moving animals interact is a relatively new area of research within movement ecology. Langrock et al. (2014) introduced a discrete-time model where each animal's movement is mixture of a biased random walk toward a group centroid and a random walk independent of this centroid. This captures the correlated movement of individuals whilst allowing for each to have a wavering affinity to the group. Continuous-time alternatives exist but suffer from the same burdens as previously discussed (Niu, Blackwell, & Skarin,

2016). Notably, models for individual-to-individual interaction are rare. Often a latent structure, such as a group centroid or advection field (Section 6.2.2), is used to induce correlation between individuals.

An aim of future work would be to extend the PDE approach to allow for individual interactions in a similar framework to Scharf et al. (2018). This would provide an efficient continuous-time alternative to the discrete-time group movement models. Additionally, it would allow for animal-animal interactions within encounter surveys; for example, in some SCR surveys, it is suspected that animals avoid each other, and this could be inferred from the observed encounter histories; however, no statistical methods exist to make such inference.

The PDE could be extended by including an advection between individuals. Let $\vec{\boldsymbol{x}}_i$ be the path travelled by individual $i = 1, \ldots, N$ and $\phi_i(\boldsymbol{x}, t)$ be the PDF of the random variable $\vec{\boldsymbol{x}}_i(t)$, the location of individual i at time t. An animal-animal interaction advection-diffusion PDE for individual i can be formulated as

$$\frac{\partial \phi_i}{\partial t} = -\sum_{j=1}^{N} \nabla \cdot \boldsymbol{\mu}_{i,j}(\boldsymbol{x}, t)\phi_j - \frac{\sigma(\boldsymbol{x}, t)^2}{2}\nabla^2\phi_i$$

where $\sigma(\boldsymbol{x}, t)$ is the diffusion rate at location $\boldsymbol{x}$ at time t. The first term contains the advection of the individual (when $j = i$) and the attractive, $\mu_{i,j} > 0$, or repulsive, $\mu_{i,j} < 0$, effect between individuals i and j. This interaction can vary spatially and temporally. Furthermore, if state-switching were included (Section 2.5.7), then individuals could influence each others state. For group movement models where individuals have two states, moving independently and moving in correlation with other individuals, allowing animals to influence each others state can capture the reality that some pairs of individuals are more likely to group together than others. Similarly, it can allow for the reality that the presence of one individual may cause the behaviour of another to change.

This model could be fit using the same HMM quadrature approach. The difference would be that the PDE for every animal would need be solved simultaneously, rather than independently, as the position of all animals at each time step will determine the movement of each individual. This will impact on computer memory requirements.

The proposed model allows for interactions between all individuals and for the strength of these interactions to vary pairwise. A more economical model could classify animals and then attempt to infer the interaction between animals in separate classes. The classical example of this is the predator-prey system where prey would avoid predator and predator

seek out prey. An SCR camera trap study could be devised to record encounters from both groups; their respective encounter histories may reflect the predator-prey interaction. Another example is the study of the relationship between conspecifics. For example, the jaguar camera study also provides records of pumas (*Puma concolor*) and there is interest in the investigation of their relationship (Harmsen et al., 2009). An inter-class interaction could be estimated from SCR observations, meaning statistical inferences can be used to provide evidence for or against biological hypotheses about animal interactions.

For many populations, animal-animal interactions drive the behaviour of the population as a whole. Furthermore, they are one of the forces that impacts on the entire community of populations within a study area. SCR surveys, especially those using continuously detecting technology, are likely to provide a wealth of information on a variety of species. Modelling how individuals interact will provide key insight into the behaviour of individuals and the behaviour of the population as a whole.

6.6 Concluding Remarks

The aim of this thesis was to take theoretical and computational steps toward incorporating animal movement models with distance sampling and spatial capture-recapture. The intention was to create a theoretical basis for future development in this direction, to overcome the computational obstacles that have prohibited this approach, and to demonstrate the benefits of this to distance sampling and spatial capture-recapture.

For statisticians, the statistical theory presented in Chapter 2 contains the key conceptual and computational elements to pursue further research. Furthermore, Chapters 3–5 illustrate how the existing methods of distance sampling and spatial capture-recapture can be seen in light of the encounter model approach. This allows researchers in both fields to link past work and current research to this framework. It is hoped that doing so will highlight the similarity in the approaches and focus research toward the common problems yet to be solved. In particular, the framework provides a rigorous way to recognise the role animal movement plays in population abundance surveys, evading the need to make an assumption of immobility.

For ecologists, this thesis demonstrates that incorporating animal movement leads to reduced bias in existing inferences made and the opportunity to make better use of the data collected. Chapters 3–5 show this by simulation studies and applications to real data. There are two motivations for practitioners to make use of the methods in this thesis. First, to account for violated assumptions. It is preferable that surveys be designed and conducted so

as to meet the assumptions of the standard methods. These methods are tried and tested, simpler to understand, and demand less time and computational effort. Yet, movement of wild animals can unavoidably lead to substantial violation of key assumptions. This thesis provides the additional model complexity required when this occurs. From this point of view, animal movement is a nuisance process that disrupts the inferences one wishes to draw. The second motivation is when interest lies in making inference about animal movement and habitat use. This can be the desire to derive information about movement from re-encounters during a spatial capture-recapture survey or to incorporate telemetry data to make joint inferences.

This thesis has focussed on key future research directions in distance sampling (Buckland et al., 2015) and spatial capture-recapture (Royle et al., 2017). As more animal populations are surveyed over longer periods and deeper research questions asked, the aim will be to further develop statistical models that rely on tenable assumptions and can provide sound, detailed, joint inference on the processes at work: detection, movement, behaviour, habitat use, survival, and more. This thesis is a contribution toward that goal.

References

Aars, J., Marques, T., Buckland, S., Andersen, M., Belikov, S., Boltunov, A., & Wiig, Ø. (2009). Estimating the barents sea polar bear subpopulation size. *Marine Mammal Science*, *25*(1), 35–52.

Åhlander, K., & Munthe-Kaas, H. (2005). Applications of the generalized fourier transform in numerical linear algebra. *BIT Numerical Mathematics*, *45*(4), 819–850.

Akaike, H. (1973). Information theory and an extension of maximum likelihood principle. In *Proc. 2nd int. symp. on information theory* (pp. 267–281).

Anderson, E., Bai, Z., Bischof, C., Blackford, L. S., Demmel, J., Dongarra, J., ... others (1999). *Lapack users' guide*. SIAM.

Au, D., & Perryman, W. (1982). Movement and speed of dolphin schools responding to an approaching ship. *Fishery bulletin*, *80*(2), 371–379.

Augustine, B. C., Royle, J. A., Kelly, M. J., Satter, C. B., Alonso, R. S., Boydston, E. E., & Crooks, K. R. (2018, 03). Spatial capture–recapture with partial identity: An application to camera traps. *The Annals of Applied Statistics*, *12*(1), 67–95.

Authier, M., Dorémus, G., Van Canneyt, O., Boubert, J.-J., Gautier, G., Doray, M., ... others (2017). Exploring change in the relative abundance of marine megafauna in the bay of biscay, 2004–2016. *Progress in Oceanography*.

Avigad, J., Dean, E., & Rute, J. (2011). Metastable convergence theorems. *arXiv preprint arXiv:1108.4400*.

Berg, H. C. (1993). *Random walks in biology*. Princeton University Press.

Besbeas, P., Freeman, S. N., Morgan, B. J., & Catchpole, E. A. (2002). Integrating mark–recapture–recovery and census data to estimate animal abundance and demographic parameters. *Biometrics*, *58*(3), 540–547.

Bird, T., Lyon, J., Nicol, S., McCarthy, M., & Barker, R. (2014). Estimating population size in the presence of temporary migration using a joint analysis of telemetry and capture–recapture data. *Methods in Ecology and Evolution*, *5*(7), 615–625.

Blackwell, P. G., Niu, M., Lambert, M. S., & LaPoint, S. D. (2016). Exact bayesian inference

for animal movement in continuous time. *Methods in Ecology and Evolution*, *7*(2), 184–195.

Borchers, D., Distiller, G., Foster, R., Harmsen, B., & Milazzo, L. (2014). Continuous-time spatially explicit capture–recapture models, with an application to a jaguar camera-trap survey. *Methods in Ecology and Evolution*, *5*(7), 656–665.

Borchers, D., Marques, T., Gunnlaugsson, T., & Jupp, P. (2010). Estimating distance sampling detection functions when distances are measured with errors. *Journal of agricultural, biological, and environmental statistics*, *15*(3), 346–361.

Borchers, D. L., Buckland, S. T., Zucchini, W., & Stephens, W. (2002). *Estimating animal abundance: closed populations* (Vol. 13). Springer Science & Business Media.

Borchers, D. L., & Cox, M. J. (2017). Distance sampling detection functions: 2d or not 2d? *Biometrics*, *73*(2), 593–602.

Borchers, D. L., & Efford, M. (2008). Spatially explicit maximum likelihood methods for capture–recapture studies. *Biometrics*, *64*(2), 377–385.

Borchers, D. L., & Marques, T. A. (2017). From distance sampling to spatial capture–recapture. *AStA Advances in Statistical Analysis*, *101*(4), 475–494.

Borchers, D. L., Stevenson, B., Kidney, D., Thomas, L., & Marques, T. (2015). A unifying model for capture–recapture and distance sampling surveys of wildlife populations. *Journal of the American Statistical Association*, *110*(509), 195–204.

Borchers, D. L., Zucchini, W., & Fewster, R. M. (1998). Mark-recapture models for line transect surveys. *Biometrics*, 1207–1220.

Borchers, D. L., Zucchini, W., Heide-Jørgensen, M., Cañadas, A., & Langrock, R. (2013). Using hidden markov models to deal with availability bias on line transect surveys. *Biometrics*, *69*(3), 703–713.

Brillinger, D. R., Preisler, H. K., Ager, A. A., & Kie, J. G. (2004). An exploratory data analysis (eda) of the paths of moving animals. *Journal of statistical planning and inference*, *122*(1-2), 43–63.

Buckland, S., & Johnston, A. (2017). Monitoring the biodiversity of regions: Key principles and possible pitfalls. *Biological Conservation*, *214*, 23–34.

Buckland, S. T. (1992). Fitting density functions with polynomials. *Applied Statistics*, 63–76.

Buckland, S. T., Anderson, D. R., Burnham, K. P., Laake, J. L., Borchers, D. L., & 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., Borchers, D. L., & Thomas,

L. (2004). *Advanced distance sampling*. Oxford University Press.

Buckland, S. T., Laake, J. L., & Borchers, D. L. (2010). Double-observer line transect methods: Levels of independence. *Biometrics*, *66*(1), 169–177.

Buckland, S. T., Oedekoven, C. S., & Borchers, D. L. (2016). Model-based distance sampling. *Journal of agricultural, biological, and environmental statistics*, *21*(1), 58–75.

Buckland, S. T., Rexstad, E. A., Marques, T. A., & Oedekoven, C. S. (2015). *Distance sampling: methods and applications*. Springer.

Cagnacci, F., Boitani, L., Powell, R. A., & Boyce, M. S. (2010). Animal ecology meets gps-based radiotelemetry: a perfect storm of opportunities and challenges. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *365*(1550), 2157–2162.

Calabrese, J. M., Fleming, C. H., & Gurarie, E. (2016). ctmm: an r package for analyzing animal relocation data as a continuous-time stochastic process. *Methods in Ecology and Evolution*, *7*(9), 1124–1132.

Calhoun, J. B., & Casby, J. U. (1958). *Calculation of home range and density of small mammals* (No. 55). US Department of Health, Education, and Welfare, Public Health Service.

Cappé, O., Moulines, E., & Rydén, T. (2005). *Inference in hidden markov models*. Springer.

Casella, G., & Berger, R. (2001). Statistical inference. *Duxbury Press*.

Chao, A. (1987). Estimating the population size for capture-recapture data with unequal catchability. *Biometrics*, 783–791.

Conn, P., & Alisauskas, R. T. (2017). Simultaneous modelling of movement, measurement error, and observer dependence in double-observer distance sampling surveys. *bioRxiv*, 126821.

Cormack, R. (1964). Estimates of survival from the sighting of marked animals. *Biometrika*, *51*(3/4), 429–438.

Cox, D. R., & Isham, V. (1980). *Point processes* (Vol. 12). CRC Press.

Daniell, P. J. (1919). Integrals in an infinite number of dimensions. *Annals of Mathematics*, *20*(4), 281-288.

Diggle, P. J. (2013). *Statistical analysis of spatial and spatio-temporal point patterns*. CRC Press.

Eddelbuettel, D., François, R., Allaire, J., Ushey, K., Kou, Q., Russel, N., ... Bates, D. (2011). Rcpp: Seamless r and c++ integration. *Journal of Statistical Software*, *40*(8), 1–18.

Eddelbuettel, D., & Sanderson, C. (2014). Rcpparmadillo: Accelerating r with high-

performance c++ linear algebra. *Computational Statistics & Data Analysis*, *71*, 1054–1063.

Efford, M. (2004). Density estimation in live-trapping studies. *Oikos*, *106*(3), 598–610.

Efford, M. (2017). secr: Spatially explicit capture-recapture models [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=secr` (R package version 3.1.0)

Efford, M., Dawson, D. K., Jhala, Y., & Qureshi, Q. (2016). Density-dependent home-range size revealed by spatially explicit capture–recapture. *Ecography*, *39*(7), 676–688.

Efford, M. G. (2014). Bias from heterogeneous usage of space in spatially explicit capture–recapture analyses. *Methods in Ecology and Evolution*, *5*(7), 599–602.

Efford, M. G., Borchers, D. L., & Byrom, A. E. (2009). Density estimation by spatially explicit capture–recapture: likelihood-based methods. In *Modeling demographic processes in marked populations* (pp. 255–269). Springer.

Ergon, T., & Gardner, B. (2014). Separating mortality and emigration: modelling space use, dispersal and survival with robust-design spatial capture–recapture data. *Methods in Ecology and Evolution*, *5*(12), 1327–1336.

Fewster, R., Stevenson, B., Borchers, D. L., et al. (2016). Trace-contrast models for capture–recapture without capture histories. *Statistical Science*, *31*(2), 245–258.

Fewster, R. M., Buckland, S. T., Burnham, K. P., Borchers, D. L., Jupp, P. E., Laake, J. L., & Thomas, L. (2009). Estimating the encounter rate variance in distance sampling. *Biometrics*, *65*(1), 225–236.

Fleming, C. H., Fagan, W. F., Mueller, T., Olson, K. A., Leimgruber, P., & Calabrese, J. M. (2015). Rigorous home range estimation with movement data: a new autocorrelated kernel density estimator. *Ecology*, *96*(5), 1182–1188.

Gardner, B., Reppucci, J., Lucherini, M., & Royle, J. A. (2010). Spatially explicit inference for open populations: estimating demographic parameters from camera-trap studies. *Ecology*, *91*(11), 3376–3383.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis* (Vol. 2). CRC press Boca Raton, FL.

Gelman, A., Rubin, D. B., et al. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, *7*(4), 457–472.

Gerrodette, T., & Forcada, J. (2005). Non-recovery of two spotted and spinner dolphin populations in the eastern tropical pacific ocean. *Marine Ecology Progress Series*, *291*, 1–21.

Gillespie, D. T. (1996). Exact numerical simulation of the ornstein-uhlenbeck process and

its integral. *Physical review E*, *54*(2), 2084.

Gleason, H. A. (1920). Some applications of the quadrat method. *Bulletin of the Torrey Botanical Club*, 21–33.

Glennie, R., Borchers, D., Murchie, M., Harmsen, B. J., & Foster, R. J. (2017). Open population maximum likelihood spatial capture-recapture. *Under Submission*. Retrieved from `http://research-repository.st-andrews.ac.uk/handle/10023/11758`

Glennie, R., Buckland, S. T., & Thomas, L. (2015). The effect of animal movement on line transect estimates of abundance. *PloS one*, *10*(3), e0121333.

Gonzalez, R. P. (2018). *Incorporating animal movement into circular plot and point transect surveys of wildlife abundance* (Unpublished doctoral dissertation).

Gurarie, E., Fleming, C. H., Fagan, W. F., Laidre, K. L., Hernández-Pliego, J., & Ovaskainen, O. (2017). Correlated velocity models as a fundamental unit of animal movement: synthesis and applications. *Movement Ecology*, *5*(1), 13.

Gurarie, E., & Ovaskainen, O. (2013). Towards a general formalization of encounter rates in ecology. *Theoretical Ecology*, *6*(2), 189–202.

Hamilton, O. N. P., Kincaid, S. E., Constantine, R., Kozmian-Ledward, L., Walker, C. G., & Fewster, R. M. (2018). Accounting for uncertainty in duplicate identification and group size judgements in mark–recapture distance sampling. *Methods in Ecology and Evolution*, *9*(2), 354–362.

Hammond, P., Lacey, C., Gilles, A., Viquerat, S., Börjesson, P., Herr, H., ... others (2017). *Estimates of cetacean abundance in european atlantic waters in summer 2016 from the scans-iii aerial and shipboard surveys* (Tech. Rep.). Wageningen Marine Research.

Hammond, P. S., Macleod, K., Berggren, P., Borchers, D. L., Burt, L., Cañadas, A., ... others (2013). Cetacean abundance and distribution in european atlantic shelf waters to inform conservation and management. *Biological Conservation*, *164*, 107–122.

Hanks, E. M., Hooten, M. B., Alldredge, M. W., et al. (2015). Continuous-time discrete-space models for animal movement. *The Annals of Applied Statistics*, *9*(1), 145–165.

Harmsen, B. J., Foster, R. J., Silver, S. C., Ostro, L. E., & Doncaster, C. P. (2009). Spatial and temporal interactions of sympatric jaguars (panthera onca) and pumas (puma concolor) in a neotropical forest. *Journal of mammalogy*, *90*(3), 612–620.

Harmsen, B. J., Foster, R. J., Silver, S. C., Ostro, L. E., & Doncaster, C. P. (2010). The ecology of jaguars in the cockscomb basin wildlife sanctuary, belize. *The biology and conservation of wild felids*, 403–416.

Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*, *57*(1), 97–109.

Hayes, R., & Buckland, S. (1983). Radial-distance models for the line-transect method. *Biometrics*, 29–42.

Hedley, S. L., & Buckland, S. T. (2004). Spatial models for line transect sampling. *Journal of Agricultural, Biological, and Environmental Statistics*, *9*(2), 181.

Hiby, A. (1982). The effect of random whale movement on density estimates obtained from whale sighting surveys. *Report of the International Whaling Commission*, *32*, 791–793.

Hines, J. E., Kendall, W. L., & Nichols, J. D. (2003). On the use of the robust design with transient capture-recapture models. *The Auk*, *120*(4), 1151–1158.

Hoffman, M. D., & Gelman, A. (2014). The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research*, *15*(1), 1593–1623.

Hooten, M. B., Buderman, F. E., Brost, B. M., Hanks, E. M., & Ivan, J. S. (2016). Hierarchical animal movement models for population-level inference. *Environmetrics*, *27*(6), 322–333.

Hooten, M. B., & Johnson, D. S. (2017). Basis function models for animal movement. *Journal of the American Statistical Association*, *112*(518), 578–589.

Hooten, M. B., Johnson, D. S., McClintock, B. T., & Morales, J. M. (2017). *Animal movement: statistical models for telemetry data*. CRC Press.

Hooten, M. B., King, R., & Langrock, R. (2017, Sep 01). Guest editor's introduction to the special issue on "animal movement modeling". *Journal of Agricultural, Biological and Environmental Statistics*, *22*(3), 224–231. Retrieved from `https://doi.org/10.1007/s13253-017-0299-0` doi: 10.1007/s13253-017-0299-0

Hooten, M. B., Scharf, H. R., Hefley, T. J., Pearse, A. T., & Weegman, M. D. (2018). Animal movement models for migratory individuals and groups. *Methods in Ecology and Evolution*.

Hooten, M. B., & Wikle, C. K. (2008). A hierarchical bayesian non-linear spatio-temporal model for the spread of invasive species with application to the eurasian collared-dove. *Environmental and Ecological Statistics*, *15*(1), 59–70.

Howe, E. J., Buckland, S. T., Després-Einspenner, M.-L., & Kühl, H. S. (2017). Distance sampling with camera traps. *Methods in Ecology and Evolution*.

Hutchinson, J., & Waser, P. M. (2007). Use, misuse and extensions of models of animal encounter. *Biological Reviews*, *82*(3), 335–359.

Illian, J. B., Martino, S., Sørbye, S. H., Gallego-Fernández, J. B., Zunzunegui, M., Esquivias, M. P., & Travis, J. M. (2013). Fitting complex ecological point process models with

integrated nested laplace approximation. *Methods in Ecology and Evolution*, *4*(4), 305–315.

IUCN. (2017). Iucn red list of threatened species. *Version 2017-3*.

Johnson, D. S., London, J. M., Lea, M.-A., & Durban, J. W. (2008). Continuous-time correlated random walk model for animal telemetry data. *Ecology*, *89*(5), 1208–1215.

Jolly, G. M. (1965). Explicit estimates from capture-recapture data with both death and immigration-stochastic model. *Biometrika*, *52*(1/2), 225–247.

Jones, A. R., Wynn, R. B., Yésou, P., Thébault, L., Collins, P., Suberg, L., ... Brereton, T. M. (2014). Using integrated land-and boat-based surveys to inform conservation of the critically endangered balearic shearwater puffinus mauretanicus in northeast atlantic waters. *Endangered Species Research*, *25*(1), 1–18.

Karanth, K. U., Nichols, J. D., Kumar, N., & Hines, J. E. (2006). Assessing tiger population dynamics using photographic capture–recapture sampling. *Ecology*, *87*(11), 2925–2937.

Kie, J. G., Ager, A. A., & Bowyer, R. T. (2005). Landscape-level movements of north american elk (cervus elaphus): effects of habitat patch structure and topography. *Landscape Ecology*, *20*(3), 289–300.

King, R., Bird, S. M., Hay, G., & Hutchinson, S. J. (2009). Estimating current injectors in scotland and their drug-related death rate by sex, region and age-group via bayesian capture—recapture methods. *Statistical Methods in Medical Research*, *18*(4), 341–359.

King, R., McClintock, B. T., Kidney, D., Borchers, D., et al. (2016). Capture–recapture abundance estimation using a semi-complete data likelihood approach. *The Annals of Applied Statistics*, *10*(1), 264–285.

Klinck, H., Mellinger, D. K., Klinck, K., Bogue, N. M., Luby, J. C., Jump, W. A., ... others (2012). Near-real-time acoustic monitoring of beaked whales and other cetaceans using a seaglider$^{\mathrm{TM}}$. *PloS one*, *7*(5), e36128.

Kyhn, L. A., Tougaard, J., Thomas, L., Duve, L. R., Stenback, J., Amundin, M., ... Teilmann, J. (2012). From echolocation clicks to animal density—acoustic sampling of harbor porpoises with static dataloggers. *The Journal of the Acoustical Society of America*, *131*(1), 550–560.

Laake, J. L., Buckland, S. T., Burnham, K. P., & Anderson, D. R. (1993). Distance user's guide v2.0 [Computer software manual]. (Colorado Cooperative Fish and Wildlife Research Unit, Colorado State University, Fort Collins)

Laake, J. L., Burnham, K. P., & Anderson, D. R. (1979). User's manual for program

transect [Computer software manual]. (Utah State University Press, Logan, UT)

Laidre, K. L., Heide-Jørgensen, M. P., & Nielsen, T. G. (2007). Role of the bowhead whale as a predator in west greenland. *Marine Ecology Progress Series*, *346*, 285–297.

Langrock, R., Hopcraft, J. G. C., Blackwell, P. G., Goodall, V., King, R., Niu, M., ... Schick, R. S. (2014). Modelling group dynamic animal movement. *Methods in Ecology and Evolution*, *5*(2), 190–199.

Langrock, R., King, R., Matthiopoulos, J., Thomas, L., Fortin, D., & Morales, J. M. (2012). Flexible and practical modeling of animal telemetry data: hidden markov models and extensions. *Ecology*, *93*(11), 2336–2342.

LeVeque, R. J. (2007). *Finite difference methods for ordinary and partial differential equations: steady-state and time-dependent problems* (Vol. 98). Siam.

Linden, D. W., Siren, A. P., & Pekins, P. J. (2017). Integrating telemetry data into spatial capture-recapture modifies inferences on multi-scale resource selection. *bioRxiv*, 131144.

Lyles, R. H., Guo, Y., & Greenland, S. (2012). Reducing bias and mean squared error associated with regression-based odds ratio estimators. *Journal of statistical planning and inference*, *142*(12), 3235–3241.

