NEW ANALYTICAL METHODS FOR CAMERA TRAP DATA

A THESIS SUBMITTED TO

THE UNIVERSITY OF KENT

IN THE SUBJECT OF STATISTICS

FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY BY RESEARCH

BY

NATOYA O. A. S. JOURDAIN

May 2017

Abstract

Density estimation of terrestrial mammals has become increasingly important in ecology, and robust analytical tools are required to provide results that will guide wildlife management. This thesis concerns modelling encounters between unmarked animals and camera traps for density estimation. We explore Rowcliffe et al. (2008) Random Encounter Model (REM) developed for estimating density of species that cannot be identified to the individual level from camera trap data. We demonstrate how REM can be used within a maximum likelihood framework to estimate density of unmarked animals, motivated by the analysis of a data set from Whipsnade Wild Animal Park (WWAP), Bedfordshire, south England. The remainder of the thesis focuses on developing and evaluating extended Random Encounter Models, which describe the data in an integrated population modelling framework. We present a variety of approaches for modelling population abundance in an integrated Random Encounter Model (iREM), where complicating features are the variation in the encounters and animal species. An iREM is a more flexible and robust parametric model compared with a nonparametric REM, which produces novel and meaningful parameters relating to density, accounting for the sampling variability in the parameters required for density estimation. The iREM model we propose can describe how abundance changes with diverse factors such as habitat type and climatic conditions. We develop models to account for induced-bias in the density from faster moving animals, which are more likely to encounter camera traps, and address the independence assumption in integrated population models. The models we propose consider a functional relationship between a camera index and animal density and represent a step forward with respect to the current simplistic modelling approaches for abundance estimation of unmarked animals from camera trap data. We illustrate the application of the models proposed to a community of terrestrial mammals from a tropical moist forest at Barro Colorado Island (BCI), Panama.

Acknowledgements

First, I would like to express my gratitude to my academic supervisors, Dr Diana J. Cole and Professor Martin S. Ridout for their continuous support, as well as Dr Marcus J. Rowcliffe for his advice and expertise and for providing the Whipsnade Wild Animal Park (WWAP) data set and the Barro Colorado Island (BCI), Panama data set. They have encouraged me to think and work independently while always being there to guide me with useful advice. It has been a pleasure working with you all, thank you! My particular thanks to Diana for your support and patience, and for being an inspiring mentor, I am especially grateful.

I am grateful to the Engineering and Physical Sciences Research Council (ESPRC), the School of Mathematics, Statistics & Actuarial Science, University of Kent, and the National Centre for Statistical Ecology (NCSE) for funding my PhD research. Thanks as well to Dr. David Borchers and Dr. Eleni Matechou for examining my thesis.

Thank you to all my family and friends, with particular thanks to my mom for your unwavering love and support, my longest running best friend Bela Dama (Savita) for your continuous support and encouragement to stay on point, and my brother Bryan for his role as big brother. And finally, to Ryan, for playing such a pivotal role during these four important years of my life, thank you for sharing it with me!!

Contents

List of Tables

List of Figures

Chapter 1

Introduction

There is considerable threat to the survival of native animal species and it is necessary to determine why these threats are occurring, how these threats are affecting population sizes and what can be done to prevent the loss of species forever. Camera trapping can be used to monitor wildlife and obtain information on behaviour, activity patterns and characteristics (Bridges et al., 2004; Ridout and Linkie, 2009), their interactions, and to determine population abundance. Camera trapping is the use of fixed cameras, triggered by infra-red sensors (Rowcliffe et al., 2011) to take detailed snapshots and/or videos of passing animals or other objects in front of them. For example, Figure 1.0.1 shows a group of diverse wildlife species surveyed with camera traps. Camera trapping is a low-cost survey tool which is widely used across the globe to study elusive medium-to-large terrestrial mammals and birds (Rowcliffe et al., 2008, 2011; Samejima et al., 2012) and small arboreal mammals (Oliveira-Santos et al., 2008). Camera trapping has been shown to be more effective and efficient than traditional distance sampling, radio telemetry and capture-mark-recapture methods as it minimizes the chance of environmental disturbances (Roberts, 2011), or wounding or stressing target animals (Cutler and Swann, 1999). These cameras can be left for up to six months (McCallum, 2013), recording the presence and behaviour of the animals in their natural environment (Swinnen et al., 2014). Leaving camera traps untouched for such long periods minimizes the risk of human disturbances, which might influence the behaviour of targeted animals (McCallum, 2013). Camera traps can operate day and night, in most weather conditions, and in difficult terrain where other methods are likely to fail (Rowcliffe et al., 2011). Camera traps can provide information on species distribution, habitat

use, and in particular, behaviour patterns of rare and highly cryptic species. Camera trapping research applications span over a wide range of areas, for example faunal checklists and detection of endangered or elusive species (Sanderson and Trolle, 2005; Rovero and De Luca, 2007); a range of species-specific or focal purpose studies on activity patterns (Tobler et al., 2009); occupancy estimation and modelling (Linkie et al., 2007); abundance and density estimation of individually recognizable species through capture-mark-recapture analysis (Karanth and Nichols, 1998), abundance and density estimation of non-recognizable species (Rowcliffe et al., 2008; Lucas et al., 2015). In camera trap studies, populations are sampled (rather than enumerated) not because they are too large to count every individual, but because the target species are elusive (Foster and Harmsen, 2012).

The focus of this thesis is on abundance and density estimation of animals from camera trap data. Up until recently, methods that employed camera traps to estimate animal abundance have been restricted to capture-recapture analysis of species with unique identifiable markings; examples include relative abundance estimation of tigers using camera trapping photographic rate