Marple, S. L., & Marple, S. L. (1987). *Digital spectral analysis: with applications* (Vol. 5). Prentice-Hall Englewood Cliffs, NJ.

Marques, F. F., & Buckland, S. T. (2003). Incorporating covariates into standard line transect analyses. *Biometrics*, *59*(4), 924–935.

Marques, F. F., Buckland, S. T., Goffin, D., Dixon, C. E., Borchers, D. L., Mayle, B. A., & Peace, A. J. (2001). Estimating deer abundance from line transect surveys of dung: sika deer in southern scotland. *Journal of Applied Ecology*, *38*(2), 349–363.

Marques, T. A. (2004). Predicting and correcting bias caused by measurement error in line transect sampling using multiplicative error models. *Biometrics*, *60*(3), 757–763.

Marques, T. A., Thomas, L., Martin, S. W., Mellinger, D. K., Ward, J. A., Moretti, D. J., ... Tyack, P. L. (2013). Estimating animal population density using passive acoustics. *Biological Reviews*, *88*(2), 287–309.

McClintock, B. T., Johnson, D. S., Hooten, M. B., Ver Hoef, J. M., & Morales, J. M. (2014). When to be discrete: the importance of time formulation in understanding animal movement. *Movement Ecology*, *2*(1), 21.

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (Vol. 37). CRC press.

McKellar, A. E., Langrock, R., Walters, J. R., & Kesler, D. C. (2014). Using mixed hidden markov models to examine behavioral states in a cooperatively breeding bird.

Behavioral Ecology, *26*(1), 148–157.

Méndez, V., Campos, D., & Bartumeus, F. (2013). *Stochastic foundations in movement ecology: anomalous diffusion, front propagation and random searches*. Springer Science & Business Media.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The journal of chemical physics*, *21*(6), 1087–1092.

Mevel, L., & Finesso, L. (2004). Asymptotical statistics of misspecified hidden markov models. *IEEE Transactions on Automatic Control*, *49*(7), 1123–1132.

Miller, D. L. (2017). Distance: Distance sampling detection function and abundance estimation [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=Distance` (R package version 0.9.7)

Miller, D. L., Burt, M. L., Rexstad, E. A., & Thomas, L. (2013). Spatial models for distance sampling data: recent developments and future directions. *Methods in Ecology and Evolution*, *4*(11), 1001–1010.

Miller Jr, R. G. (2011). *Survival analysis* (Vol. 66). John Wiley & Sons.

Mitchell, A. R., & Griffiths, D. F. (1980). *The finite difference method in partial differential equations*. John Wiley.

Moler, C., & Van Loan, C. (2003). Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. *SIAM review*, *45*(1), 3–49.

Morales, J. M., Haydon, D. T., Frair, J., Holsinger, K. E., & Fryxell, J. M. (2004). Extracting more out of relocation data: building movement models as mixtures of random walks. *Ecology*, *85*(9), 2436–2445.

Nathan, R., Getz, W. M., Revilla, E., Holyoak, M., Kadmon, R., Saltz, D., & Smouse, P. E. (2008). A movement ecology paradigm for unifying organismal movement research. *Proceedings of the National Academy of Sciences*, *105*(49), 19052–19059.

Newson, S. E., Evans, K. L., Noble, D. G., Greenwood, J. J., & Gaston, K. J. (2008). Use of distance sampling to improve estimates of national population sizes for common and widespread breeding birds in the uk. *Journal of Applied Ecology*, *45*(5), 1330–1338.

Niu, M., Blackwell, P. G., & Skarin, A. (2016). Modeling interdependent animal movement in continuous time. *Biometrics*, *72*(2), 315–324.

Okubo, A., & Levin, S. A. (2013). *Diffusion and ecological problems: modern perspectives* (Vol. 14). Springer Science & Business Media.

Otis, D. L., Burnham, K. P., White, G. C., & Anderson, D. R. (1978). Statistical inference from capture data on closed animal populations. *Wildlife monographs*(62), 3–135.

Ovaskainen, O., Rekola, H., Meyke, E., & Arjas, E. (2008). Bayesian methods for analyzing movements in heterogeneous landscapes from mark–recapture data. *Ecology*, *89*(2), 542–554.

Palka, D., & Hammond, P. (2001). Accounting for responsive movement in line transect estimates of abundance. *Canadian journal of fisheries and aquatic sciences*, *58*(4), 777–787.

Parton, A., Blackwell, P. G., & Skarin, A. (2016). Bayesian inference for continuous time animal movement based on steps and turns. In *International conference on bayesian statistics in action* (pp. 223–230).

Patterson, T. A., Parton, A., Langrock, R., Blackwell, P. G., Thomas, L., & King, R. (2017). Statistical modelling of individual animal movement: an overview of key methods and a discussion of practical challenges. *AStA Advances in Statistical Analysis*, *101*(4), 399–438.

Patterson, T. A., Thomas, L., Wilcox, C., Ovaskainen, O., & Matthiopoulos, J. (2008). State–space models of individual animal movement. *Trends in ecology & evolution*, *23*(2), 87–94.

Pedersen, M. W., Patterson, T. A., Thygesen, U. H., & Madsen, H. (2011). Estimating animal behavior and residency from movement data. *Oikos*, *120*(9), 1281–1290.

Pettex, E., David, L., Authier, M., Blanck, A., Dorémus, G., Falchetto, H., ... others (2016). Using large scale surveys to investigate seasonal variations in seabird distribution and abundance. part i: The north western mediterranean sea. *Deep Sea Research Part II: Topical Studies in Oceanography*.

Pledger, S. (2000). Unified maximum likelihood estimates for closed capture–recapture models using mixtures. *Biometrics*, *56*(2), 434–442.

Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). Coda: Convergence diagnosis and output analysis for mcmc. *R News*, *6*(1), 7–11. Retrieved from `https://journal.r-project.org/archive/`

Pollock, K. H., Nichols, J. D., Brownie, C., & Hines, J. E. (1990). Statistical inference for capture-recapture experiments. *Wildlife monographs*, 3–97.

Preisler, H. K., Ager, A. A., Johnson, B. K., & Kie, J. G. (2004). Modeling animal movements using stochastic differential equations. *Environmetrics*, *15*(7), 643–657.

Quigley, F. R. P. L. P. E. S. R., H., & Harmsen, B. (2017). Panthera onca. *The IUCN Red List of Threatened Species*.

R Core Team. (2017). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Rabinowitz, A. R., & Jr, B. (1986). Ecology and behaviour of the jaguar (panthers onca) in belize, central america. *Journal of Zoology*, *210*(1), 149–159.

Rexstad, E., & Burnham, K. P. (1991). *User's guide for interactive program capture.* Color. Cooperative Fish and Wildlife Research Unit.

Roberts, G. O., Gelman, A., Gilks, W. R., et al. (1997). Weak convergence and optimal scaling of random walk metropolis algorithms. *The annals of applied probability*, *7*(1), 110–120.

Ronconi, R. A., & Burger, A. E. (2009). Estimating seabird densities from vessel transects: distance sampling and implications for strip transects. *Aquatic Biology*, *4*(3), 297–309.

Royle, J. A., Chandler, R. B., Gazenski, K. D., & Graves, T. A. (2013). Spatial capture–recapture models for jointly estimating population density and landscape connectivity. *Ecology*, *94*(2), 287–294.

Royle, J. A., Chandler, R. B., Sollmann, R., & Gardner, B. (2013). *Spatial capture-recapture.* Academic Press.

Royle, J. A., Chandler, R. B., Sun, C. C., & Fuller, A. K. (2013). Integrating resource selection information with spatial capture–recapture. *Methods in Ecology and Evolution*, *4*(6), 520–530.

Royle, J. A., Fuller, A. K., & Sutherland, C. (2016). Spatial capture–recapture models allowing markovian transience or dispersal. *Population ecology*, *58*(1), 53–62.

Royle, J. A., Fuller, A. K., & Sutherland, C. (2017). Unifying population and landscape ecology with spatial capture–recapture. *Ecography*.

Royle, J. A., & Young, K. V. (2008). A hierarchical model for spatial capture–recapture data. *Ecology*, *89*(8), 2281–2289.

Rudin, W. (1987). *Real and complex analysis.* Tata McGraw-Hill Education.

Russell, J. C., Hanks, E. M., & Haran, M. (2016). Dynamic models of animal movement with spatial point process interactions. *Journal of agricultural, biological, and environmental statistics*, *21*(1), 22–40.

Saad, Y. (1992a). Analysis of some krylov subspace approximations to the matrix exponential operator. *SIAM Journal on Numerical Analysis*, *29*(1), 209–228.

Saad, Y. (1992b). *Numerical methods for large eigenvalue problems.* Manchester University Press.

SAMBAH. (2016). *Static acoustic monitoring of the baltic sea harbour porpoise* (Tech. Rep.). Retrieved from `http://www.sambah.org/SAMBAH-Final-Report-FINAL-for-website-April-2017.pdf`

Sanderson, C., & Curtin, R. (2016). Armadillo: a template-based c++ library for linear algebra. *Journal of Open Source Software*.

Scharf, H., Hooten, M. B., & Johnson, D. S. (2017). Imputation approaches for animal movement modeling. *Journal of Agricultural, Biological and Environmental Statistics*, *22*(3), 335–352.

Scharf, H. R., Hooten, M. B., Johnson, D. S., & Durban, J. W. (2018). Process convolution approaches for modeling interacting trajectories. *Environmetrics*, *29*(3), e2487.

Schick, R. S., Loarie, S. R., Colchero, F., Best, B. D., Boustany, A., Conde, D. A., ... Clark, J. S. (2008). Understanding movement data and movement processes: current and emerging directions. *Ecology letters*, *11*(12), 1338–1350.

Schwarz, C. J., & Seber, G. A. (1999). Estimating animal abundance: review iii. *Statistical Science*, 427–456.

Scott, M. D., & Chivers, S. J. (2009). Movements and diving behavior of pelagic spotted dolphins. *Marine Mammal Science*, *25*(1), 137–160.

Seber, G. A. (1965). A note on the multiple-recapture census. *Biometrika*, *52*(1/2), 249–259.

Seber, G. A. (1986). A review of estimating animal abundance. *Biometrics*, 267–292.

Seber, G. A. (1992). A review of estimating animal abundance ii. *International Statistical Review/Revue Internationale de Statistique*, 129–166.

Seber, G. A. F. (1982). The estimation of animal abundance and related parameters.

Secretariat, C. (2010). The strategic plan for biodiversity 2011-2020 and the aichi biodiversity targets..

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*(3/4), 591–611.

Sidje, R. B. (1998). Expokit: a software package for computing matrix exponentials. *ACM Transactions on Mathematical Software (TOMS)*, *24*(1), 130–156.

Sidje, R. B., & Hansen, N. R. (2012). Package 'exporkit'.

Sollmann, R., Gardner, B., Parsons, A. W., Stocking, J. J., McClintock, B. T., Simons, T. R., ... O'Connell, A. F. (2013). A spatial mark–resight model augmented with telemetry data. *Ecology*, *94*(3), 553–559.

Spear, L. B., Ainley, D. G., Hardesty, B. D., Howell, S. N., & Webb, S. W. (2004). Reducing biases affecting at-sea surveys of seabirds: use of multiple observer teams. *Marine Ornithology*, *32*(2), 147–157.

Standart, I., & Subcommittee, P. (2017). *Guidelines for using the iucn red list categories and criteria.* Version.

Sutherland, C., Fuller, A. K., & Royle, J. A. (2015). Modelling non-euclidean movement and landscape connectivity in highly structured ecological networks. *Methods in Ecology and Evolution*, *6*(2), 169–177.

Tao, T. (2011). *An introduction to measure theory* (Vol. 126). American Mathematical Soc.

Thomas, L., Buckland, S. T., Rexstad, E. A., Laake, J. L., Strindberg, S., Hedley, S. L., ... Burnham, K. P. (2010). Distance software: design and analysis of distance sampling surveys for estimating population size. *Journal of Applied Ecology*, *47*(1), 5–14.

Turchin, P. (1998). *Quantitative analysis of movement*. Sinauer assoc. Sunderland (mass.).

Turnock, B. J., & Quinn, T. J. (1991). The effect of responsive movement on abundance estimation using line transect sampling. *Biometrics*, 701–715.

Uhlenbeck, G. E., & Ornstein, L. S. (1930). On the theory of the brownian motion. *Physical review*, *36*(5), 823.

Ver Hoef, J. M. (2012). Who invented the delta method? *The American Statistician*, *66*(2), 124–127.

White, G. C., & Burnham, K. P. (1999). Program mark: survival estimation from populations of marked animals. *Bird study*, *46*(sup1), S120–S139.

Worton, B. (1987). A review of models of home range for animal movement. *Ecological modelling*, *38*(3-4), 277–298.

Yapp, W. (1956). The theory of line transects. *Bird study*, *3*(2), 93–104.

Yuan, Y., Bachl, F., Lindgren, F., Brochers, D., Illian, J., Buckland, S., ... Gerrodette, T. (2016). Point process models for spatio-temporal distance sampling data. *arXiv preprint arXiv:1604.06013*.

Zhang, X., Wang, Q., & Saar, W. (2017). *Openblas: An optimized blas library*.

Zucchini, W., MacDonald, I. L., & Langrock, R. (2016). *Hidden markov models for time series: an introduction using r* (Vol. 150). CRC press.

Appendix A

In this appendix, the computational algorithms are described for the methods presented in this thesis. The algorithms are written in pseudo-code. For each application, these methods were implemented using the statistical programming language `R (3.4.4)` (R Core Team, 2017). For efficiency, many of the algorithms were implemented in `C++98` with the linear algebra library `armadillo (9.100)` (Sanderson & Curtin, 2016) and numerical libraries `OpenBLAS (0.2.20)` (Zhang, Wang, & Saar, 2017) and `LAPACK (3.1)` (Anderson et al., 1999). The `C++` implementations were used within `R` through the packages `Rcpp (0.12.16)` (Eddelbuettel et al., 2011) and `RcppArmadillo (0.8.400)` (Eddelbuettel & Sanderson, 2014).

The appendix is structured in the same order as the thesis: path integration, distance sampling, continuous-time spatial capture-recapture, and discrete-time spatial capture-recapture. For each section, the main algorithms are described.

A1: Path Integration

Algorithm 1 shows the method to compute a path integral using the hidden Markov model (HMM) forward algorithm as described in Section 2.5.2. This algorithm relies upon four other algorithms to get the initial distribution, get the transition rate matrix (TRM), weight the probability density vector, and update the probability density vector.

1.1 HMM quadrature

The overall idea is that a single vector $\boldsymbol{\phi}$ contains the probability density within each cell in the grid created over space; for each time step this probability is first weighted by the likelihood of any observations made during that time and then updated by the movement of individuals across the grid. This movement is described by the transition rate matrix, $\boldsymbol{R}$. When path integration is used for each application, the initial distribution, weighting,

and updating is specified differently.

Algorithm 1: Path integration

Input:

- $\boldsymbol{\theta}_m, \boldsymbol{\theta}_d$: parameter values for movement and observation processes, respectively;
- G: total number of grid cells, G_x is number in each row, G_y number in each column;
- Δx: spatial grid cell width and length;
- Δt: time-step;
- T: total time to integrate over.

Output: I, value of the path integral approximation.

$\boldsymbol{\delta} \leftarrow$ **Get Initial Distribution**$(G, \Delta x)$;
$\boldsymbol{R} \leftarrow$ **Get Trm**$(G, \boldsymbol{\theta}_m, \Delta x)$;
$\tilde{T} \leftarrow \lfloor \frac{T}{\Delta t} \rfloor$;
$\boldsymbol{\phi} \leftarrow \boldsymbol{\delta}$;
$I \leftarrow 0$;
for $t \leftarrow 0$ **to** $\tilde{T}$ **do**
    $\boldsymbol{\psi} \leftarrow$ **Weight**$(t, \boldsymbol{\phi}, \boldsymbol{\theta}_d, \Delta x, \Delta t)$;
    $\boldsymbol{\phi} \leftarrow$ **Update**$(t, \boldsymbol{\psi}, \Delta t, \boldsymbol{R}, G)$;
    $s \leftarrow \sum_{i=1}^{G} \boldsymbol{\phi}_i$;
    $I \leftarrow I + \log(s)$;
    $\boldsymbol{\phi} \leftarrow \boldsymbol{\phi}/s$;
end
return $\exp(I)$

Algorithm 2 shows how the HMM algorithm is adapted to compute state-switching path integrals as described in Section 2.5.7. Here, a matrix $\boldsymbol{\phi}$ is constructed where each row $\boldsymbol{\phi}[b,]$ gives the probability density in each grid cell for individuals in state b. For each time step, each entry in this matrix is weighted by the likelihood of any observations made given an individual occupied that grid cell and inhabited that state. Afterwards, the matrix is updated to allow for individuals to move across the grid or change state. Notice, a state-switching path integral requires a state-switching TRM and matrix initial distribution to be specified.

Algorithm 2 also uses functions **Vectorise** and **UnVectorise**. These functions are used to take the matrix form for $\boldsymbol{\phi}$ where each row represents the spatial probability density for a state and each column the distribution over states for a single grid cell and collapses this into a single row vector: this row vector is the concatenation of the rows of the matrix

form. The **UnVectorise** function does the opposite operation. This is necessary because the **Update** function multiplies this vectorised form by the state-switching transition rate matrix, $\boldsymbol{R}$.

Algorithm 2: State-Switching Path Integral

Input:

- $\boldsymbol{\theta}_m^{(b)}, \boldsymbol{\theta}_d^{(b)}$: parameter values for movement and observation respectively where $^{(b)}$ denotes the subset associated with behaviour b;
- G: total number of grid cells, G_x is number in each row, G_y number in each column;
- B: number of states;
- Δx: spatial grid cell width and length;
- Δt: time-step;
- T: total time to integrate over.

Output: I, value of the path integral approximation.

$\boldsymbol{\delta} \leftarrow$ **Get Initial Distribution**$(G, B, \Delta x)$;
$\boldsymbol{R} \leftarrow$ **Get State-Switching Trm**$(G, B, \boldsymbol{\theta}_m, \Delta x)$;
$\tilde{T} \leftarrow \lfloor \frac{T}{\Delta t} \rfloor$;
$\boldsymbol{\phi} \leftarrow \boldsymbol{\delta}$;
$I \leftarrow 0$;
for $t \leftarrow 0$ **to** $\tilde{T}$ **do**
    foreach *state* b **do** $\boldsymbol{\psi}[b,] \leftarrow$ **Weight**$(t, \boldsymbol{\phi}[b,], \boldsymbol{\theta}_d^{(b)}, \Delta x, \Delta t)$;
    $\boldsymbol{\psi} \leftarrow$ **Vectorise**$(\boldsymbol{\psi})$;
    $\boldsymbol{\phi} \leftarrow$ **Update**$(t, \boldsymbol{\psi}, \Delta t, \boldsymbol{R}, G)$;
    $\boldsymbol{\phi} \leftarrow$ **UnVectorise**$(\boldsymbol{\phi})$;
    $s \leftarrow \sum_{i=1}^{G} \boldsymbol{\phi}_i$;
    $I \leftarrow I + \log(s)$;
    $\boldsymbol{\phi} \leftarrow \boldsymbol{\phi}/s$;
end
return I

1.2 Transition Rate Matrix

The transition rate matrix (TRM) is derived as a numerical approximation to the advection-diffusion partial differential equation (Section 2.5.2). Here, the algorithms are given to compute the TRM for Brownian motion (Algorithm 3), Ornstein-Uhlenbeck motion (Algorithm

4), and state-switching movement (Algorithm 5) on a spatial grid.

The state-switching algorithm can be used to compute the TRM for state-switching between any movement models for which a state-dependent TRM can be computed; thus, this algorithm can be used to create a TRM for state-switching Brownian or Ornstein-Uhlenbeck motion.

The TRM should be stored as a sparse matrix. When the spatial grid is large in extent or has a high resolution, the size of the TRM is large; thus, storing this matrix as a sparse matrix has a computational advantage. There are several implementations of sparse matrices. In this thesis, all TRMs were stored in column-major form using the `spmat` data type available in the `C++` library `armadillo`. For TRMs that are known to be Toeplitz, an alternative method used was to store only the first row and column of the matrix.

It is assumed in the algorithms given for the TRM that any entries not assigned a value have default value zero.

1.4 Weight

The vector giving the probability density within each grid cell is weighted by the likelihood of an observation given the individual occupied that grid cell. This likelihood is denoted $f_{\boldsymbol{\theta}}(s,t)$ where s is the grid cell, t is the time of the observation, and $\boldsymbol{\theta}$ the observation process parameters. This function will change depending on the application. Algorithm 6 gives the procedures for a general f.

1.5 Krylov approximation

Algorithm 7 shows how to compute the Krylov approximation as described in Section 2.5.4. The algorithm is used to compute $\exp(\boldsymbol{R}\Delta t)\boldsymbol{\phi}$ for a transition rate matrix $\boldsymbol{R}$ and a row vector $\boldsymbol{\phi}$. This is a simplified version of the algorithm. The full algorithm also includes steps to compute the error of this approximation *a posteriori*. If this error exceeds a chosen tolerance, then the time-step is split into smaller time-steps and the approximation refined. The tolerance used throughout this thesis was 1×10^{-10}. The algorithm is implemented in the `ExpoKit` library for Fortran and Matlab (Sidje, 1998); it is also implemented for `R` (Sidje & Hansen, 2012). For this thesis, the algorithm was re-implemented in `C++98` using the `armadillo` library.

The computational savings are made in Line 5 of Algorithm 7. Here, a vector is multiplied by the transition rate matrix, $\boldsymbol{R}$. When stored as a sprase matrix, this multiplication is

Algorithm 3: Get Trm: Brownian Motion

Input:

- G: total number of grid cells, G_x is number in each row, G_y number in each column;
- σ: standard deviation of Brownian motion;
- Δx: spatial grid cell width and length.

Output: $\boldsymbol{R}$, sparse transition rate matrix for Brownian motion.

$\boldsymbol{R} \leftarrow$Sparse Matrix filled with zeros;
$r \leftarrow \dfrac{\sigma^2}{2\Delta x^2}$;
for $i \leftarrow 0$ **to** G_x **do**
  for $j \leftarrow 0$ **to** G_y **do**
    $a \leftarrow 0$;
    $s \leftarrow i + G_x j$;
    if $i < G_x - 1$ **then**
      $\boldsymbol{R}[s, s+1] \leftarrow r$;;
      $a \leftarrow a + r$;
    end
    if $i > 0$ **then**
      $\boldsymbol{R}[s, s-1] \leftarrow r$;
      $a \leftarrow a + r$;
    end
    if $j < G_y - 1$ **then**
      $\boldsymbol{R}[s, s+G_x] \leftarrow r$;
      $a \leftarrow a + r$;
    end
    if $j > 0$ **then**
      $\boldsymbol{R}[s, s-G_x] \leftarrow r$;
      $a \leftarrow a + r$;
    end
    $\boldsymbol{R}[s, s] \leftarrow -a$;
  end
end
return $\boldsymbol{R}$

Algorithm 4: Get Trm: Ornstein-Uhlenbeck motion

Input:

- G: total number of grid cells, G_x is number in each row, G_y number in each column;
- σ: standard deviation of Brownian motion;
- α: attraction to activity centre;
- $\boldsymbol{\mu}$: activity centre;
- Δx: spatial grid cell width and length.

Output: $\boldsymbol{R}$, sparse transition rate matrix for Ornstein-Uhlenbeck motion.

$\boldsymbol{R} \leftarrow$Sparse Matrix filled with zeros;
$r \leftarrow \dfrac{\sigma^2}{2\Delta x^2}$;
$v \leftarrow -\dfrac{\alpha}{2\Delta x}$;
for $i \leftarrow 0$ **to** G_x **do**
  $v_x \leftarrow v(i\Delta x - \boldsymbol{\mu}_x)$;
  for $j \leftarrow 0$ **to** G_y **do**
    $v_y \leftarrow v(j\Delta x - \boldsymbol{\mu}_y)$;
    $a \leftarrow 0$;
    $s \leftarrow i + G_x j$;
    if $i < G_x - 1$ **then**
      $\boldsymbol{R}[s, s+1] \leftarrow r + v_x$;;
      $a \leftarrow a + r$;
    end
    if $i > 0$ **then**
      $\boldsymbol{R}[s, s-1] \leftarrow r - v_x$;
      $a \leftarrow a + r$;
    end
    if $j < G_y - 1$ **then**
      $\boldsymbol{R}[s, s+G_x] \leftarrow r + v_y$;
      $a \leftarrow a + r$;
    end
    if $j > 0$ **then**
      $\boldsymbol{R}[s, s-G_x] \leftarrow r - v_y$;
      $a \leftarrow a + r$;
    end
    $\boldsymbol{R}[s, s] \leftarrow -a$;
  end
end
return $\boldsymbol{R}$

Algorithm 5: Get Trm: state-switching

Input:

- G: total number of grid cells, G_x is number in each row, G_y number in each column;
- B: number of states;
- $\boldsymbol{\theta}^{(b)}$: movement parameters for state b;
- $\boldsymbol{\lambda}$: behaviour switching transition rate matrix;
- Δx: spatial grid cell width and length.

Output: $\boldsymbol{R}$, sparse transition rate matrix for state-switching motion.

$\boldsymbol{R} \leftarrow$Sparse Matrix of size $GB \times GB$ filled with zeros;
foreach *state* b **do**
  $\boldsymbol{R}[(b-1)G : bG, (b-1)G : bG] \leftarrow$ Get State Dependent Trm$(G, \boldsymbol{\theta}^{(b)}, \Delta x)$;
end
for $s \leftarrow 0$ **to** G **do**
  foreach *pair of states* (b, d) **do**
    $\boldsymbol{R}[s + bG, s + dG]) = \boldsymbol{R}[s_0 + bG, s_0 + dG] + \boldsymbol{\lambda}[b, d]$;
  end
end
return $\boldsymbol{R}$

Algorithm 6: Weight

Input:

- t: time;
- $\boldsymbol{\phi}$: probability distribution over grid;
- $\boldsymbol{\theta}$: parameters for observation process;
- G: number of grid cells;
- f: function that returns probability of observation at time t for each grid cell.

Output: $\boldsymbol{\phi}$, probability distribution weighted by likelihood of observation.

if *no observation made* **then return** $\boldsymbol{\phi}$;
for $s \leftarrow 0$ **to** G **do**
  $\boldsymbol{\phi}[s] \leftarrow \boldsymbol{\phi}[s] f_{\boldsymbol{\theta}}(s, t)$;
end
return $\boldsymbol{\phi}$

tractable for large spatial grids. When the matrix is Toeplitz, an alternative computational approach is to store only the first row and column of $\boldsymbol{R}$ and use the two-dimensional fast Fourier transform to perfrom the multiplication in Line 5. This is described in Section 2.5.5.

Algorithm 7: Update by simplified Krylov approximation

Input:

- t: time;
- $\boldsymbol{\phi}$: probability distribution over grid;
- $\boldsymbol{R}$: transition rate matrix;
- Δx: spatial grid cell width and length;
- Δt: time step;
- G: number of grid cells;
- K: dimension of Krylov subspace;
- $\boldsymbol{e}$: vector of length K filled with zeros, except first entry is a one.

Output: $\boldsymbol{\psi}$, probability distribution updated by TRM.

```
1  V ← G × K matrix;
2  H ← K × K matrix;
3  V[, 1] ← φ‖φ‖^{-1};
4  for j ← 0 to K do
5  |   w ← V[, j − 1]R;
6  |   for i ← 0 to j do
7  |   |   H[i, j − 1] ← w · V[, i];
8  |   |   w ← w − H[i, j − 1]V[, i];
9  |   end
10 |   H[j, j − 1] ← ‖w‖;
11 |   w ← w(H[j, j − 1])^{-1};
12 |   V[, j] ← w;
13 end
14 ψ ← ‖φ‖V exp(Hdt)e;
15 return ψ
```

A2: Distance Sampling

Section 3.2 describes how distance sampling is a type of encounter model; the new models are termed movement in distance sampling models (MDS). The algorithims in Section A1 are used to compute the necessary path integral. The HMM quadrature requires that the initial distribution, weighting, and updating be specified. In this section, each of these elements are described. Afterwards, the likelihood computation and abundance estimation algorithms are given. For each transect, space is discretised into square cells of length Δx. This space contains the entire transect and a buffer region around the transect. The buffer B is the distance added to all sides of the rectangular space.

The initial distribution for both line and point transects is uniform over the spatial grid. That is, each element of $\boldsymbol{\phi}$ is $\frac{1}{G}$ where G is the number of grid points. In distance sampling analysis, the assumption is made that transects are placed at random with respect to the position of individuals.

The updating step is performed using the Krylov approximation (Algorithm 7) and the Brownian motion transition rate matrix (Algorithm 3).

2.1 Weighting

For MDS models, the vector $\boldsymbol{\phi}$ at time step t contains for each grid cell s the probability density that an individual occupies grid cell s and has not been detected up to that time, termed the survival probability. At each time-step, this vector is weighted by the survival probability. One can think of this as the probability density in each grid cell being thinned by the detection process.

Survival probability depends on the hazard of detection which itself depends on the relative distance between the observer and the individual. Algorithm 8 describes how to compute the hazard of detection for an individual that is in position (x, y) relative to the observer. This algorithm uses the hazard of the form cr^{-d} where r is the radial distance separating the individual and the observer, and (c, d) are parameters. For point transects, this is a simple multiplication; for line transects, the hazard must be integrated over the distance travelled by the observer along the line. For the form of hazard considered here, this has a closed form expression given in Glennie et al. (2015).

The survival probability for each grid cell is given by computing the average hazard of detection over the cell (Algorithm 9). There is not a closed form expression for this, so the two-dimensional trapezoidal rule is used. The integration points are a $m \times m$ subgrid of

Algorithm 8: Calculate Hazard of Detection

Input:

- x: perpendicular distance from observer;
- y: forward distance from observer;
- Δt: time step;
- v: observer speed;
- d: detection shape parameter;
- s: detection scale parameter;
- tol: tolerance below which numbers are assumed to be zero;
- B: compute Beta function;
- I: compute incomplete Beta function, CDF of a Beta distributed variable.

Output: h, hazard of detection.

$r_0 \leftarrow \sqrt{x^2 + y^2}$;
$c \leftarrow s^d$;
if *point transect* **then** $h \leftarrow \dfrac{\Delta t c}{r_0^d}$;
else if *line transect* **then**
  if $y < 0$ **then return** 0;
  $\beta \leftarrow 0.5(d-1)$;
  $y_1 \leftarrow y - v\Delta t$;
  if $y_1 < 0$ **then** $y_1 \leftarrow 0$;
  $r_1 \leftarrow \sqrt{x^2 + y_1^2}$;
  if $r_1 < tol$ **then return** ∞;
  if $(\|x\| < tol)$ **then**
    if $\|d-1\| < tol$ **then** $h \leftarrow c(\log(r_1) - \log(r_0))$;
    else $h \leftarrow \dfrac{c}{d-1}\left(r_1^{-\beta} - r_0^{-\beta}\right)$;
  end
  else
    $h \leftarrow \left(I\left(\dfrac{x^2}{r_1^2}, \beta, 0.5\right) - I\left(\dfrac{x^2}{r_0^2}, \beta, 0.5\right)\right) B(\beta, 0.5) \dfrac{c}{2\|x\|^{d-1}}$;
  end
end
return h

equally spaced points within the grid cell.

2.2 Likelihood

For distance sampling surveys, individuals and transects are assumed to be independent. Thus, the likelihood is a product of likelihood contributions for each individual. Algorithm 10 describes how to compute the likelihood for a given survey. This algorithm is similar to the general HMM quadrature (Algorithm 1), but includes additional steps in lines 11–16. This is because the MDS likelihood for a single individual involves the product of a path integral (as in Algorithm 1) with a hazard of detection (Section 3.2.1).

The variable l stores the log-likelihood value for the entire survey. The vector $\boldsymbol{\phi}$ contains the probability density over space for the current transect being processed. Line 11 states that if an individual was detected during time step t, then this observation makes a contribution to the likelihood which, as stated on Line 13, is a product of the hazard of detection (f) and the survival probability up to that time within that grid cell ($\boldsymbol{\phi}(g_{i,j})$). The variable h is used to scale the vector $\boldsymbol{\phi}$ during the HMM algorithm to prevent numerical underflow; it is included in Line 13 to reverse this scaling. Finally $\log(\Delta x^2)$ is included in Line 13 as the individual could occupy any location within the grid cell.

Recall that the MDS model is conditional on an individual being detected at some time during the transect. Line 24 is equivalent to dividing the likelihood contribution for each individual by the probability of being detected on the particular transect that individual was seen upon. The detection probability, p, is computed iteratively on Line 18: for each grid cell, the difference between probability of survival at time-step $t-1$ and time-step t is the probability that an individual within that grid cell is detected; these probabilities are then multiplied by $\boldsymbol{I}$ pointwise (the Hadamard product). The vector $\boldsymbol{I}$ contains, for each grid cell, a one if the grid cell is inside the transect and a zero if not. Multiplication by $\boldsymbol{I}$ is necessary because an individual is only recorded and included in the analysis if it is detected within the transect; it is possible that an individual may be have been detected outside the transect, but excluded from the analysis — this is common when truncating data in distance sampling analyses.

2.3 Abundance Estimation

Once the detection and movement parameters have been estimated, Algorithm 11 describes how to compute a design-based estimate of abundance. For each transect j in the survey, $\boldsymbol{p}_j$ can be interpreted as the proportion of the spatial grid that is sampled; hence, $\frac{\boldsymbol{n}_j}{\boldsymbol{p}_j}$ is an

Algorithm 9: Calculate Probability of Surviving Detection

Input:

- t: time;
- $\boldsymbol{\theta}$: detection parameters;
- G: number of cells in total and G_x, G_y directions;
- Δt: time step;
- Δx: spatial grid size;
- v: observer speed;
- m: number of integration points.

Output: $\boldsymbol{s}$, survival probability for each grid cell.

$\boldsymbol{s} \leftarrow G \times G$ matrix of zeros;
$\boldsymbol{z} \leftarrow$ observer position at time t;
$y_0 \leftarrow 0$;
if *line transect* **then** $y_0 \leftarrow \lfloor \frac{\boldsymbol{z}_y}{\Delta x} \rfloor$;
for $i \leftarrow 0$ **to** G_x **do**
    for $j \leftarrow y_0$ **to** G_y **do**
        $g \leftarrow i + jG_x$;
        $x \leftarrow i\Delta x - \boldsymbol{z}_x$;
        $y \leftarrow j\Delta x - \boldsymbol{z}_y$;
        $\boldsymbol{f} \leftarrow$ vector of length m^2;
        for $k \leftarrow 0$ **to** m **do**
            $x_1 \leftarrow x + \frac{k}{m}\Delta x$;
            for $l \leftarrow 0$ **to** m **do**
                $y_1 \leftarrow y + \frac{l}{m}\Delta x$;
                $\boldsymbol{f}_{k+lm} \leftarrow h(x_1, y_1, \boldsymbol{\theta})$;
            end
        end
        $\boldsymbol{s}_g \leftarrow$Trapezoidal Rule($\boldsymbol{f}$);
        $\boldsymbol{s}_g \leftarrow \exp(-\boldsymbol{s}_g)$;
    end
end
return $\boldsymbol{s}$

Algorithm 10: Log-Likelihood for Distance Sampling with Movement

Input:

- $\boldsymbol{\theta}_d, \boldsymbol{\theta}_m$: detection and movement parameters respectively;
- $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\tau}, \boldsymbol{g}$: matrix of perpendicular distances, forward distances, detection times, and grid cell for ith detection on jth transect;
- $\boldsymbol{T}$: time taken to survey each transect;
- $\boldsymbol{n}$: number of individuals seen on each transect;
- $\boldsymbol{v}$: speed of observer along each transect;
- G: number of grid cells in total and G_x, G_y in each direction;
- $\boldsymbol{I}_j$: vector with 1 for grid cells inside transect j and 0 for those outside;
- Δx: spatial grid cell size;
- Δt: time-step.

Output: l, log-likelihood for distance sampling with movement.

$l \leftarrow 0$;
$\boldsymbol{R} \leftarrow$**Get Trm**$(G, \boldsymbol{\theta}_m, \Delta x)$;
foreach *transect* j **do**
  $\boldsymbol{\phi} \leftarrow$ CalcInitialDistribution$(G, \Delta x)$;
  $\tilde{T}_j \leftarrow \lfloor \frac{\tilde{T}}{\Delta t} \rfloor$;
  $i \leftarrow 1$;
  $h \leftarrow 0$;
  $c \leftarrow$ TRUE;
  $p \leftarrow 0$;
  for $t \leftarrow 0$ **to** $\tilde{T}_j$ **do**
    while c *and* $\tau_{i,j} \leq t\Delta t$ **do**
      $f \leftarrow$Calculate hazard seen at $(\boldsymbol{x}_{i,j}, \boldsymbol{y}_{i,j}, \boldsymbol{\tau}_{i,j})$;
      $l \leftarrow l + \log(f) + \log(\boldsymbol{\phi}(\boldsymbol{g}_{i,j})) + h - \log(\Delta x^2)$;
      $i \leftarrow i + 1$;
      if $i >$ *number of individuals seen on transect* j **then** $c \leftarrow$ FALSE;
    end
    $\boldsymbol{\psi} \leftarrow$**Weight**$(t, \boldsymbol{\phi}, \boldsymbol{\theta}_d, v, \Delta x, \Delta t)$;
    $p \leftarrow p + \exp(h) \sum_{g=1}^{G} ((\boldsymbol{\phi} - \boldsymbol{\psi}) \circ \boldsymbol{I}_j)$;
    $\boldsymbol{\phi} \leftarrow$**Update**$(t, \boldsymbol{\psi}, \Delta t, \boldsymbol{R}, G)$;
    $s \leftarrow \sum_{i=1}^{G} \boldsymbol{\phi}_i$;
    $h \leftarrow h + \log(s)$;
    $\boldsymbol{\phi} \leftarrow \boldsymbol{\phi}/s$;
  end
  $l \leftarrow l - n_i \log(p)$;
end
return $\boldsymbol{l}$

estimate of abundance derived from data on transect j. Over all transects, $\sum_{j=1}^{J} \boldsymbol{p}_j$ is the total proportion of the spatial grid that is sampled. The abundance estimator is thus as given in Line 17 of Algorithm 11.

2.4 Applications

The key algorithms to fit distance sampling with animal movement models for point and line transects are implemented in the `R` package `moveds`. Appendix B provides a guide to using the package. The computational details for the analysis of the spotted dolphin and shearwater case studies are also given in Appendix B.

Algorithm 11: Estimate Abundance in Distance Sampling with Movement

Input:

- $\boldsymbol{\theta}_d, \boldsymbol{\theta}_m$: detection and movement parameters respectively;
- J: number of transects;
- $\boldsymbol{n}$: number of individuals seen on each transect;
- $\boldsymbol{v}$: speed of observer along each transect;
- G: number of grid cells in total and G_x, G_y in each direction;
- $\boldsymbol{I}$: vector with 1 for grid cells inside transect and 0 for those outside;
- Δx: spatial grid cell size;
- Δt: time-step.

Output: $\hat{N}$, abundance estimate.

$l \leftarrow 0$;
$\boldsymbol{R} \leftarrow$**Get Trm**$(G, \boldsymbol{\theta}_m, \Delta x)$;
$p \leftarrow$ zero vector of length J;
foreach *transect* j **do**
  $\boldsymbol{\phi} \leftarrow$ CalcInitialDistribution$(G, \Delta x)$;
  $\tilde{T}_j \leftarrow \lfloor \frac{\tilde{T}}{\Delta t} \rfloor$;
  $h \leftarrow 0$;
  for $t \leftarrow 0$ **to** $\tilde{T}_j$ **do**
    $\boldsymbol{\psi} \leftarrow$**Weight**$(t, \boldsymbol{\phi}, \boldsymbol{\theta}_d, v, \Delta x, \Delta t)$;
    $\boldsymbol{p}[j] \leftarrow \boldsymbol{p}[j] + \exp(h) \sum_{g=1}^{G} ((\boldsymbol{\phi} - \boldsymbol{\psi}) \circ \boldsymbol{I}_j)$;
    $\boldsymbol{\phi} \leftarrow$**Update**$(t, \boldsymbol{\psi}, \Delta t, \boldsymbol{R}, G)$;
    $s \leftarrow \sum_{i=1}^{G} \boldsymbol{\phi}_i$;
    $h \leftarrow h + \log(s)$;
    $\boldsymbol{\phi} \leftarrow \boldsymbol{\phi}/s$;
  end
end
$\hat{N} \leftarrow \dfrac{\sum_{j=1}^{J} \boldsymbol{n}_j}{\sum_{j=1}^{J} \boldsymbol{p}_j}$;
return $\hat{N}$

A3: Continuous-time spatial capture-recapture

In Section 4.3, continuous-time spatial capture-recapture (SCR) surveys are shown to be a type of encounter model. For a single individual, its recorded location at detectors over time are observed. The likelihood of these observations is a product of the encounter intensity and the exponential of a path integral (Section 4.3.1).

The likelihood is similar to the HMM quadrature algorithm (Algorithm 1): it requires an initial distribution for the individual's location, a transition rate matrix, and a weighting function. Space and time are both discretised.

Conceptually in SCR surveys, individuals have an activity range around which they move. In continuous-time, this is described by the Ornstein-Uhlenbeck process, hence the transition rate matrix is computed using Algorithm 4. This depends on the activity centre of the individual, $\boldsymbol{z}$. Here, the described algorithms assume this activity centre is known. Appendix C describes how to estimate this activity centre from SCR data. Given this, the initial distribution of the individual is a bivariate Gaussian around this activity centre.

The weighting function is the probability an individual would survive detection during the time-step, given the number and placement of detectors active in that time-step. The equation for S_t is given in Section 4.3.1. This accumulated hazard $\boldsymbol{h}_j$ is computed for detector j at all grid points using Algorithm 9 where Line 19 is omitted. These hazards are then summed across all J detectors $\boldsymbol{h} = \sum_{j=1}^{J} \boldsymbol{h}_j$. The weighting probability is then given by $\exp(-\boldsymbol{h})$ for each grid cell in space. The probability of being detected in a time-step is thus $1 - \exp(-\boldsymbol{h})$.

3.1 Likelihood

Algorithm 12 describes how to compute the likelihood given data on a single individual in a continuous-time SCR survey and their activity centre. The TRM, $\boldsymbol{R}$, depends on the activity centre of an individual as its movement will be biased toward this centre. Also, the TRM depends on a vector $\boldsymbol{I}$ that has a one or zero for each grid cell, indicating whether that grid cell is accessible or not to the individual. This allows one to account for coastlines or impenetrable boundaries of the study area. The initial distribution also depends on this, as individuals cannot exist at points where $\boldsymbol{I}$ is zero.

In Algorithm 12, at time-step t, the vector $\boldsymbol{\phi}$ contains the probability density of what has been observed up to time t: the individual has been seen some number of times at certain detectors and eluded detection otherwise. At each time-step, this probability density is

weighted by the survival probability using the **Weight** function when the individual has eluded detection by all detectors. If the individual is seen, then Lines 11–16 show that the probability density is instead weighted by the complement of the survival probability, the detection probability.

Notice, that Algorithm 12 is a direct application of the forward algorithm for hidden Markov models (see Algorithm A.1.3 in Zucchini et al. (2016)). The transition probability matrix is $\mathbf{\Gamma} = \exp(\boldsymbol{R}\Delta t)$ and the probability of observations in each time-step, stored by Zucchini et al. (2016) in a matrix, is either the survival probability (no detections) or the detection probability.

3.2 Inference

Once the movement and detection parameters are estimated, Section 4.3.6 describes the inferences that can be made on the movement of an individual and the interaction between individuals. In this section, it is described how to compute the time-spent distribution for an individual and the spatio-temporal overlap between individuals.

Recall from the previous section that the approximated likelihood is a direct application of the forward algorithm for hidden Markov models. Similarly, the approximation of the inferences proposed in Section 4.3.6 are made using the algorithms that exist for hidden Markov models.

Zucchini et al. (2016) present their Appendix A `R` code for these algorithms; this code can be treated as pseudo-code.

Time-spent distribution

Recall from Section 4.3.6, that the time spent distribution is a function $\mathcal{T}$ of space such that $\mathcal{T}(\boldsymbol{x})$ is the estimated amount of time during the survey the individual spent in location $\boldsymbol{x}$. It is defined as the integral over the predictive distribution of where the individual is for each time given the data.

$$\mathcal{T}_i(\boldsymbol{x}) = \int_0^T [\vec{\boldsymbol{x}}_i(t) = \boldsymbol{x} \mid \boldsymbol{r}] \, \mathrm{d}t$$

Hence, to compute this quantity one must compute the conditional distribution of the hidden location of an individual over time and space. This is approximated by computing the approximate distribution for each time-step and grid cell. Algorithm A.1.11 in Zucchini et al. (2016) can be used to compute the predictive distribution for the hidden state (the

Algorithm 12: Log-Likelihood for continuous-time spatial capture-recapture with movement

Input:

- $\boldsymbol{\theta}_d, \boldsymbol{\theta}_m$: detection and movement parameters respectively;
- $\boldsymbol{z}$: individual's activity centre;
- $\boldsymbol{\tau}, \boldsymbol{g}$: matrix of detection times and grid cell for ith detection;
- $\boldsymbol{T}$: survey duration;
- G: number of grid cells in total and G_x, G_y in each direction;
- $\boldsymbol{I}$: vector with 1 for grid cells inside survey area and 0 for those outside;
- Δx: spatial grid cell size;
- Δt: time-step.

Output: l, log-likelihood value

$l \leftarrow 0$;
$\boldsymbol{R} \leftarrow$**Get Trm**$(G, \boldsymbol{z}, \boldsymbol{\theta}_m, \Delta x, \boldsymbol{I})$;
$\boldsymbol{\phi} \leftarrow$ CalcInitialDistribution$(G, \boldsymbol{\theta}_m, \boldsymbol{z}, \Delta x, \boldsymbol{I})$;
$\tilde{T} \leftarrow \lfloor \frac{\tilde{T}}{\Delta t} \rfloor$;
$i \leftarrow 1$;
$c \leftarrow$ TRUE;
$p \leftarrow 0$;
for $t \leftarrow 0$ **to** $\tilde{T}_j$ **do**
  seen $\leftarrow$ FALSE;
  while c *and* $\tau_{i,j} \leq t\Delta t$ **do**
    seen $\leftarrow$ TRUE;
    $f \leftarrow$Calculate survival probability seen at $(\boldsymbol{x}_{i,j}, \boldsymbol{y}_{i,j}, \boldsymbol{\tau}_{i,j})$;
    $f \leftarrow 1 - f$;
    $\boldsymbol{\psi} \leftarrow \boldsymbol{\phi} \circ f$;
    $i \leftarrow i + 1$;
    if $i >$ *number of captures in survey* **then** $c \leftarrow$ FALSE;
  end
  if *not seen* **then** $\boldsymbol{\psi} \leftarrow$**Weight**$(t, \boldsymbol{\phi}, \boldsymbol{\theta}_d, \Delta x, \Delta t)$;
  $\boldsymbol{\phi} \leftarrow$**Update**$(t, \boldsymbol{\psi}, \Delta t, \boldsymbol{R}, G)$;
  $s \leftarrow \sum_{i=1}^{G} \boldsymbol{\phi}_i$;
  $l \leftarrow l + \log(s)$;
  $\boldsymbol{\phi} \leftarrow \boldsymbol{\phi}/s$;
end
return $\boldsymbol{l}$

spatial location of the individual). This algorithm relies on the computation of the forward and backward probabilities.

The forward probabilities can be stored in a matrix $\boldsymbol{F}$ where each row corresponds to a time-step and each column to a grid cell. The forward probabilities can be computed using Algorithm 12; row t of $\boldsymbol{F}$ is the vector $\boldsymbol{\phi}$ at the beginning of iteration t.

The backward probabilities $\boldsymbol{B}$ are computed similarly to Algorithm A.1.8 in Zucchini et al. (2016). The only differences from Algorithm 12 is that the loop is run backwards in time, the initial distribution is a uniform distribution, and $\boldsymbol{\phi}$ is a column vector. This has one complication: in the forward algorithm, the transition probability matrix is multiplied by a row vector, $\boldsymbol{\phi\Gamma}$, and this is appproximated using the Krylov subspace approximation in Algorithm 7. For the backwards algorithm, the multiplication is a matrix with a column vector, $\boldsymbol{\Gamma\phi}$. To perform this multiplication using Algorithm 7, one must take the transposition of this product and compute $\boldsymbol{\phi}^t\boldsymbol{\Gamma}^t$ using Algorithm 7, and then take the transposition of this as the result.

The predictive distribution for each time-step in each grid cell is thus computed using $\boldsymbol{F}$ and $\boldsymbol{B}$ as in Algorithm A.1.11 in Zucchini et al. (2016). This produces a matrix $\boldsymbol{P}$ where the (i, j) entry is the probability an individual occupied cell j at time-step i given the observed data and estimated parameters. The time-spent distribution over the spatial grid is the column sums of $\boldsymbol{P}$.

Animal-Animal interactions

The previous section describes how to compute the predictive distribution for each time-step and each grid cell. Let $\boldsymbol{P}_i$ be the matrix approximating the continuous predictive distribution over space and time for individual i and $\mathcal{T}_i$ the derived time-spent distribution.

Inter-individual interaction can be quantified as the spatial overlap between individuals. Spatial overlap between individual i and individual j is computed as the total area of cells where the pointwise multiplication of their time-spent distributions is positive.

Another measure of inter-individual interaction is to compute at what distance individuals spend at least $p\%$ of their time that distance or futher apart. Algorithm 13 presents one way to compute this distance.

Algorithm 13: Get spatio-temporal separation distance

Input:

- $\boldsymbol{P}_i, \boldsymbol{P}_j$: spatio-temporal predictive distributions for individuals i and j;
- p: level at which to compute the distance individuals spend at least $p\%$ of their apart;
- Δx: spatial grid length and width;
- T: total survey time.

Output: l, separation distance between individuals i and j.

$d \leftarrow 0$;
$\mathcal{T}_i \leftarrow$ column sums of $\boldsymbol{P}_i$;
$c \leftarrow$ TRUE;
while c **do**
    $d \leftarrow d + \Delta x$;
    $s \leftarrow 0$;
    foreach *grid cell* g *where* $\mathcal{T}_i > 0$ **do**
        Compute distance between g and every other grid cell;
        $s \leftarrow s+$ sum of $\boldsymbol{P}_j[, k]$ for all cells k where cells g and k are less than d apart;
    end
    $s \leftarrow s/T$;
    if $s \geq 1 - p$ **then** $c \leftarrow$ FALSE;
end
return d

3.3 Application

The algorithms presented in this appendix show how to compute the likelihood contribution from observations of a single individual when the activity centre of this individual is known.

Appendix C describes how these algorithms can be used within a Markov chain Monte Carlo algorithm to estimate the detection, movement, and density parameters. Appendix C further describes the computational details for the application of this approach to the jaguar case study in Section 4.4.

A4: Discrete-time spatial capture-recapture

Section 5.3 describes how the multi-dimensional integral that arises in discrete-time SCR with moving activity centres can be computed efficiently using the Krylov approximation (Algorithm 7). This idea is applied to three applications in Chapter 5: SCR with moving activty centres, SCR with transience and resident individuals, and finally to Cormack-Jolly-Seber SCR with moving activity centres. In this appendix, the algorithms for all three likelihoods are presented in pseudo-code. These algorithms are implemented in the `R` package `openpopscr`. The use of this package is described in Appendix D.

4.1 Discrete-time SCR with movement

SCR surveys are conducted with different detector types. Here the algorithm is presented for proximity detectors, e.g. DNA hair snares as in the bear example in Section 5.4.

Algorithm 14 describes how to compute the likelihood for detection, movement, and density parameters given observed captures across individuals and detectors. The initial distribution function here is assumed to be uniform across the study area as in distance sampling. The transition rate matrix is computed using Algorithm 3. The update function is an application of the Krylov approximation (Algorithm 7) where the time-step $\boldsymbol{\Delta t}_k$ is the duration between occasions k and $k+1$.

The vector $\boldsymbol{\phi}$ is used to store for each grid cell the likelihood of the observations up to that point for that individual given its activity centre resided within that grid cell. The weight function multiplies each of these probabilites by the probability of the observed capture history for each occasion. For proximity detectors, this weighting is either the detection probability at each distance, if the individual was seen on that detector, or the survival probability at that distance if it was not seen on this detector. Algorithm 9 describes

how to compute the survival probability for each grid cell given a hazard function. The most common hazard function to use is the half-normal as described in Borchers and Efford (2008). The weighting is then carried as in Algorithm 6 with f being the survival probability or the detection probability (the complement of the survival probability) depending on the capture history observed.

Notice, that in Algorithm 14, Lines 4–13 compute the log-likelihood l of the detection and movement parameters given the observed capture histories. In SCR surveys, the likelihood from which inference is drawn is the conditional likelihood, conditional on an individual being detected at least once during the survey. Lines 14–22 compute this probability, p. Line 23 contains the contribution made by the density parameter.

4.2 Discrete-time SCR with transients and residents

In Section 5.5, a discrete-time SCR model was presented for populations with a mixture of transient individuals, whose activity centres move between occasions, and resident individuals, whose activity centres are static. This was described as a mixture model. The likelihood can be computed as the likelihood of a mixture model also. Algorithm 14 can be used to compute the likelihood contribution for each individual when that individual's activity centre moves over time, denote this contribution for a single individual by L_T. If Line 9 of Algorithm 14 is omitted, that is, if the probability density is not updated, then individuals do not move and the algorithm is equivalent to that used to compute the likelihood for a conventional SCR model. Denote the likelihood contribution for a single individual under this algorithm to be L_R. It follows that if ψ is the probability of being a transient individual, the likelihood for the full model is

$$L = \psi L_T + (1 - \psi) L_R$$

In practice, rather than perform the iterations in Algorithm 14 twice, it is more efficient to simulatenously perform the algorithm on two vectors $\boldsymbol{\phi}_T$ and $\boldsymbol{\phi}_R$ where the former is updated through time and the later is not. The same approach can be taken when computing the detection probability p in Lines 14–23.

4.3 Cormack-Jolly-Seber SCR with movement

In Section 5.6, the Cormack-Jolly-Seber (CJS) model is presented. Here, the algorithm to compute the log-likelihood of this model is given. It is similar to the state-switching

Algorithm 14: Log-Likelihood for discrete-time spatial capture-recapture with movement

Input:

- $\boldsymbol{\theta}_d, \boldsymbol{\theta}_m$: detection and movement parameters respectively;
- D: density parameter;
- A: area of survey region;
- n: number of individuals seen;
- $\boldsymbol{c}$: capture history for each individual i for detector j on discrete occasion k;
- G: number of grid cells in total and G_x, G_y in each direction;
- $\boldsymbol{I}$: vector with 1 for grid cells inside survey area and 0 for those outside;
- Δx: spatial grid cell size;
- $\boldsymbol{\Delta t}$: time intervals between occasions.

Output: l, log-likelihood value.

$l \leftarrow 0$;
$\boldsymbol{R} \leftarrow$**Get Trm**$(G, \boldsymbol{\theta}_m, \Delta x, \boldsymbol{I})$;
$\boldsymbol{\delta} \leftarrow$ **CalcInitialDistribution**$(G, \boldsymbol{\theta}_m, \Delta x, \boldsymbol{I})$;
foreach *individual* **do**
    $\boldsymbol{\phi} \leftarrow \boldsymbol{\delta}$;
    foreach *occasion* **do**
        $\boldsymbol{c}_{i,k} \leftarrow$ capture history of individual on that occasion;
        $\boldsymbol{\psi} \leftarrow$ **Weight**$(\boldsymbol{\phi}, \boldsymbol{\theta}_d, \boldsymbol{c}_{i,k})$;
        $\boldsymbol{\phi} \leftarrow$ **Update**$(\boldsymbol{\Delta t}_k, \boldsymbol{\psi}, \boldsymbol{R}, G)$;
        $s \leftarrow \sum_{i=1}^{G} \boldsymbol{\phi}_i$;
        $l \leftarrow l + \log(s)$;
    end
end
$p \leftarrow 0$;
$\boldsymbol{\phi} \leftarrow \boldsymbol{\delta}$;
$\boldsymbol{e} \leftarrow$ empty capture history;
foreach *occasion* **do**
    $\boldsymbol{\psi} \leftarrow$ **Weight**$(\boldsymbol{\phi}, \boldsymbol{\theta}_d, \boldsymbol{e})$;
    $\boldsymbol{\phi} \leftarrow$ **Update**$(\boldsymbol{\Delta t}_k, \boldsymbol{\psi}, \boldsymbol{R}, G)$;
    $s \leftarrow \sum_{i=1}^{G} \boldsymbol{\phi}_i$;
    $p \leftarrow p + \log(s)$;
end
$p \leftarrow 1 - \exp(p)$;
$l_n \leftarrow n \log(DA) - DAp - \log(n!)$;
$l \leftarrow l + l_n$;
return l

algorithms (Algorithms 2 and 16). In CJS surveys, individuals can either be alive or dead. The CJS algorithm is simpler than the state-switching algorithms because individuals can only move when alive and cannot move whilst dead; hence only one update is required for each time iteration.

Algorithm 15 gives the steps to compute the likelihood of the Cormack-Jolly-Seber model with moving activity centres. The initial distribution, transition rate matrix, and weighting are as described in Section 4.1. There are three new elements to this algorithm: the variable d, the parameter α, and the variable p.

The vector $\boldsymbol{\phi}$ contains for each grid cell the probability that an individual is alive and occupies that grid cell. The variable d contains the probability an individual is dead. This variable need not be spatial because the observation process is invariant to space for dead individuals, that is, once dead only empty capture histories (no detections) can be observed. The variable d undergoes the same steps as the vector $\boldsymbol{\phi}$: it is weighted and updated. The weighting is simple: if an individual has been seen on a given occasion then it cannot be dead and so d must be zero, this is shown on Line 12. If an individual is not seen on an occasion, this is a certainty when it is dead and so d is weighted by one, it does not change value. The value of d is not updated spatially. It is however updated through time by the survival process; in Line 15, the value of d is increased by the total probability an individual is alive in an occasion and then dies. This probability depends on the duration between occasions $\boldsymbol{\Delta t}$. Similarly, the probability of occupying each grid cell is multiplied by the survival probability in Line 16.

The variable p is used to compute the probability an individual was seen by at least one detection during the first occasion that it was seen on. This is because CJS models are conditional on the first time an individual is seen. This is why the likelihood is divided by this probability in Line 21.

Algorithm 15: Log-Likelihood for Cormack-Jolly-Seber spatial capture-recapture with movement

Input:

- $\boldsymbol{\theta}_d, \boldsymbol{\theta}_m$: detection and movement parameters respectively;
- α: survival probability, that is, probability an individual is alive in the next occasion given it is alive in the current occasion;
- $\boldsymbol{c}$: capture history for each individual for detector on each discrete occasion;
- $\boldsymbol{s}$: first occasion that individuals are available for detection;
- $\boldsymbol{e}$: last occasion that individuals are available for detection, possibly due to removal from the population;
- G: number of grid cells in total and G_x, G_y in each direction;
- $\boldsymbol{I}$: vector with 1 for grid cells inside survey area and 0 for those outside;
- Δx: spatial grid cell size;
- $\boldsymbol{\Delta t}$: time intervals between occasions.

Output: l, log-likelihood value.

$l \leftarrow 0$;
$\boldsymbol{R} \leftarrow$**Get Trm**$(G, \boldsymbol{\theta}_m, \Delta x, \boldsymbol{I})$;
$\boldsymbol{\delta} \leftarrow$ **CalcInitialDistribution**$(G, \boldsymbol{\theta}_m, \Delta x, \boldsymbol{I})$;
$\boldsymbol{e} \leftarrow$ empty capture history;
foreach *individual* **do**
    $\boldsymbol{\phi} \leftarrow \boldsymbol{\delta}$;
    $d \leftarrow 0$;
    foreach *occasion between* $\boldsymbol{s}_i$ *and* $\boldsymbol{e}_i$ **do**
        $\boldsymbol{c}_{i,k} \leftarrow$ capture history of individual on that occasion;
        $\boldsymbol{\psi} \leftarrow$ **Weight**$(\boldsymbol{\phi}, \boldsymbol{\theta}_d, \boldsymbol{c}_{i,k})$;
        if *first time seen* **then** $p \leftarrow$ **Weight**$(\boldsymbol{\phi}, \boldsymbol{\theta}_d, \boldsymbol{e})$;
        if *seen on this occasion* **then** $d \leftarrow 0$;
        $\boldsymbol{\phi} \leftarrow$ **Update**$(\boldsymbol{\Delta t}_k, \boldsymbol{\psi}, \boldsymbol{R}, G)$;
        $s \leftarrow \sum_{i=1}^{G} \boldsymbol{\phi}_i$;
        $d \leftarrow d + s(1 - \alpha^{\boldsymbol{\Delta t}_k})$;
        $\boldsymbol{\phi} \leftarrow \boldsymbol{\phi}\alpha^{\boldsymbol{\Delta t}_k}$;
        $s \leftarrow s + d$;
        $l \leftarrow l + \log(s)$;
    end
    $p \leftarrow 1 - p$;
    $l \leftarrow l - \log(p)$;
end
return l

Appendix B

In this appendix, the `R` package `moveds` is described. This package implements the distance sampling with animal movement model for point and line transects. It can be used to estimate the detection function accounting for animal movement either by specifying a movement rate or providing auxiliary movement data. The package can also provide an estimate of abundance and a measure of goodness-of-fit.

First, this appendix contains two examples that show how to use the package for line and point transect surveys (Sections B1 and B2). Second, the computational details for the spotted dolphin and shearwater applications are discussed (Sections B3 and B4).

The package can be installed using the R package `devtools`:

```
devtools::install_github("r-glennie/moveds", build_vignettes = TRUE)
```

It can then be loaded in the usual way:

```
library(moveds)
```

B1: Line Transects

I load some simulated line transect data in the required format.

```
data("line_example_dsdat")
str(line_example_dsdat)

# List of 2
#  $ obs  :'data.frame': 334 obs. of  4 variables:
#   ..$ transect: int [1:334] 1 2 2 2 2 3 3 3 4 4 ...
#   ..$ x       : num [1:334] 0.307 -4.093 -10.792 -29.28 1.045 ...
```

```
#   ..$ y       : num [1:334] 9.47 17.63 3.66 9.81 5.98 ...
#   ..$ t       : num [1:334] 509 171 192 541 885 ...
#  $ trans: num [1:100, 1:2] 1 2 3 4 5 6 7 8 9 10 ...

obs <- line_example_dsdat$obs
trans <- line_example_dsdat$trans
```

The object `obs` contains the records of detections made during the line transect survey.

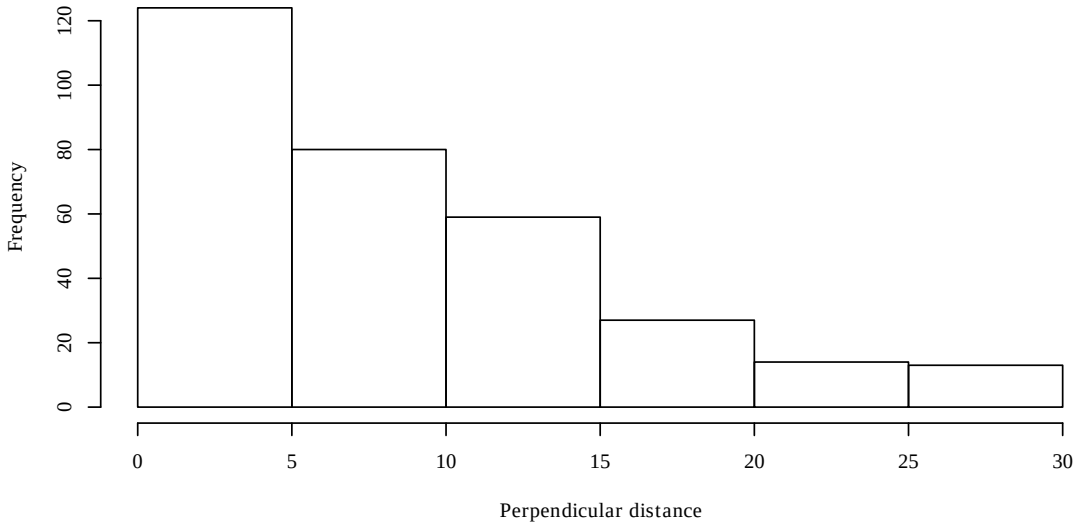
```
summary(obs)

#     transect            x                   y                   t
#  Min.   :  1.00   Min.   :-29.9963   Min.   :  0.1164   Min.   :  0.9096
#  1st Qu.: 22.00   1st Qu.: -7.1687   1st Qu.:  4.1608   1st Qu.:273.8548
#  Median : 49.00   Median :  0.2989   Median :  8.8777   Median :528.4980
#  Mean   : 49.55   Mean   : -0.2354   Mean   : 13.0759   Mean   :513.9855
#  3rd Qu.: 76.00   3rd Qu.:  6.5858   3rd Qu.: 14.9528   3rd Qu.:757.1722
#  Max.   :100.00   Max.   : 28.4892   Max.   :141.9650   Max.   :997.0250

nrow(obs)

# [1] 334
```

There are 334 detections made with 95% of perpendicular distances less than 23.3 distance units.

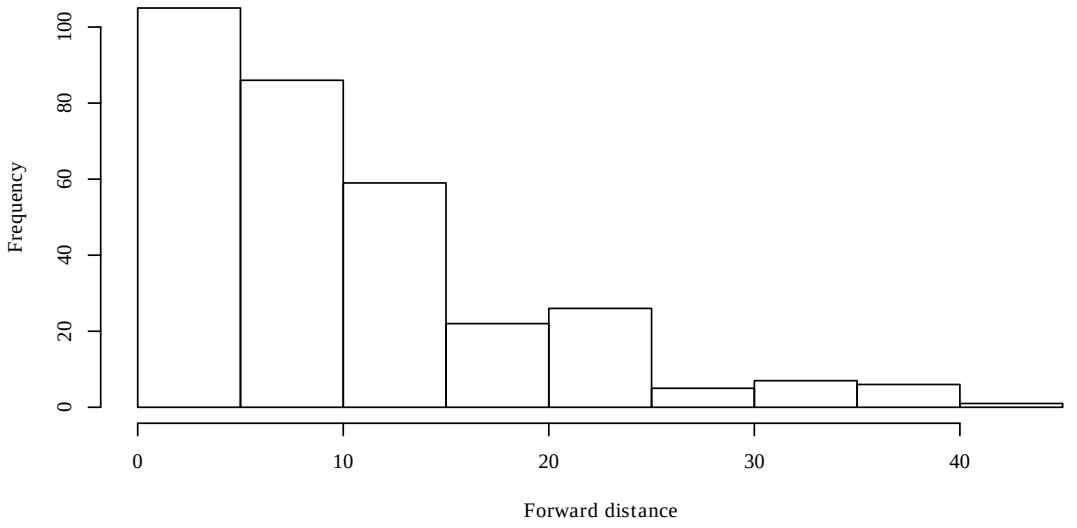
There are a few large forward distances recorded, to make things better to plot, let's look only at those less than the 95% quantile.

```
y.quantile <- quantile(obs$y, prob = 0.95)
subobs <- obs[obs$y <= y.quantile,]
```

```
hist(abs(subobs$x), main = "", xlab = "Perpendicular distance")
```




```
hist(subobs$y, main = "", xlab = "Forward distance")
```



The object `trans` contains the transect ID numbers and the length of each transect, termed the effort. In this simulation, all transects had equal length. In total, the survey consisted of 100 transects.

Single sightings of an individual do not provide any information on how individuals move. Auxiliary movement information is required. You can either provide a fixed value for the diffusion parameter or provide movement data for this parameter to be estimated.

If you want to fix the diffusion parameter, you can simulate movement data based on a fixed value to see if distances travelled seem reasonable for the study species.

For this analysis, I use simulated tag data on five individuals:

```
data("line_example_movedat")
str(line_example_movedat[[1]])

#  num [1:101, 1:3] 103.1 97.1 98.3 84.3 87.7 ...
#  - attr(*, "dimnames")=List of 2
#   ..$ : NULL
#   ..$ : chr [1:3] "x" "y" "observation.times"
```

2D CDS

You can fit a two-dimensional conventional distance sampling (CDS) model using the `moveds` package. First, I set up the data into the format required for the `mds` function. One important variable is `aux` which contains required information: the region width and length, the truncation distance, the observer's average speed (in the simulated survey, the observer had speed 1 distance unit per time unit), and the transect type (0 for line transects, 1 for point transects).

```
aux <- c(1000, 1000, 30, 1, 0)
```

The first object required by the `mds` function is the list corresponding to the distance data.

```
ds <- list(data = obs,
           transect = trans,
           delta = c(2.5, 1000),
           aux = aux,
           buffer = 0,
           hazardfn = 1,
           move = 0)
```

The variable `buffer` is only relevant to models where individuals move; here, I am fitting a CDS model, so individuals do not move. The hazard function I use is isotropic and is equivalent to the hazard-rate model used in the 1D case.

The `delta` variable controls the discretisation of space and time. Calculations are performed on a grid with spacing 2.5. For time, there is no need to discretise since the time discretisation only affects the movement of indviduals, not the observer, so I set that to the maximum 1000, the entire survey time. Note, that the time discretisation should not exceed the time it takes to survey a transect.

Finally, I set `move` to 0 because I want to fit a model where the individuals are assumed not to move.

The second object `mds` requires corresponds to the movement data. Here, I am fitting a CDS model, where individual's don't move, so I will just provide some dummy variables:

```
move <- list(data = NULL)
```

To fit the model, you must give it starting values: the first value is always the detection scale and the second the detection shape. It is a good idea to try a few different sets of starting values to ensure the global maximum has been found. The `moveds` package uses `nlm` to maximise the log-likelihood.

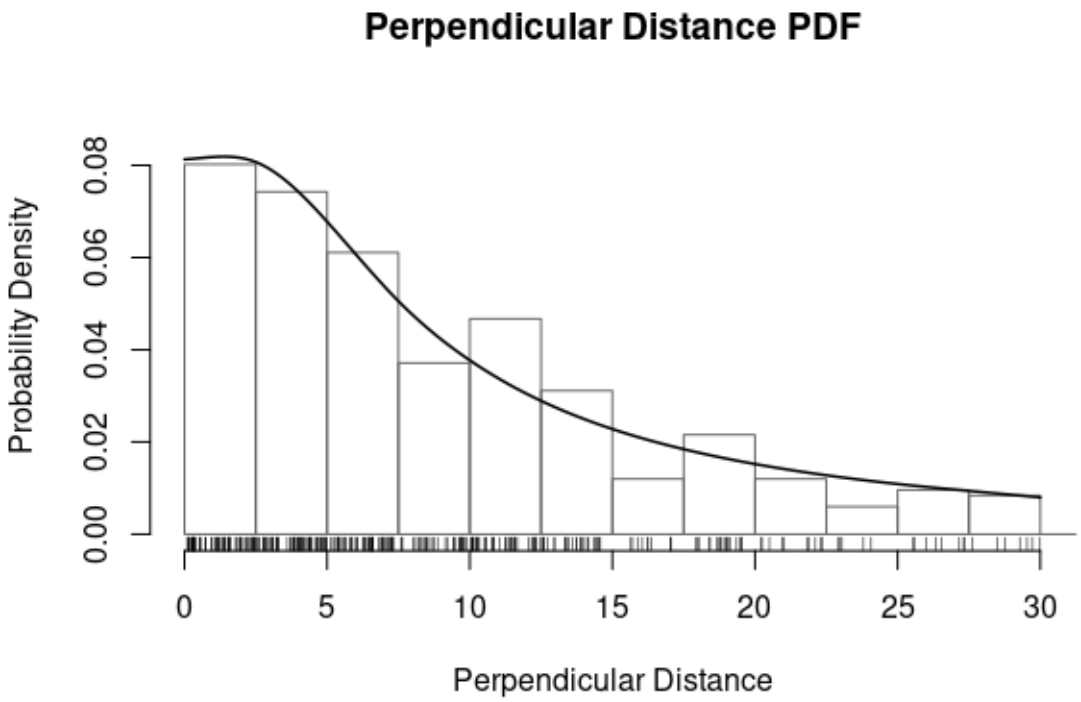
```
cds2d <- mds(ds, move, start = c(s = 4, d = 3))
```

```
summary(cds2d)

# Distance Sampling with movement model analysis
# Number of observations:  334
# Truncation distance:  30
#
# Detection Model: Isotropic radial hazard
# Movement: None
#
# Parameter Estimates:
#   Estimate     SE      LCL      UCL
# s   3.0685 0.0478   2.7940   3.3700
# d   2.4868 0.0246   2.3696   2.6099
```

```
# N 133.2640 9.0922 115.4436 151.0844
# D   0.0001 0.0000   0.0001   0.0002
#
# Mean detection probability:  0.0251
# Loglik:  -4761.92   AIC:  9527.83
```

The abundance estimate is 133.26. The probability of detection in the survey area is 0.025. To plot the fit of the model, you can use the plot command:

```
plot(cds2d)
```



The goodness-of-fit of the model in the perpendicular can be tested by the Kolmogorov-Smirnov test:

```
cds2d.gof <- mds.gof(cds2d)
```

```
cds2d.gof

# $x.ks
#
#  One-sample Kolmogorov-Smirnov test
#
# data:  abs(ddata[, 3])
# D = 0.047847, p-value = 0.429
# alternative hypothesis: two-sided
#
#
# $y.ks
#
#  One-sample Kolmogorov-Smirnov test
#
# data:  ydat
# D = 0.062665, p-value = 0.1658
# alternative hypothesis: two-sided
```

The fit of the model appears to be adequate.

MDS model

The MDS model allows individuals to move by Brownian motion. The data is setup similarly to the CDS 2D model. The `aux` variable is unchanged.

```
aux <- c(1000, 1000, 30, 1, 0)
```

The `ds` object is changed to use a `buffer` around the transect and to allow individuals to move during the survey with the `move` variable. Finally, I set a time-step so that individual movement can be approximated through time using the variable `delta`.

```
ds <- list(data = obs,
           transect = trans,
           delta = c(2.5, 1),
           aux = aux,
```

```
            buffer = 5,
            hazardfn = 1,
            move = 1)
```

What is the `buffer` variable for? Space is discretised into cells, in particular, the space inside the transect is split into grid cells. But, when individuals move they can enter and exit the transect; movement outside the transect must be accounted for. To do this, a boundary of grid cells is added that contains all individuals outside the transect area; individuals then move from these boundary cells into the transect by assuming a uniform distribution of individuals relative to the transect. This approximation has a small affect when the detection probability is small or zero from the boundary outward. In some cases, the detection probability may still be significant near the edge of the transect, so I can add a buffer around the transect to ensure that the boundary of the grid is at a distance where detection probability is very low.

The second object `move` contains the movement data.

```
move <- list(data = line_example_movedat)
```

If I had no movement data, I could fix the movement parameter to be a specific value.

```
move <- list(fixed.sd = 2)
```

The model is fit similarly to CDS2D. For starting values, you set the detection parameters and then the movement parameter. I fit the model using the `mds` function. This can take time, depending on your computer and software setup.

Fitting the model:

```
mds2d <- mds(ds, move, start = c(s = 4, d = 3, sd = 2.5))
```

```
summary(mds2d)

# Distance Sampling with movement model analysis
# Number of observations:  334
# Truncation distance:  30
#
# Detection Model: Isotropic radial hazard
```

```
# Movement: Estimated diffusion
#
# Parameter Estimates:
#    Estimate     SE     LCL      UCL
# s    5.3641 0.0404  4.9559   5.8059
# d    3.1305 0.0286  2.9600   3.3109
# sd   2.4088 0.0219  2.3076   2.5144
# N   98.9954 6.1910 86.8612 111.1296
# D    0.0001 0.0000  0.0001   0.0001
#
# Mean detection probability:  0.0337
# Loglik:  -8202.74   AIC:  16411.47
```

The estimated abundance is 99. The CDS2D estimate was 35% larger. The detection probability for the MDS model was much larger 0.0337 than the CDS model.

These data were simulated with true parameters $N = 100$, $s = 5$, $d = 3$, and $sd = 2.5$.

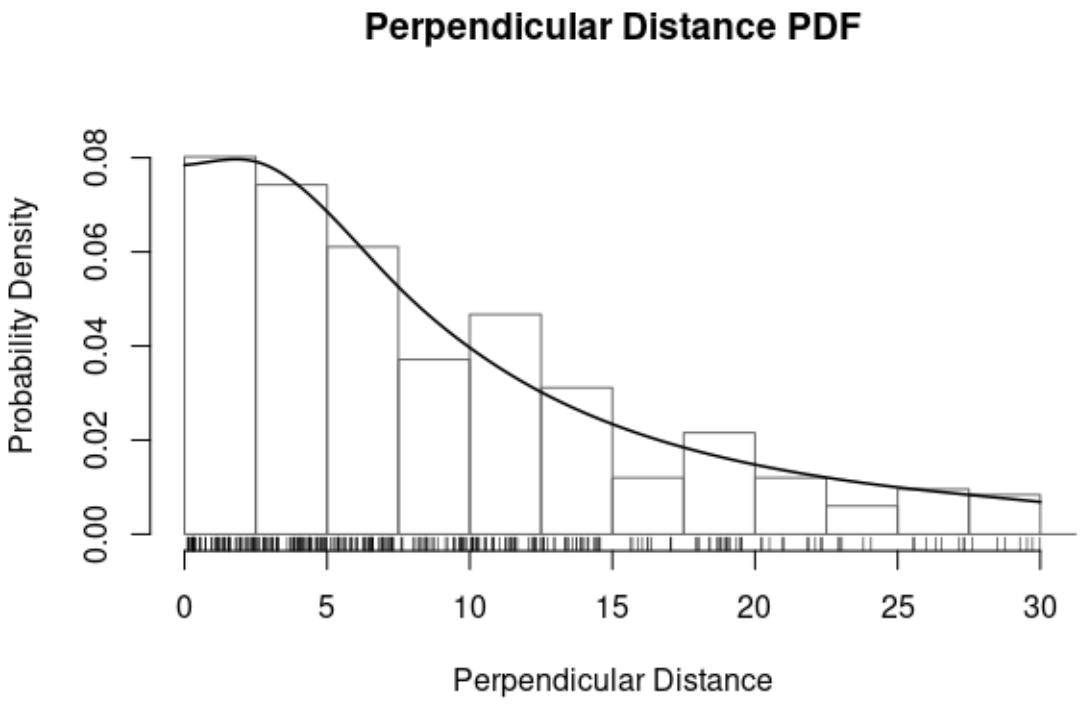
Again, I can plot the fitted PDF for the MDS model:

```
plot(mds2d)
```

Goodness-of-fit also appears adequate:

```
mds2d.gof <- mds.gof(mds2d)
```

```
mds2d.gof

# $x.ks
#
#  One-sample Kolmogorov-Smirnov test
#
# data:  abs(ddata[, 3])
# D = 0.032386, p-value = 0.8749
# alternative hypothesis: two-sided
#
#
```



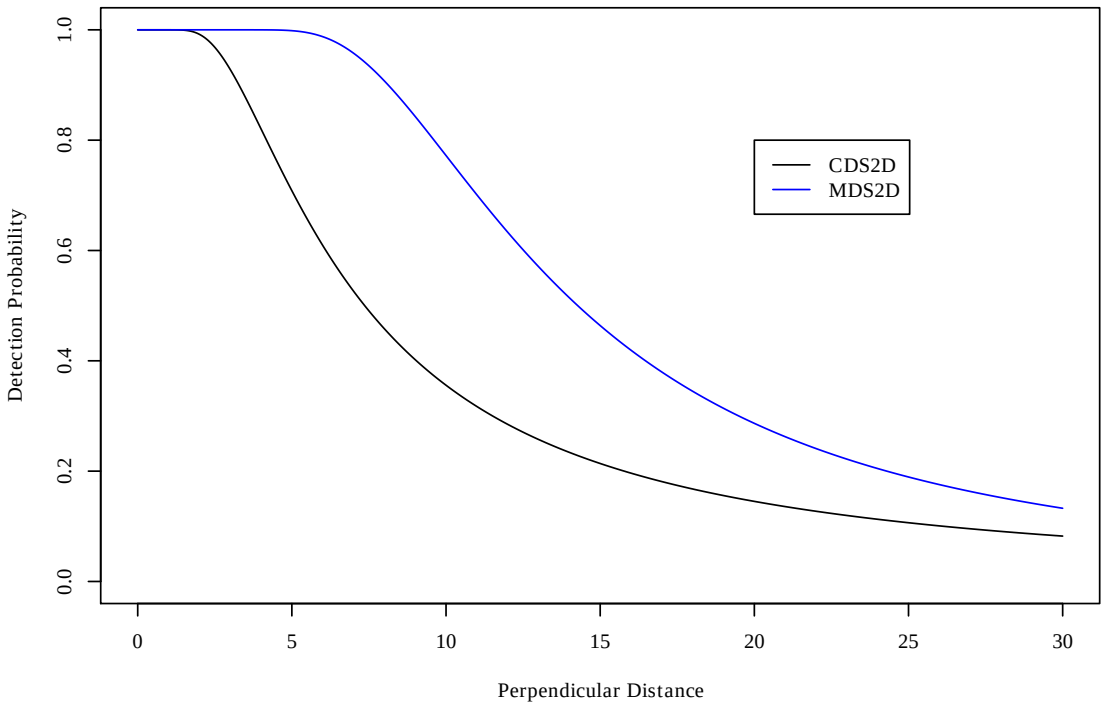

```
# $y.ks
#
#  One-sample Kolmogorov-Smirnov test
#
# data:  ydat
# D = 0.033269, p-value = 0.8743
# alternative hypothesis: two-sided
```

When using the isotropic hazard function, the function `s2sigmab` can be used to convert the estimated s and d parameters from the `mds` function to the b and σ parameters reported in the `Distance` package.

The estimated detection functions from the CDS and MDS methods are markedly different:

```
x <- seq(0, 30, 0.01)

g <- function(x, b, sigma) {1 - exp(-(x/sigma)^(-b))}
```

```
est.cds2d <- s2sigmab(cds2d$result[1,1], cds2d$result[2,1])
g.cds2d <- g(x, est.cds2d[1], est.cds2d[2])

est.mds2d <- s2sigmab(mds2d$result[1,1], mds2d$result[2,1])
g.mds2d <- g(x, est.mds2d[1], est.mds2d[2])

plot(x, g.cds2d,
     type = "l",
     xlab = "Perpendicular Distance",
     ylab = "Detection Probability",
     ylim = c(0, 1))
lines(x, g.mds2d, col = "blue")

legend(20, 0.8, c("CDS2D", "MDS2D"), col = c("black", "blue"), lty = 1)
```



MDS2D has a estimated detection function that has a wider shoulder and slopes off less than the CDS detection function; if the individuals did not move, CDS methods would estimate

the detection function as described by MDS. The problem is that when individuals do move, this is not recognised in CDS model formulation, fitting, or validation, and the estimated detection function, biased from movement, is used.

B2: Point Transects

The analysis for point transects using the `moveds` package is similar to that for line transects. I load some simulated point transect data in the required format.

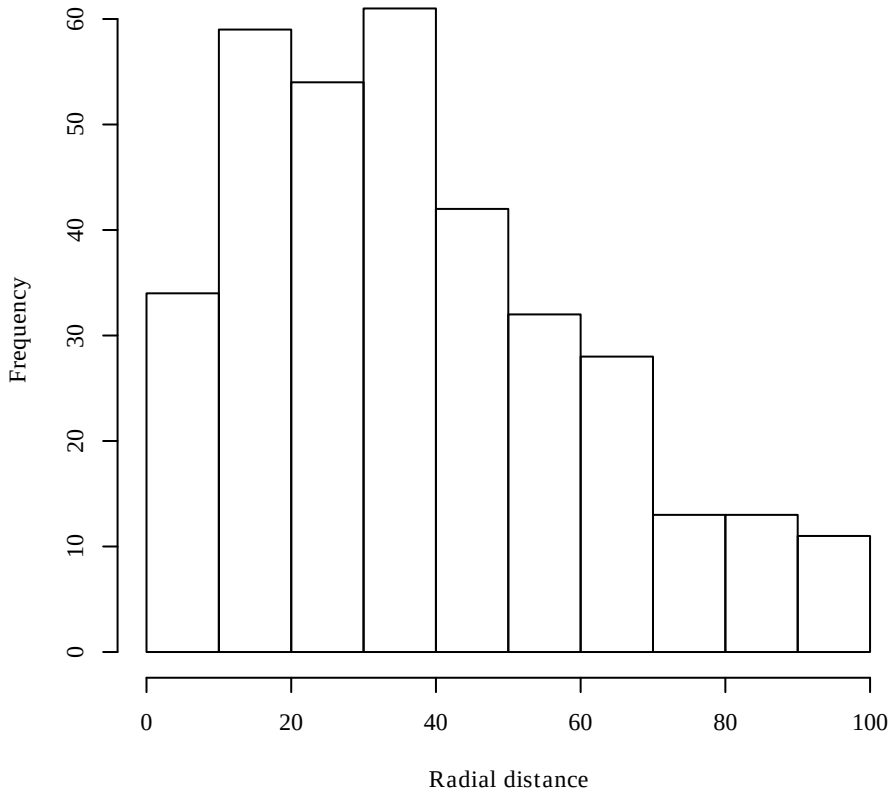
```
data("point_example_dsdat")
str(point_example_dsdat)

# List of 2
#  $ obs  :'data.frame': 347 obs. of  4 variables:
#   ..$ transect: int [1:347] 1 1 1 2 2 2 2 3 3 3 ...
#   ..$ x       : num [1:347] -17.76 11.289 11.222 -3.639 0.301 ...
#   ..$ y       : num [1:347] -37.51 16.84 -58.25 -1.72 15.57 ...
#   ..$ t       : num [1:347] 4.799 66.741 269.055 0.436 7.967 ...
#  $ trans: num [1:100, 1:2] 1 2 3 4 5 6 7 8 9 10 ...

obs <- point_example_dsdat$obs
trans <- point_example_dsdat$trans
```

The object `obs` contains the records of detections made during the point transect survey. There was 347 detections made within point transects of radius 100 distance units.

Let's look at these radial distances.

```
distances <- sqrt(obs$x^2 + obs$y^2)
```

```
hist(distances, main = "", xlab = "Radial distance")
```



The object `trans` contains the transect ID numbers and the time spent surveying each point, termed the effort. In this simulation, all transects had equal duration. In total, the survey consisted of 100 transects.

For this analysis, I use simulated tag data, stored in the object `movedat`

```
data("point_example_movedat")
str(point_example_movedat[[1]])

#  num [1:1001, 1:3] 26.4 26.3 27 27.2 26.5 ...
#  - attr(*, "dimnames")=List of 2
#   ..$ : NULL
#   ..$ : chr [1:3] "x" "y" "observation.times"
```

CDS 2D

For a 2D CDS model, I, again, specify the `aux` variable: the region width and length, the truncation distance, the observer's average speed (in point transect surveys, this is zero), and the transect type (0 for line transects, 1 for point transects).

```
aux <- c(1000, 1000, 100, 0, 1)
```

The `ds` and `move` objects are setup similarly to the line transect example.

```
ds <- list(data = obs,
           transect = trans,
           aux = aux,
           delta = c(5, 5*60),
           buffer = 0,
           hazardfn = 1,
           move = 0)
```

```
move <- list(data = NULL)
```

To fit the model, you must give it starting values: the first value is always the detection scale and the second the detection shape.

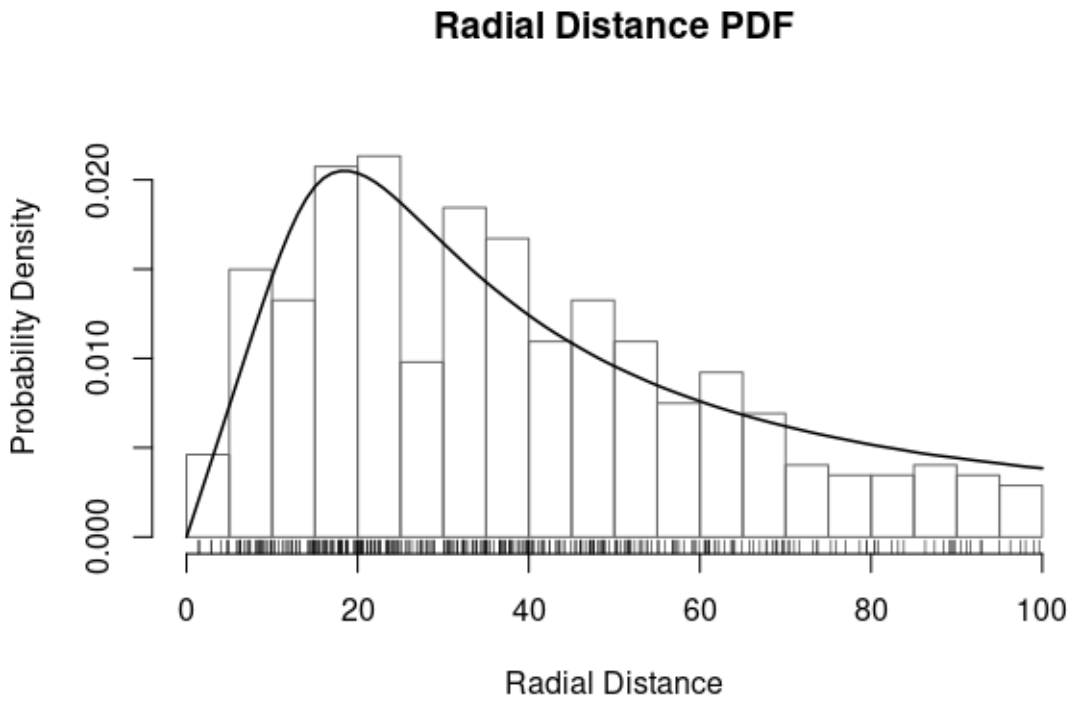
```
cds2d <- mds(ds, move, start = c(s = 4, d = 3))
```

```
summary(cds2d)

# Distance Sampling with movement model analysis
# Number of observations:  347
# Truncation distance:  100
#
# Detection Model: Isotropic radial hazard
# Movement: None
#
# Parameter Estimates:
#   Estimate      SE      LCL      UCL
```

```
# s   2.3476  0.0813   2.0016   2.7533
# d   2.4782  0.0334   2.3210   2.6461
# N 747.8614 72.4484 605.8652 889.8576
# D   0.0007  0.0001   0.0006   0.0009
#
# Mean detection probability:  0.0046
# Loglik:  -5239.96   AIC:  10483.93
```

The abundance estimate is 747.86. The probability of detection in the survey area is 0.005. To plot the estimated PDF, you can use the `plot` command.

```
plot(cds2d)
```



The goodness-of-fit of the model by the Kolmogoriv-Smirnov test:

```
cds2d.gof <- mds.gof(cds2d)
```

```
cds2d.gof

# $r.ks
#
#  One-sample Kolmogorov-Smirnov test
#
# data:  rdat
# D = 0.041804, p-value = 0.5791
# alternative hypothesis: two-sided
```

The fit of the model appears to be adequate.

MDS model

Again, the `ds` object is changed to use a `buffer` around the transect and to allow individuals to move during the survey with the `move` variable. Finally, I set a time-step so that individual movement can be approximated through time using the variable `delta`.

```
ds <- list(data = obs,
           transect = trans,
           aux = aux,
           delta = c(5, 1),
           buffer = 5,
           hazardfn = 1,
           move = 1)
```

The second object `move` contains the movement data.

```
move <- list(data = point_example_movedat)
```

I fit the model using the `mds` function:

```
mds2d <- mds(ds, move, start = c(s = 4, d = 3, sd = 1))
```

```
summary(mds2d)

# Distance Sampling with movement model analysis
# Number of observations:  347
# Truncation distance:  100
#
# Detection Model: Isotropic radial hazard
# Movement: Estimated diffusion
#
# Parameter Estimates:
#    Estimate      SE      LCL      UCL
# s    4.4676  0.0727   3.8746   5.1515
# d    2.9234  0.0340   2.7349   3.1249
# sd   1.0046  0.0050   0.9948   1.0145
# N  500.0186 40.4806 420.6782 579.3591
# D    0.0005  0.0000   0.0004   0.0006
#
# Mean detection probability:  0.0069
# Loglik:  -33722.9   AIC:  67451.79
```

```
mds2d.N <- mds2d$result[4,1]
```

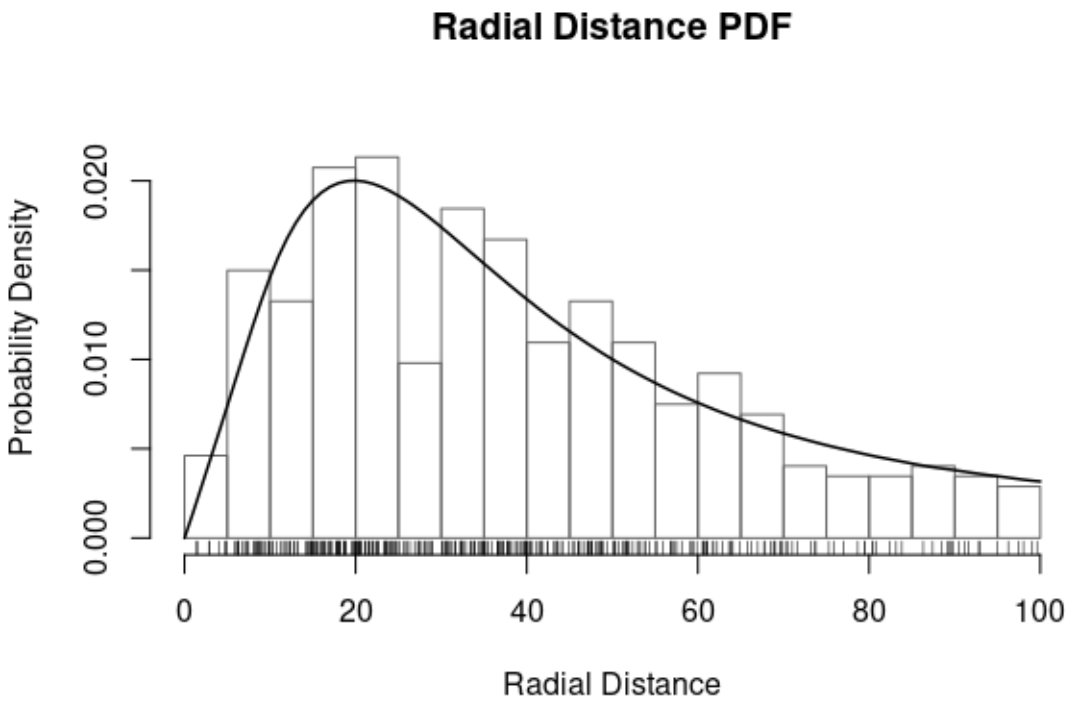
The estimated abundance is 500.02. The CDS2D estimate is 50% larger. The detection probability for the MDS model was larger 0.0069 than the CDS model.

These data were simulated with true parameters $N = 500$, $s = 5$, $d = 3$, and $sd = 1$.

Again, I can plot the fitted PDF for the MDS model:

```
plot(mds2d)
```

Goodness-of-fit also appears adequate:




```
mds2d.gof <- mds.gof(mds2d)
```

```
mds2d.gof

# $r.ks
#
#  One-sample Kolmogorov-Smirnov test
#
# data:  rdat
# D = 0.02716, p-value = 0.96
# alternative hypothesis: two-sided
```

B3: Spotted Dolphin Application

In Section 3.5, MDS is applied to the Eastern Tropical Pacific survey on a population of spotted dolphins. Only surveying and detections in conditions of Beaufort state two or

less were used in the analysis. The algorithms presented in Appendix A were implemented as in the `moveds` package; however, the implementation was adapted to account for the interrupted effort during the ETP surveys and the varying speed of the observer.

Frequently, throughout the surveying, observers would go off effort for short periods of time; usually, this is to approach a detected dolphin school and improve species identification and school size estimation. Surveying is then resumed thereafter. To account for this, Algorithm 10 was adapted. A vector $\boldsymbol{e}$ contained a one or zero for each time-step, indicating whether the observer was on (one) or off (zero) effort. Line 17 of Algorithm 10 was then amended as

$$\textbf{If } \boldsymbol{e}_t = 1 \textbf{ then } \boldsymbol{\psi} \leftarrow \textbf{Weight}(t, \boldsymbol{\psi}, \boldsymbol{\theta}_d, v, \Delta x, \Delta t)$$

That is, the probability density is weighted by the detection process only when the observer is on effort. Thus, when the observer is off effort for a short period of time, the probability density is still updated and so movement of individuals during that time is accounted for.

The model was fit with dx = 0.25 kilometres and a time-step dt of 1 minute. To assess stability of the approximation, the estimated abundance was re-calculated for the maximum likelihood estimates over a range of values for dx. An estimate of the percentage relative error in the approximation was calculated by taking the estimated abundance for the smallest dx as the true value. Figure S1 shows the estimated percentage relative error against spatial discretisation. The time-step was also halved and the estimated abundance changed by less than 1%.

The fit of the movement model to the nineteen spotted dolphin tags was assessed by computing residuals from the cumulative distribution function. Figure S2 shows the quantile-quantile plots for the residuals.

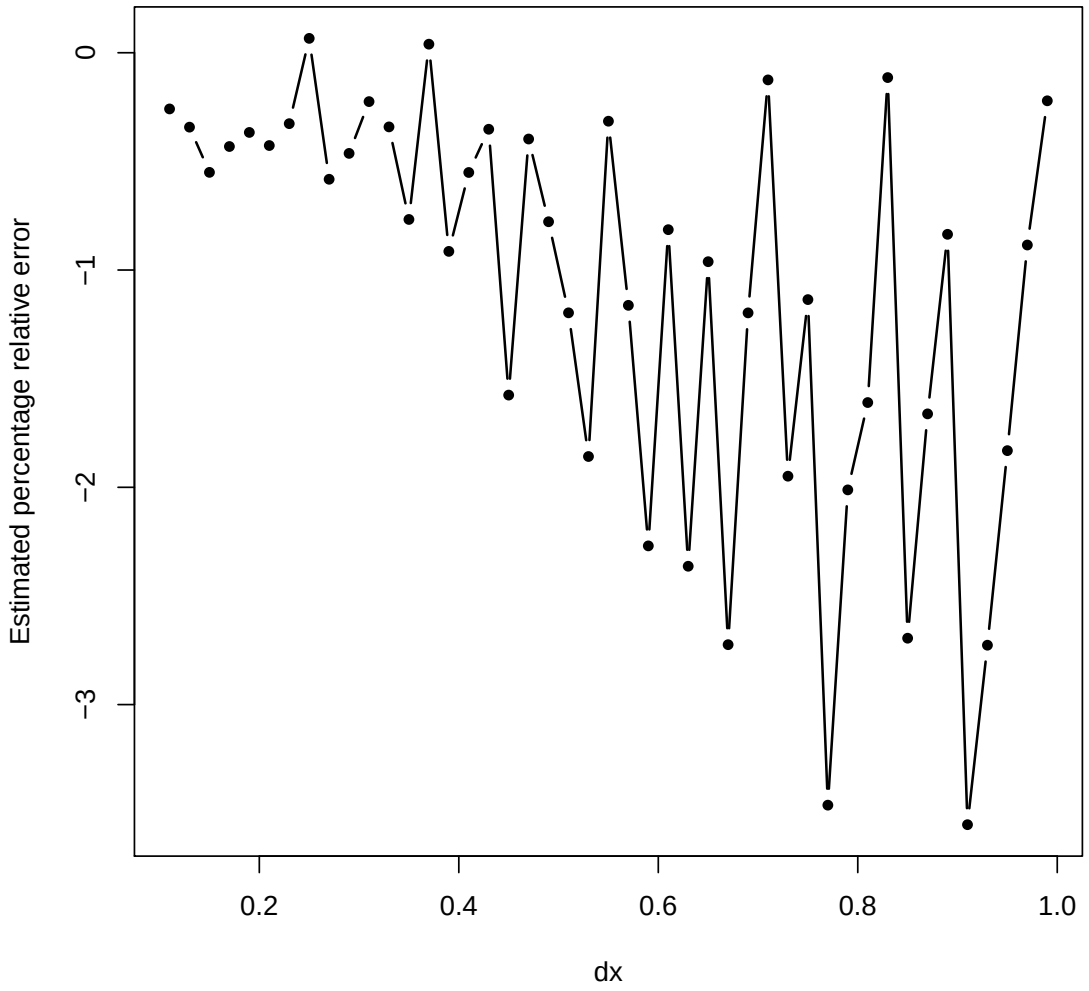


Figure S1 *Estimated percentage relative error in approximation against spatial grid size for spotted dolphin case study.*

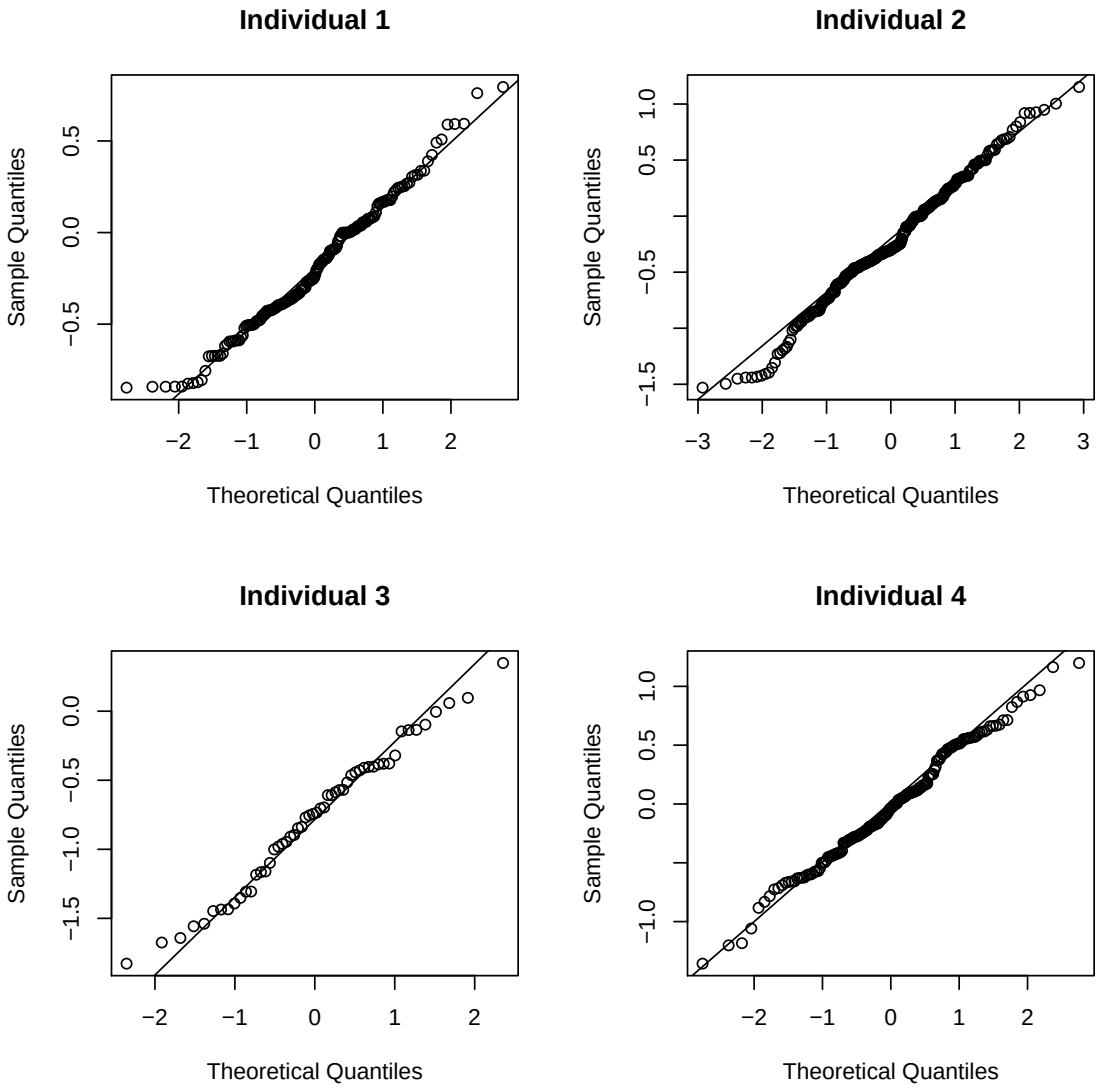


Figure S2 *Quantile-quantile plot of residuals from the movement model fitted to tag data on nineteen spotted dolphins.*

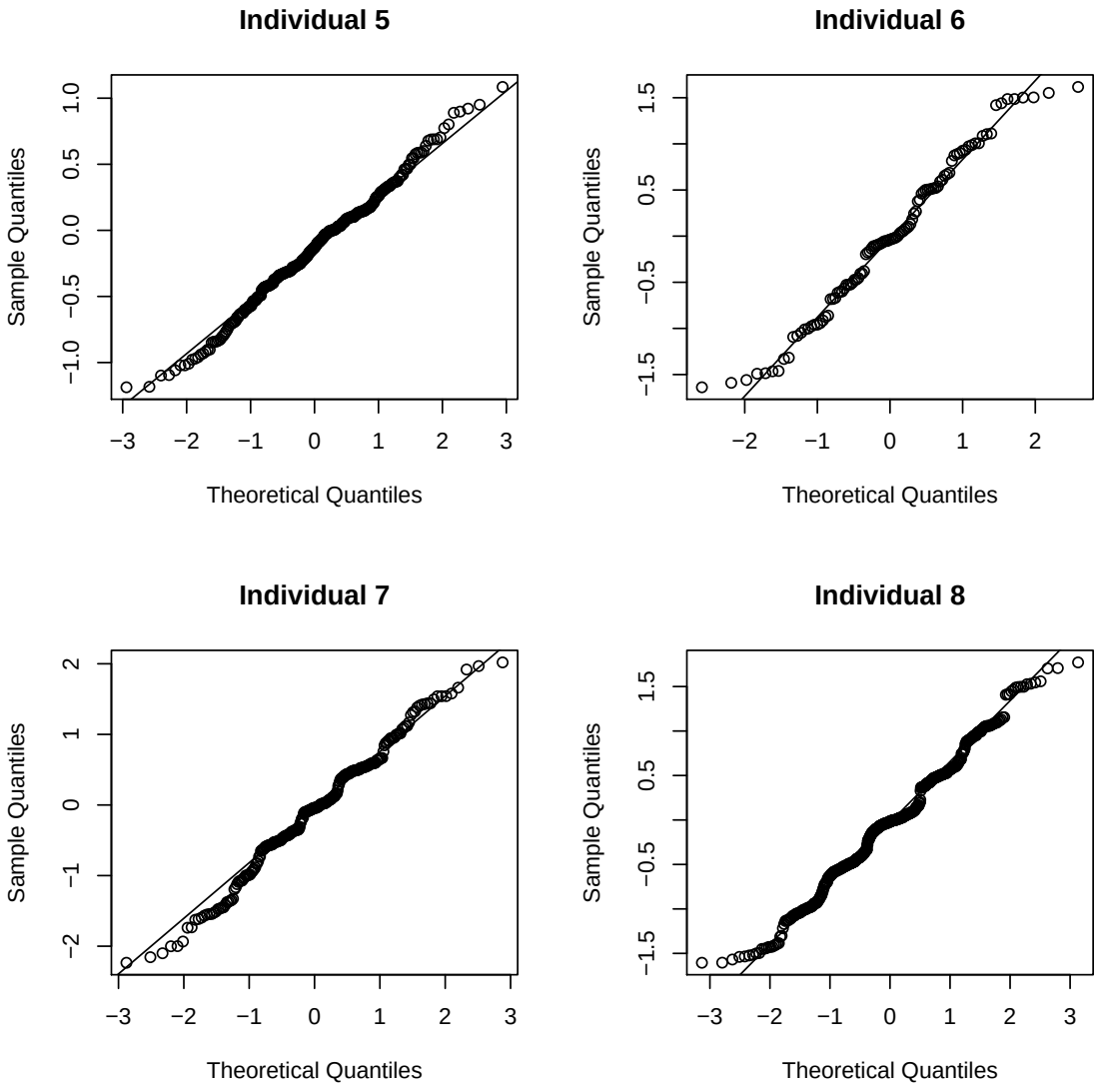


Figure S2 *Quantile-quantile plot of residuals from the movement model fitted to tag data on nineteen spotted dolphins.*

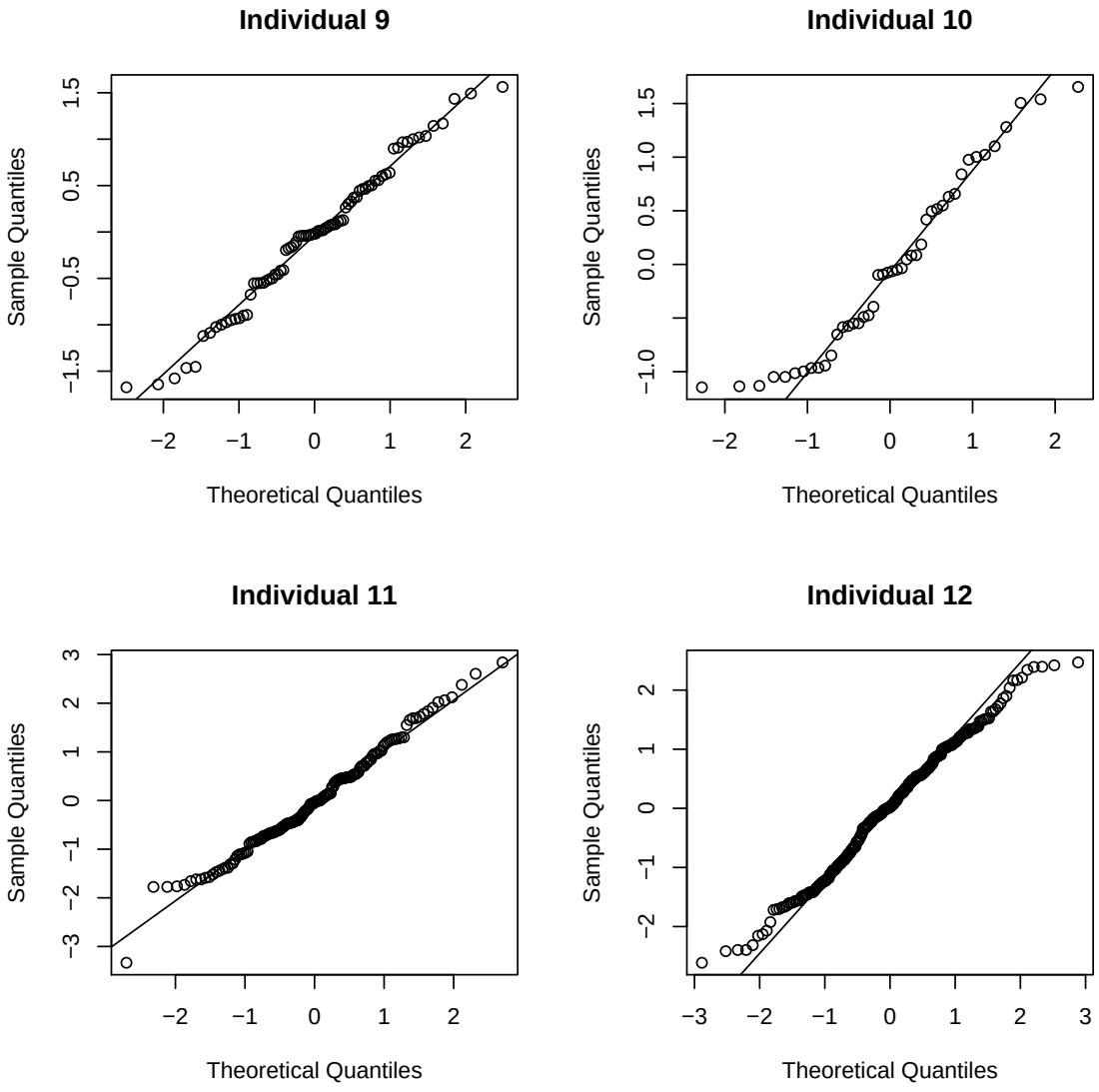


Figure S2 *Quantile-quantile plot of residuals from the movement model fitted to tag data on nineteen spotted dolphins.*

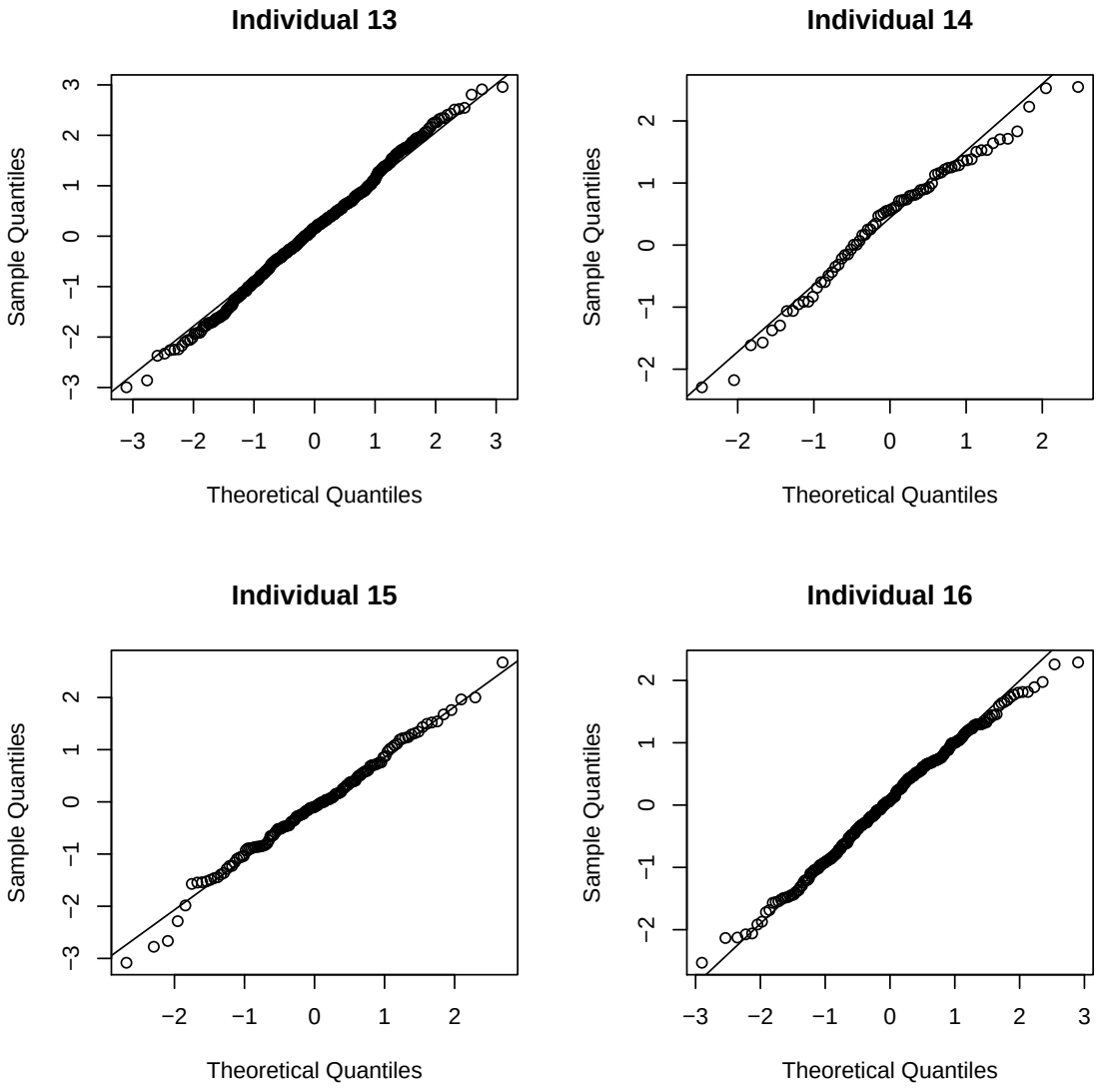


Figure S2 *Quantile-quantile plot of residuals from the movement model fitted to tag data on nineteen spotted dolphins.*

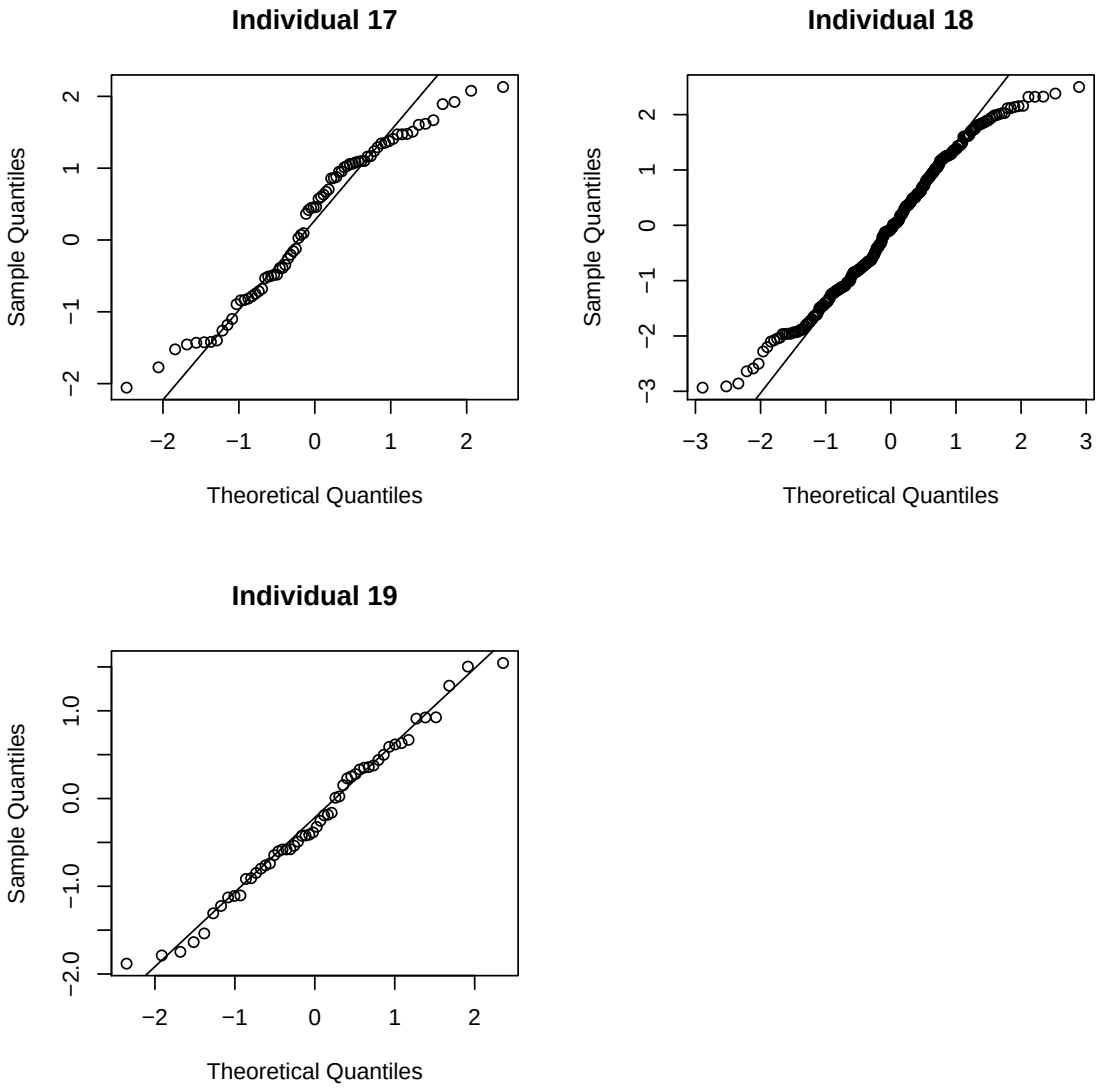


Figure S2 *Quantile-quantile plot of residuals from the movement model fitted to tag data on nineteen spotted dolphins.*

B4: Shearwater Application

In Section 3.7.1, a behaviour-switching MDS model is applied to line transect surveys on a population of Balearic Shearwaters in the Bay of Biscay. Behaviour-switching MDS models are not yet implemented in the `moveds` package. Algortihm 10 was adapted in a similar way to Algorithm 2 to compute the behaviour-switching model. There was few breaks in effort during the survey of a transect and so no adaptation for interrupted effort was used.

Algorithm 16 was implemented in `R` and `C++98` using `Rcpp` and `armadillo`. The probability density $\boldsymbol{\phi}$ is a matrix where each row is the probability density over space for a particular behaviour. In the Shearwater survey, behaviours were assumed to be observed as flying or on the sea surface. The transition rate matrix is computed using Algorithm 5 where each behaviour exhibits a different rate of Brownian motion. In Algorithm 16, the **Vectorise** and **UnVectorise** steps, as described in Algorithm 2, are omitted for brevity.

The initial distribution is a matrix where for each row, representing a behaviour, individuals are assumed to be uniformly distributed across space with respect to the transect; the rows are then weighted by the equilibrum distribution of the behaviour-switching Markov chain. The equilibrum distribution is found using the algorithm given by Zucchini et al. (2016).

The model was fit with dx equal to 50 metres and a time-step of 10 minutes. Similar to the spotted dolphin case study, the abundance estimate was re-calculated at the maximum likelihood estimate for a range of values of dx to assess stability of the approximation. Figure S3 shows the estimated percentage relative error agaisnt spatial discretisation. Havling the time-step also caused a change in estimated abundance of less than 1%.

The goodness-of-fit of the behaviour-switching, velocity-based movement model (Section 3.7.1) was assessed by computing the pseudo-residuals according to Zucchini et al. (2016). The quantile-quantile plots are given in Figure S4.

Algorithm 16: Log-Likelihood for Behaviour-switching Distance Sampling with Movement

Input:

- $\boldsymbol{\theta}_d, \boldsymbol{\theta}_m$: detection and movement parameters respectively where $^{(b)}$ denotes the subset corresponding to behaviour b;
- $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\tau}, \boldsymbol{b}, \boldsymbol{g}$: matrix of perpendicular distances, forward distances, detection times, behaviours and grid cell for ith detection on jth transect;
- $\boldsymbol{T}$: time taken to survey each transect;
- $\boldsymbol{n}$: number of individuals seen on each transect;
- $\boldsymbol{v}$: speed of observer along each transect;
- G: number of grid cells in total and G_x, G_y in each direction;
- B: number of behaviours;
- $\boldsymbol{I}$: vector with 1 for grid cells inside transect and 0 for those outside;
- Δx: spatial grid cell size;
- Δt: time-step.

Output: l, log-likelihood

$l \leftarrow 0$;
$\boldsymbol{R} \leftarrow$**Get Trm**$(G, \boldsymbol{\theta}_m, \Delta x)$;
foreach *transect* j **do**
  $\boldsymbol{\phi} \leftarrow$ CalcInitialDistribution$(G, B, \Delta x)$;
  $\tilde{T}_j \leftarrow \lfloor \frac{\tilde{T}}{\Delta t} \rfloor$;
  $i \leftarrow 1$;
  $h \leftarrow 0$;
  $c \leftarrow$ TRUE;
  $p \leftarrow 0$;
  for $t \leftarrow 0$ **to** $\tilde{T}_j$ **do**
    while c *and* $\tau_{i,j} \leq t\Delta t$ **do**
      $f \leftarrow$Calculate hazard seen at $(\boldsymbol{x}_{i,j}, \boldsymbol{y}_{i,j}, \boldsymbol{\tau}_{i,j})$;
      $l \leftarrow l + \log(f) + \log(\boldsymbol{\phi}(\boldsymbol{b}_{i,j}, \boldsymbol{g}_{i,j})) + h - \log(\Delta x^2)$;
      $i \leftarrow i + 1$;
      if $i >$ *number of individuals seen on transect* j **then** $c \leftarrow$ FALSE;
    end
    foreach *behaviour* b **do** $\boldsymbol{\psi}[b,] \leftarrow$**Weight**$(t, \boldsymbol{\phi}[b,], \boldsymbol{\theta}_d^{(b)}, v, \Delta x, \Delta t)$;
    $p \leftarrow p + \exp(h) \sum_{g=1}^{G} ((\boldsymbol{\phi} - \boldsymbol{\psi}) \circ \boldsymbol{I}_j)$;
    $\boldsymbol{\phi} \leftarrow$**Update**$(t, \boldsymbol{\psi}, \Delta t, \boldsymbol{R}, G)$;
    $s \leftarrow \sum_{i=1}^{G} \boldsymbol{\phi}_i$;
    $h \leftarrow h + \log(s)$;
    $\boldsymbol{\phi} \leftarrow \boldsymbol{\phi}/s$;
  end
  $l \leftarrow l - n_i \log(p)$;
end
return $\boldsymbol{l}$

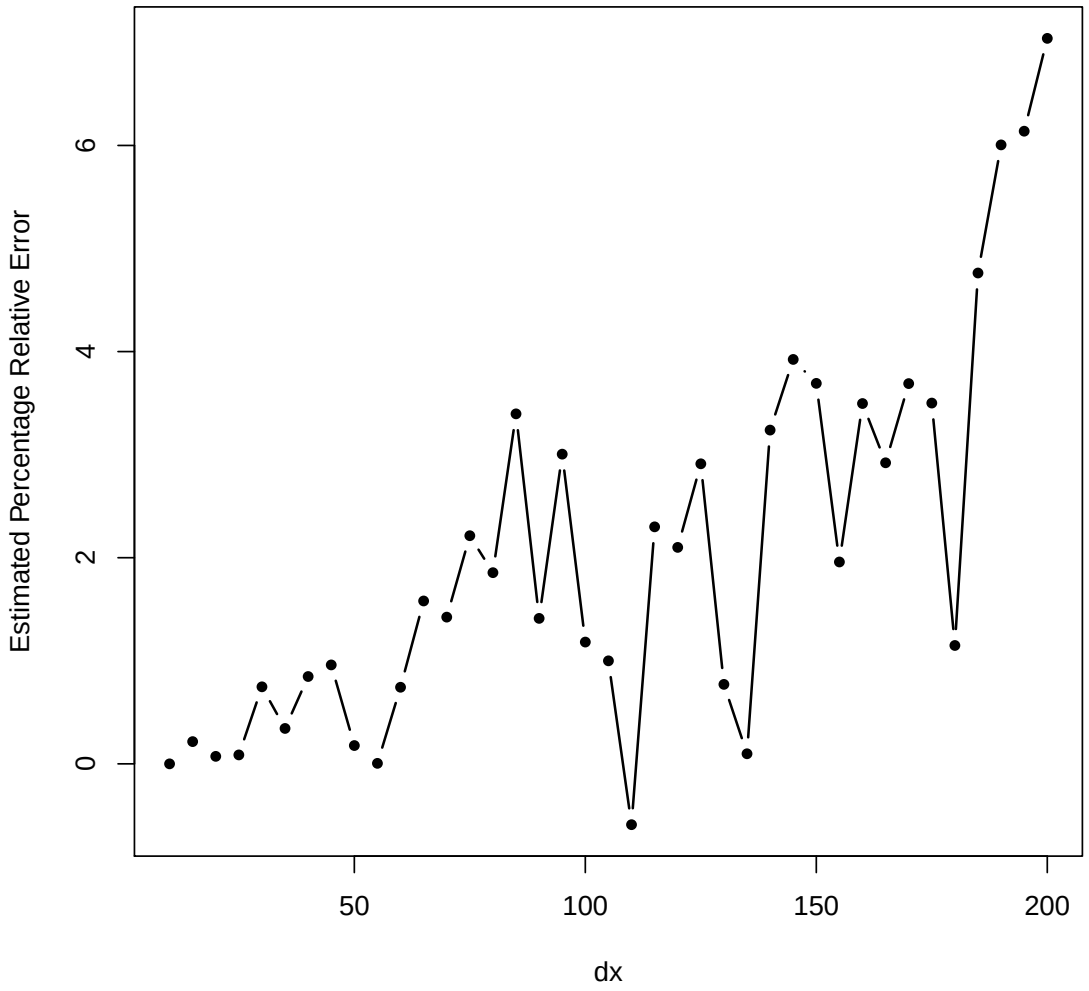


Figure S3 *Estimated percentage relative error in approximation against spatial grid size for shear-water case study.*

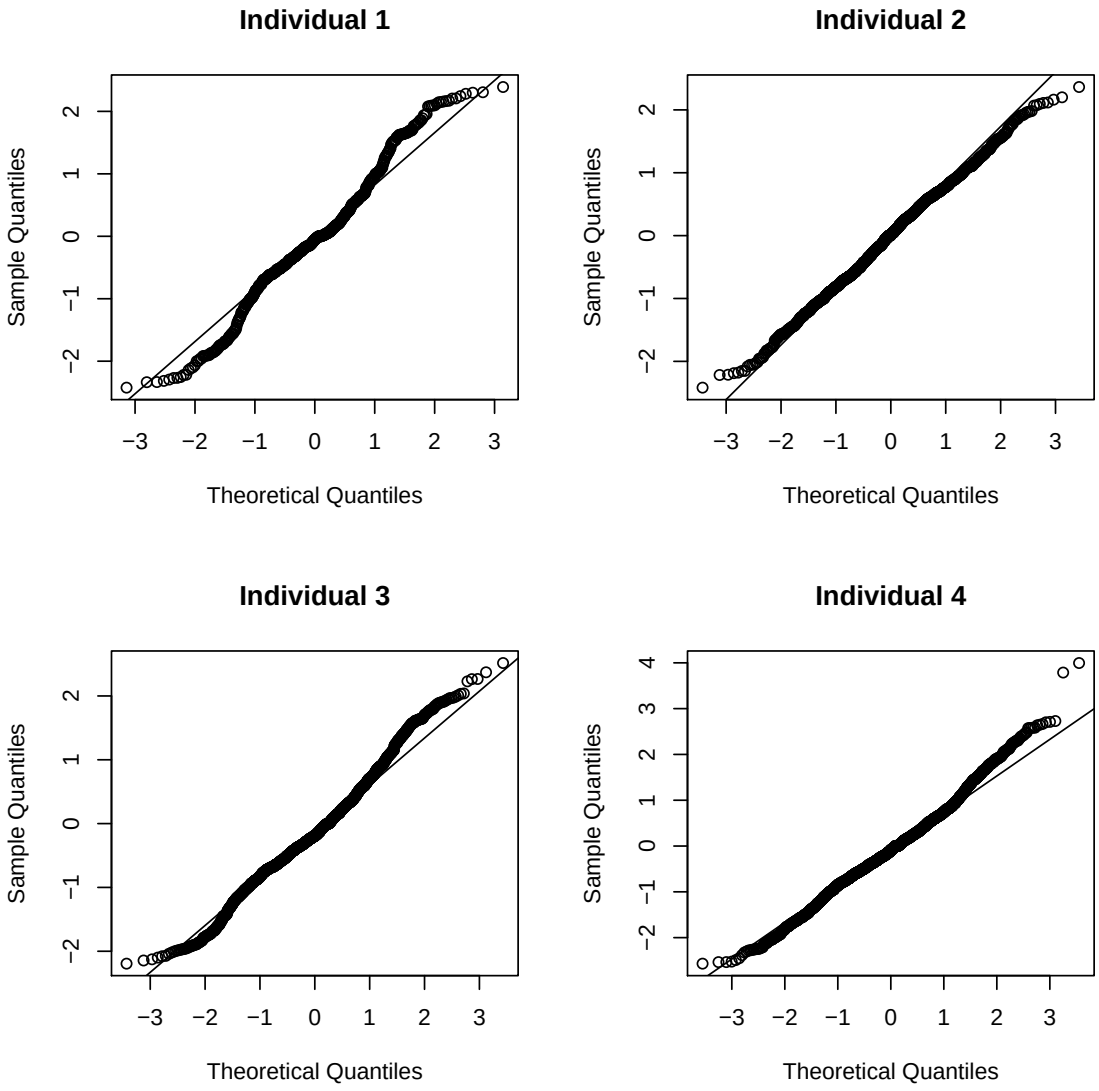


Figure S4 *Quantile-quantile plot of residuals from the movement model fitted to tag data on seven shearwaters.*

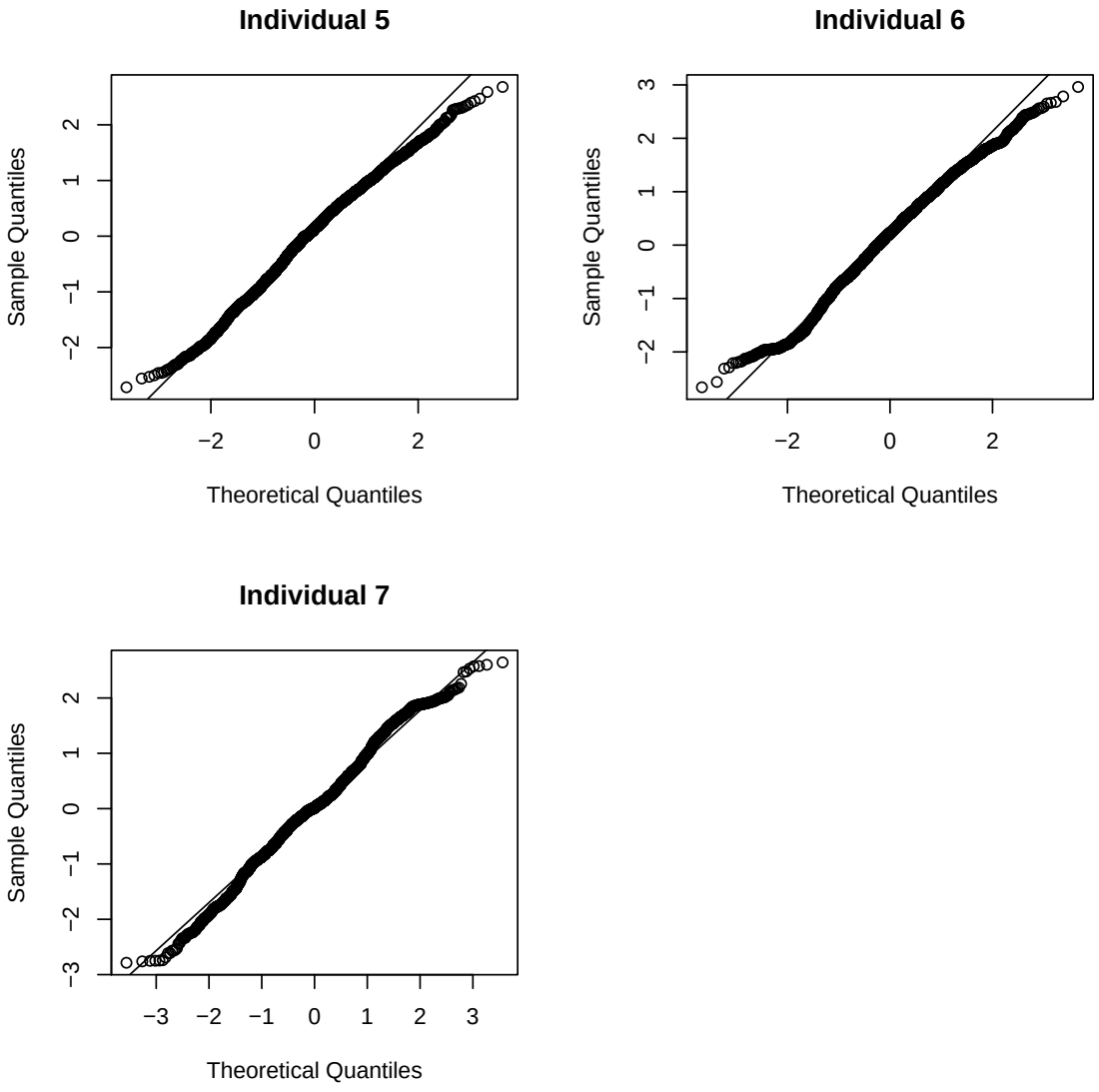


Figure S4 *Quantile-quantile plot of residuals from the movement model fitted to tag data on seven shearwaters.*

Appendix C

In this appendix, the algorithm and computational details for the jaguar case study in Section 4.4 are given. Appendix A, Section 3 gives the key algorithms to compute the likelihood of the continuous-time spatial capture-recapture (SCR) model for a single individual whose activity centre is known.

To estimate the detection, movement, and density parameters along with the unknown activity centre of each individual, a Markov chain Monte Carlo algorithm is used. This algorithm uses data augmentation as described for SCR by Royle and Young (2008).

C1: MCMC Algorithm

Algorithm 17 gives pseudo-code for the MCMC algorithm. The priors are also specified for each parameter: the encounter rate, λ, the movement rate off trails σ_0, the movement rate on trails σ_1, the attraction to activity centres α, and the inclusion parameter ψ.

The idea behind data augmentation is that a meta-population M of individuals are simulated. Each individual exists in the real population with probability ψ. The variable e_i is one when individual i exists in the population and zero otherwise. Hence, the real population size is $N = \sum_{i=1}^{M} e_i$. Clearly, if an individual is captured at least once during the survey, $e_i = 1$. Re-captures of individuals give information on the detection parameters and this determines the likelihood that a certain number of individuals in the meta-population existed and were not captured. This is equivalent to fitting a zero-inflated SCR model where the zeroes are the empty capture histories for the individuals in the meta-population.

For each individual, an activity centre $\boldsymbol{z}$ is sampled uniformly from the study area A. The likelihood for the other parameters can then be computed conditional on the sampled $\boldsymbol{z}$.

The MCMC algorithm used was a simple random walk Metropolis-Hastings sampler (Hastings, 1970; Metropolis, Rosenbluth, Rosenbluth, Teller, & Teller, 1953). Parameters were proposed by Gaussian proposal distributions where the standard deviation was tuned until the mean acceptance rate was approximately 25% (Roberts, Gelman, Gilks, et al., 1997).

The likelihood of the continuous-time SCR model conditional on known activity centres is computed using Algorithm 12. The time-step Δt and grid size Δx were selected as those for which the likelihood over all individuals at the starting values for the parameters changed by less than 1%. The selected values were a grid size of 500 metres and a time-step of 1 hour between April and July 2013.

Algorithm 17: Markov chain Monte Carlo for continuous-time SCR with movement

Input:

- $\boldsymbol{\theta}_d, \boldsymbol{\theta}_m$: detection and movement parameters respectively
- $\boldsymbol{z}$: individual's activity centre
- $\boldsymbol{c}$: capture histories for each individual
- $\boldsymbol{T}$: survey duration
- M: meta-population size
- G: number of grid cells in total and G_x, G_y in each direction
- $\boldsymbol{I}$: vector with 1 for grid cells inside survey area and 0 for those outside
- Δx: spatial grid cell size
- Δt: time-step

Output: l: log-likelihood value

$\lambda \sim \text{U}(0.0001, 0.005)$;
$\sigma_0 \sim \text{U}(0.01, 0.1)$;
$\sigma_1 \sim \text{U}(0.01, 0.1)$;
$\alpha \sim \text{U}(0.0001, 0.001)$;
$\psi \sim U(0, 1)$;
foreach *individual* i **do**
    $e_i \mid \psi \sim \text{Bern}(\psi)$;
    $\boldsymbol{z}_i \sim U(A)$;
    $\boldsymbol{c}_i \mid e_i, \boldsymbol{z}_i \sim$ cts-time SCR model
end
$N \leftarrow \sum_{i=1}^{M} e_i$;

C2: MCMC Output

Four chains were simulated with a meta-population of size 100 and were run for 10000 iterations each. Figure S5 shows the MCMC trace plots for each parameter; visual inspection

suggests convergence occurs after 1000 iterations. To be conservative, a 5000 iteration burn-in was used. Figure S6 shows the Gelman-Rubin shrink factor (Gelman, Rubin, et al., 1992) against the number of iterations of each chain. The potential scale reduction factor for each parameter was less than 1.1 as computed using the `gelman.diag` function in the R package `coda 0.19` (Plummer, Best, Cowles, & Vines, 2006). This is evidence that the chains have converged.

For each iteration of the chain, the activity range for each individual was computed as described in Appendix A, Section 3.2. This provides a posterior for the realised activity range of each individual. The time-spent distributions given in Figure 4.3 of the main text is a plot of the posterior mean for each grid cell. The spatial and spatio-temporal overlap between individuals was computed using the posterior means of each parameter only as it was computationaly prohibitive to compute spatio-temporal overlap for every iteration of the MCMC chain.

C3: Goodness-of-fit

Goodness-of-fit of the movement model to the tag data is assessed by computing pseudo-residuals. Under the assumed movement model, the probability of an individual moving to each location given its current location is given by the HMM algorithm (Algorithm 1). This is termed the redistribution kernel in other contexts (Hooten, Johnson, et al., 2017). The probability of an individual moving to a certain new location depends not only on the distance but also the habitat type of that location and of the intervening space. For the jaguar movement model, locations that are connected to the individual's current location by trails or rivers are more likely than those than are not.

Each recorded movement is then an independent realisation from this distribution and so the cumulative distribution function evaluated for each movement ought to produce pseudo-residuals that are uniformly distributed. Applying the Gaussian quantile function to these ought to produce Gaussian distributed residuals. This provides a way to test the goodness-of-fit of the movement model visually and numerically. Figure S7 shows the quantile-quantile plots of the computed residuals. The Shapiro-Wilks test rejects the hypothesis the residuals for individual 2 are Gaussian, but indicates there is no evidence of non-normality for individual 1. From the quantile-quantile plot, the movement distribution for individual 2 appears to have heavier tails than expected under the model. This may be due to individual 2 not having a single activity centre; the recorded locations of individual 2 indicate that this individual may have multiple centres of attraction.

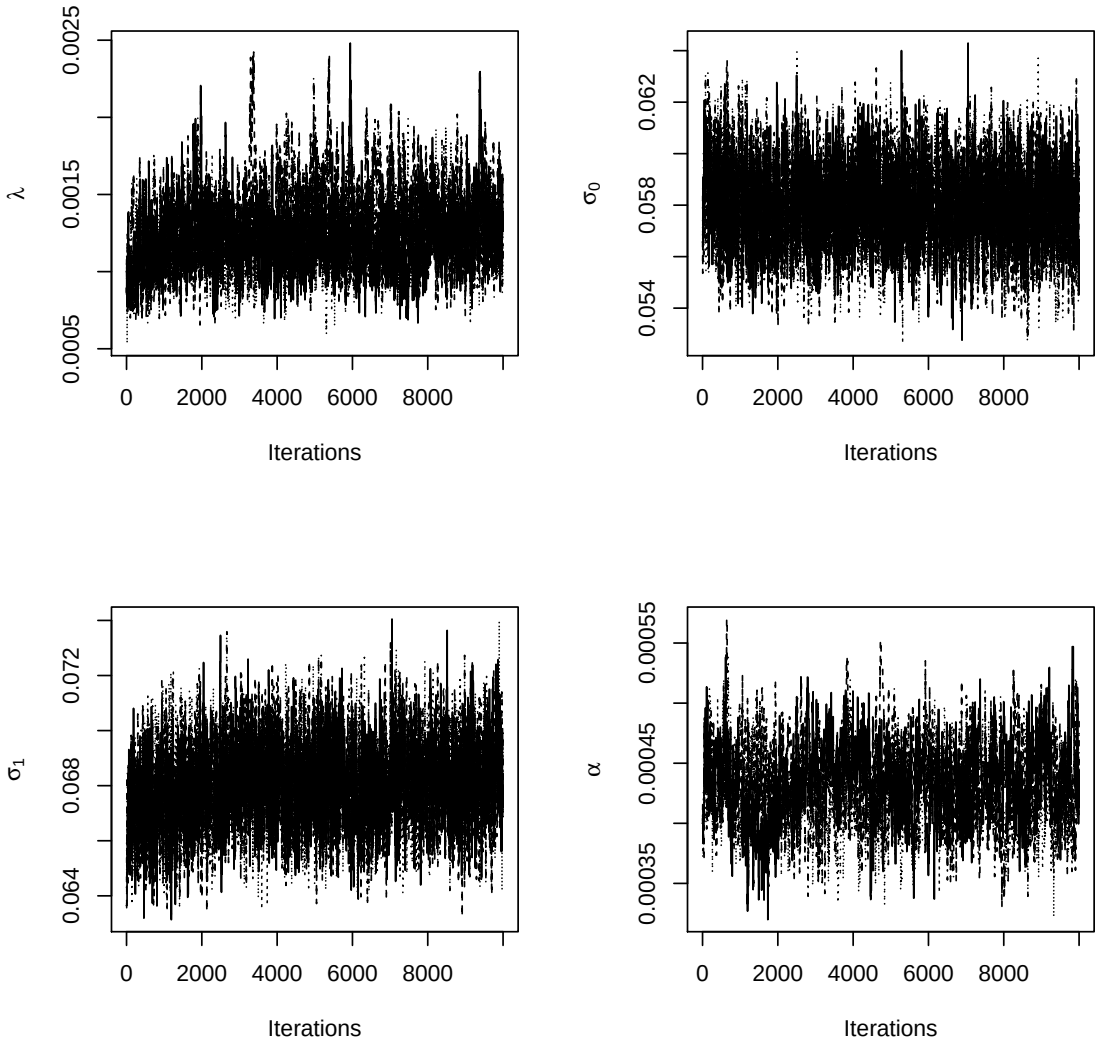


Figure S5 *Value of each parameter against number of iterations of Markov chain for encounter rate λ, movement rate off trail σ_0, movement rate on trail σ_1, and attraction to activity centre α for jaguar case study.*

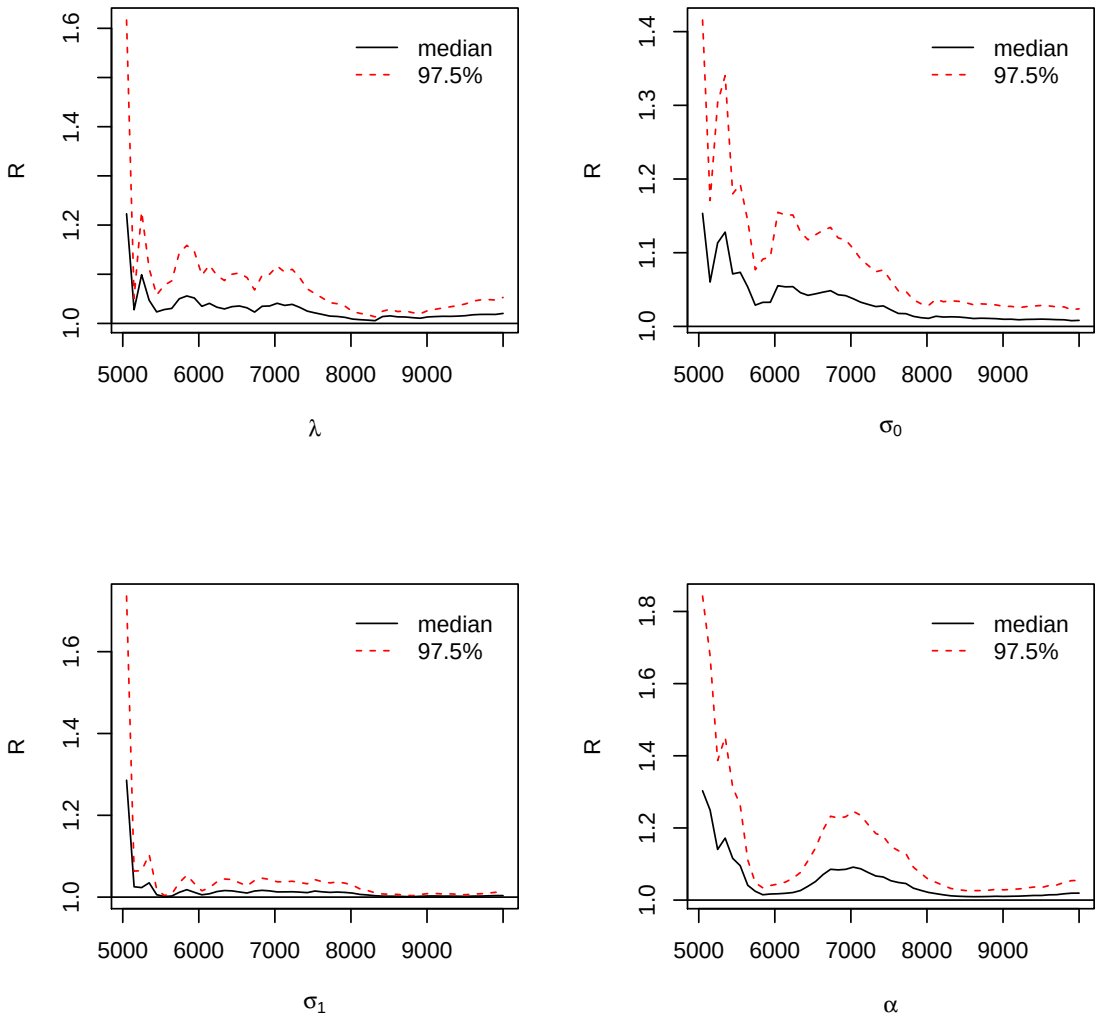


Figure S6 *Gelman-Rubin shrink factor against number of iterations of Markov chain for encounter rate λ, movement rate off trail σ_0, movement rate on trail σ_1, and attraction to activity centre α for jaguar case study.*

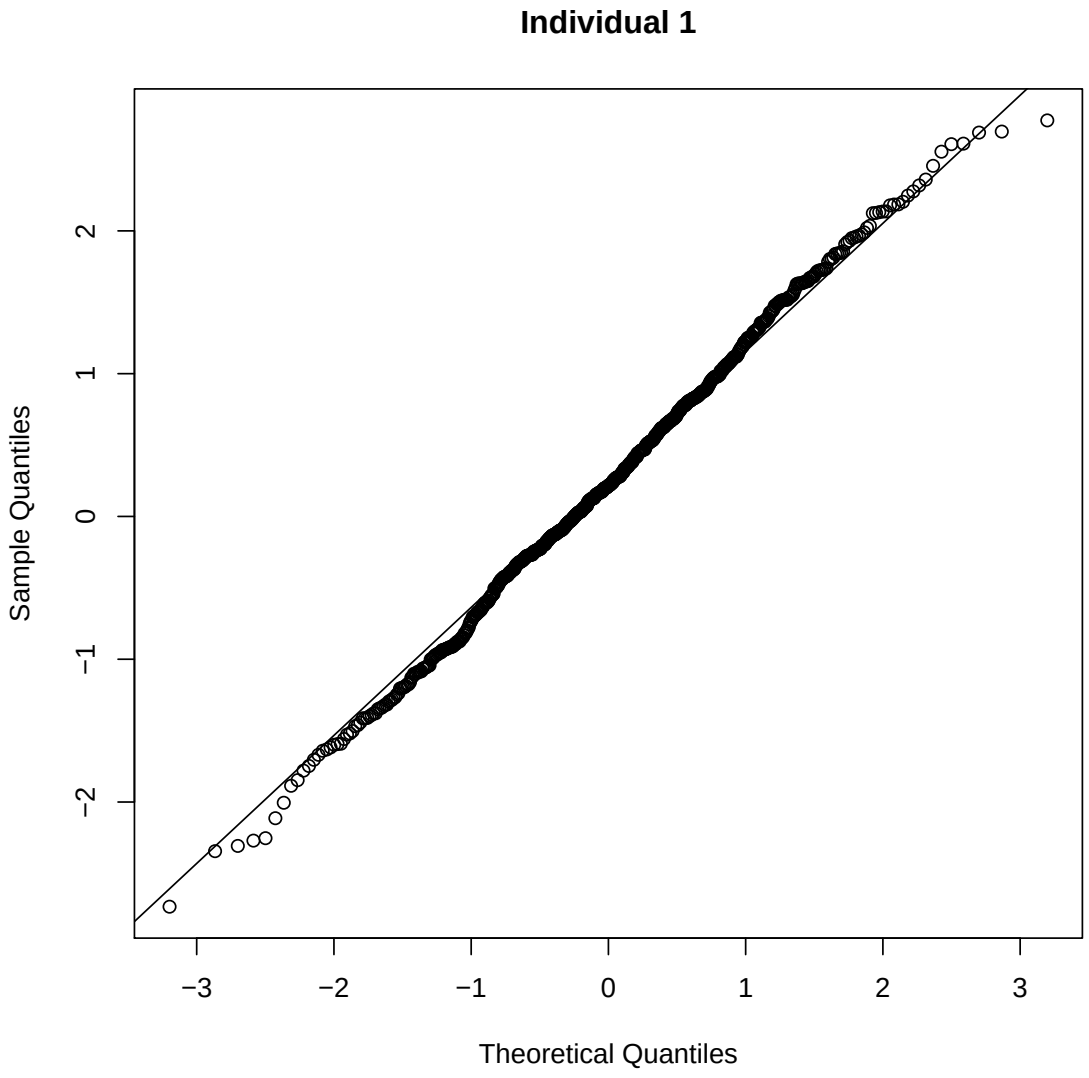


Figure S7 *Quantile-quantile plots of pseudo-residuals from the movement distributions of two tagged jaguars.*

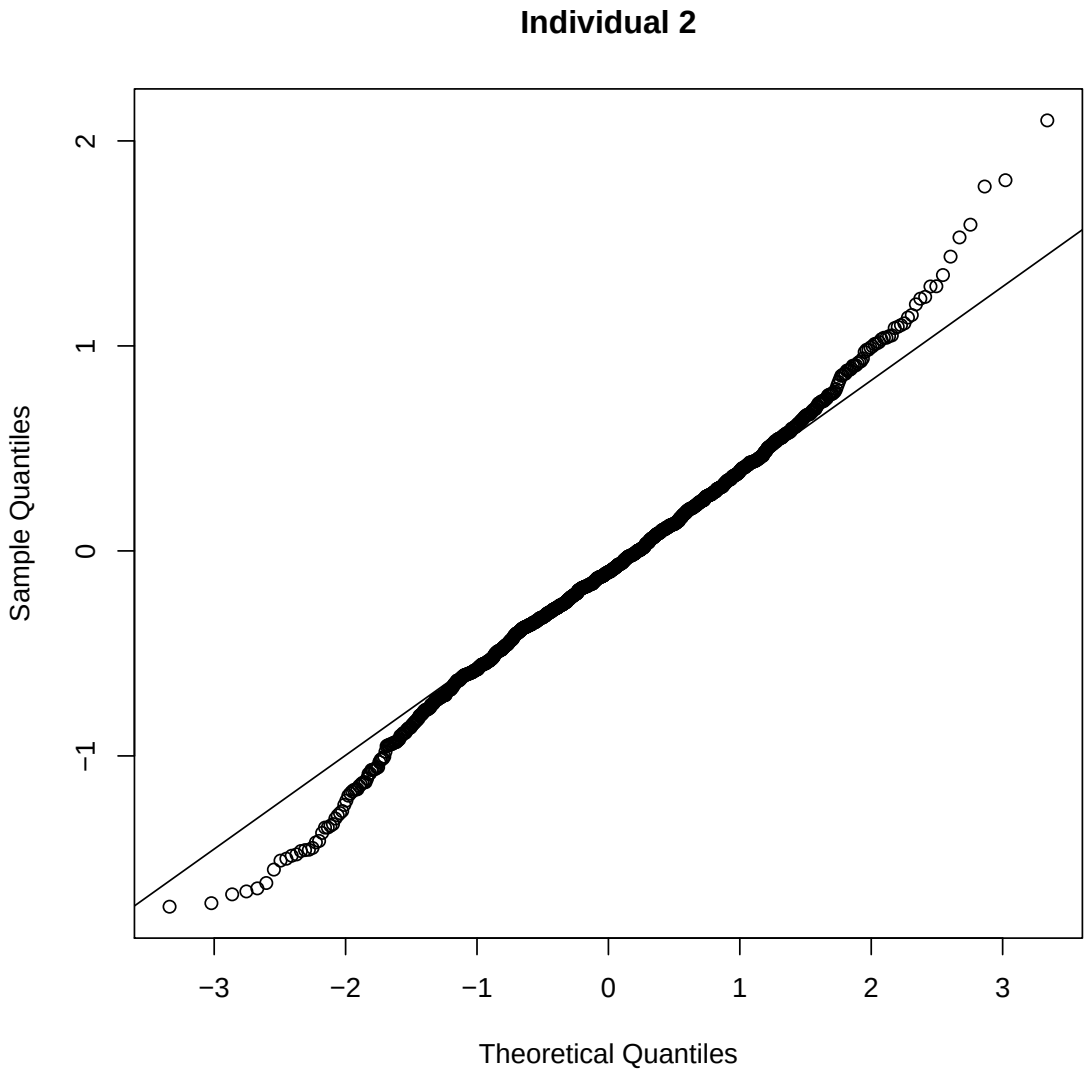


Figure S7 *Quantile-quantile plots of pseudo-residuals from the movement distributions of two tagged jaguars.*

Appendix D

In this appendix, the `R` package `openpopscr v1.1.0` is described. This package implements the discrete-time spatial capture-recapture models described in this thesis. It can be used to fit discrete-time SCR models with moving activity centres and open population SCR models with movement, such as the Cormack-Jolly-Seber model.

Here, two examples are given on how the package can be used for SCR with moving activity centres and to fit Cormack-Jolly-Seber SCR models with moving activity centres.

The package can be installed using the R package `devtools`:

```
devtools::install_github("r-glennie/openpopscr@v1.1.0",
                         build_vignettes = TRUE)
```

It can then be loaded in the usual way:

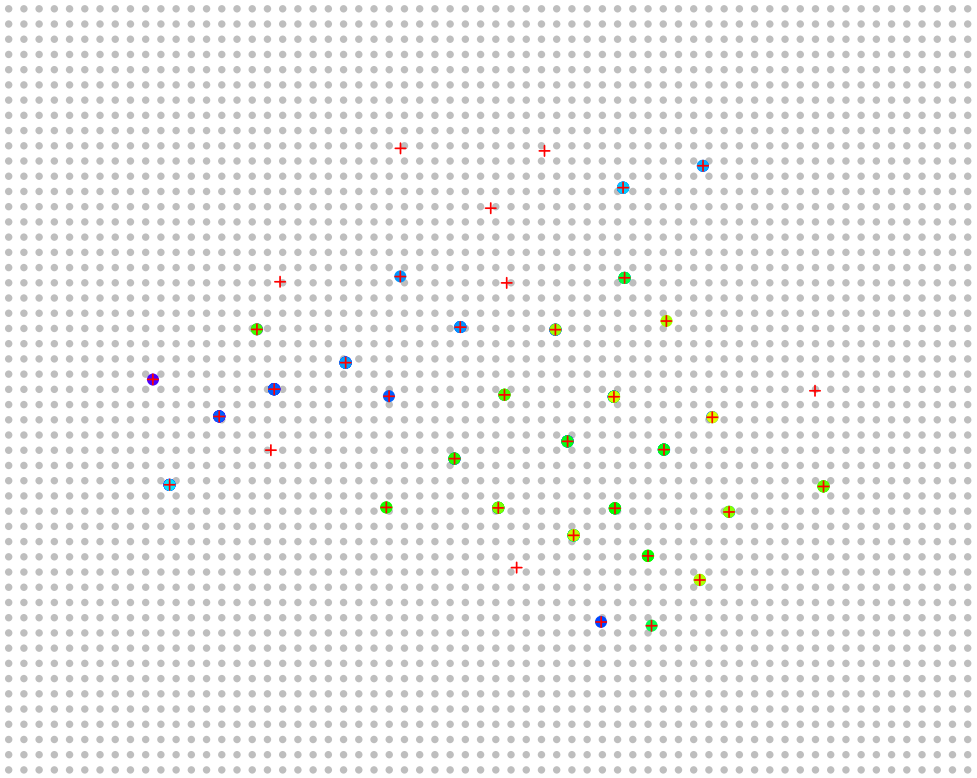
```
library(openpopscr)
```

D1: SCR with moving activity centres

As an example, I use the bear data set as described in Section 5.4. The data are stored in an `ScrData` object called `beardat`. For information on how to create this object, see the vignette on the `ScrData` class within the package.

In this survey, 47 unique individual were detected at least once. More summary information can be found by typing the name of the object into the console:

```
beardat
```



#	1	2	3	4	5	6	7	8	Total
# n	9	8	20	18	17	17	20	11	120
# u	9	5	13	6	5	2	5	2	47
# f	19	11	7	2	2	2	4	0	47
# M(t+1)	9	14	27	33	38	40	45	47	47
# losses	0	0	0	0	0	0	0	0	0
# detections	10	8	25	24	25	21	25	13	151
# detectors visited	7	6	14	15	19	14	18	10	103
# detectors used	38	38	38	38	38	38	38	38	304

I fit to this data set both a conventional SCR model and a SCR model with moving activity centres; both can be fit using the `openpopscr` package.

Models of class `ScrModel` can be used to fit SCR Models. To create a ScrModel, you need to specify a formula for each detection parameter and a starting value for all parameters, both as lists. The function `get_start_value` can be used to produce a reasonable guess for each parameter.

```
# set each parameter to be a constant
form <- list(lambda0 ~  1, sigma ~ 1)

# get start values
start <- get_start_values(beardat)

# check starting values are reasonable
start

# $sigma
# [1] 1708.49
#
# $lambda0
# [1] 0.2410844
#
# $D
# [1] 0.1551417
```

Before fitting the model, I create the model object:

```
scrmod <- ScrModel$new(form, beardat, start, num_cores = 4)
```

Once the model object has been created, you can fit the model using the fit function:

```
scrmod$fit()
```

If you have the output printed, the checking convergence message reports whether convergence has been reached and if it has not it reports the code output by `optim` that tells you why it failed to converge.

Once a model has been fit, you can type its name into the console and some summary results are printed:

```
scrmod

# PARAMETER ESTIMATES (link scale)
#                      Estimate Std. Error    LCL    UCL
# lambda0.(Intercept)   -2.222    0.13620 -2.489 -1.955
# sigma.(Intercept)      7.591    0.06573  7.462  7.720
# D                     -1.784    0.15380 -2.086 -1.483
```

The parameters are reported on the link scale. The package `openpopscr` uses the log-link function for all of these parameters. So, you would take the exponential of these reported numbers to obtain the parameter values on the response scale.

Alternatively, you can use the `get_par` function:

```
scrmod$get_par("lambda0", k = 1, j = 1)

#         [,1]
# 1 0.1083706

scrmod$get_par("sigma", k = 1, j = 1)

#       [,1]
# 1 1979.674

scrmod$get_par("D")

# [1] 0.1678821
```

I specify $k = 1, j = 1$ in the functions above for σ, λ_0 because by default `get_par` returns the value of these parameters for every occasion and detector. This is useful when they vary, but here they are constants, so I only want their value on a single occasion for a single detector.

I now fit the transient model. The transient model class is called `ScrTransientModel`. We must also now specify a start value for the movement parameter.

```
# specify formulas for parameters
form <- list(lambda0 ~ 1,
             sigma  ~ 1,
             sd ~ 1)

start <- list(lambda0 = 0.1, sigma = 2000, sd = 1000, D = 0.11)
```

```
transmod <- ScrTransientModel$new(form, beardat, start, num_cores = 4)
```

Note from the output given when creating the model object that a rectangular mesh is created based on the mesh supplied by the user. This is because the current algorithm that implements the movement model requires a rectangular study space. Note, however, that individuals may only reside and move around within the mesh supplied by the user; the extra mesh points created to produce a rectangular study space are not accessible. Also note, that the rectangular mesh will be larger and contain more mesh points than the user supplied mesh, so you may be required to reduce the resolution of the user supplied mesh in order to make computations with the rectangular mesh feasible.

I now fit the model as usual:

```
# fit model
transmod$fit()
```

```
# look at results
transmod

# PARAMETER ESTIMATES (link scale)
#                     Estimate Std. Error    LCL    UCL
# lambda0.(Intercept)   -1.916    0.18860 -2.286 -1.546
# sigma.(Intercept)      7.442    0.08668  7.272  7.612
# sd.(Intercept)         6.889    0.27280  6.354  7.423
# D                     -1.801    0.15510 -2.105 -1.497

# look at parameters on response scale
transmod$get_par("lambda0", k = 1)
```

```
#         [,1]
# 1 0.1472039

transmod$get_par("sigma", k = 1)

#       [,1]
# 1 1706.519

transmod$get_par("sd", k = 1)

#       [,1]
# 1 981.0167

transmod$get_par("D")

# [1] 0.1651229
```

D2: CJS with moving activity centres

In this section, I show how CJS SCR data can be simulated using the `openpopscr` package. I then fit a CJS model with no moving activity centres and one with moving activity centres. The models can then be compared by AIC to determine whether the movement model is necessary.

Simulating data

To simulate a survey, I require true detection and survival parameters, a detector layout, and a mesh.

```
# set true parameters
true_par <- list(lambda0 = 2, sigma = 20, phi = 0.7, sd = 20)

# make detectors array
detectors <- make.grid(nx = 7, ny = 7, spacing = 20, detector = "count")

# make mesh
```

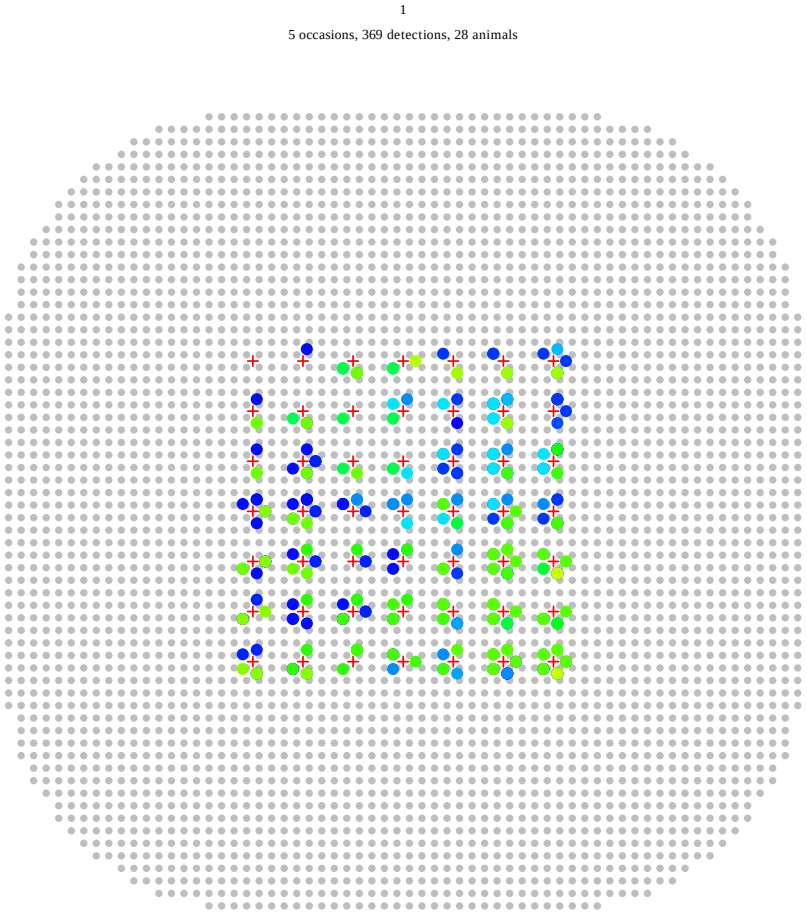
```
mesh <- make.mask(detectors,
                  buffer = 100,
                  nx = 64,
                  ny = 64,
                  type = "trapbuffer")

# set number of occasions to simulate
n_occasions <- 5

# set number of individuals
N <- 100
```

The data are simulated using the `simulate_cjs_openscr` function where whether or not individual activity centre's move can be specified.

```
# simulate ScrData
scrtransdat <- simulate_cjs_openscr(true_par,
                                    N,
                                    n_occasions,
                                    detectors,
                                    mesh,
                                    move = TRUE,
                                    seed = 95811)
```

In the simulated survey, 28 unique individual were detected at least once. More summary information can be found by typing the name of the object into the console:

```
scrtransdat
```




```
#                      1  2  3  4  5 Total
# n                   20 15  8  9  7    59
# u                   20  3  1  2  2    28
# f                   14  5  3  4  2    28
# M(t+1)              20 23 24 26 28    28
# losses               0  0  0  0  0     0
# detections         116 91 67 61 34   369
# detectors visited   36 36 30 34 20   156
# detectors used      49 49 49 49 49   245
```

I will first fit a stationary CJSModel to these data:

```
# set formulae and start values
stat_form <- list(lambda0 ~ 1, sigma ~ 1, phi ~ 1)
start <- get_start_values(scrtransdat, model = "CjsModel")
# create model object
stat <- CjsModel$new(stat_form, scrtransdat, start, num_cores = 4)
# fit model
stat$fit()
```

```
# look at results
stat

# PARAMETER ESTIMATES (link scale)
#                     Estimate Std. Error      LCL    UCL
# lambda0.(Intercept)   0.2341    0.09510  0.04766 0.4205
# sigma.(Intercept)     3.1910    0.04837  3.09600 3.2850
# phi.(Intercept)       0.7128    0.39670 -0.06463 1.4900

# look at parameters on response scale
stat$get_par("lambda0", k = 1)

#       [,1]
# 1 1.263717

stat$get_par("sigma", k = 1)

#       [,1]
# 1 24.30213

stat$get_par("phi", k = 1)

#        [,1]
# 1 0.6710296
```

I now fit the transient model. The transient model class is called `CjsTransientModel`. We must also now specify a start value for the movement parameter.

```
# specify formulas for parameters
form <- list(lambda0 ~ 1,
             sigma  ~ 1,
             phi ~ 1,
             sd ~ 1)

start <- get_start_values(scrtransdat, model = "CjsTransientModel")
```

```
trans <- CjsTransientModel$new(form, scrtransdat, start, num_cores = 4)
```

I now fit the model as usual:

```
# fit model
trans$fit()
```

```
# look at results
trans

# PARAMETER ESTIMATES (link scale)
#                     Estimate Std. Error      LCL    UCL
# lambda0.(Intercept)   0.6603    0.09784  0.46850 0.8521
# sigma.(Intercept)     3.0040    0.04522  2.91500 3.0920
# phi.(Intercept)       0.7992    0.42580 -0.03539 1.6340
# sd.(Intercept)        3.1470    0.15920  2.83500 3.4590

# look at parameters on response scale
trans$get_par("lambda0", k = 1)

#       [,1]
# 1 1.935364

trans$get_par("sigma", k = 1)

#       [,1]
# 1 20.15743
```

```
trans$get_par("sd", k = 1)

#       [,1]
# 1 23.26205

trans$get_par("phi", k = 1)

#        [,1]
# 1 0.6897964
```

The survival probability is marginally higher and the estimate of activity range reduced by accounting for transience.

The two models can be compared by AIC. The transient model is preferred.

```
AIC(stat, trans)

#       df       AIC
# stat   3 1098.1665
# trans  4  995.6119
```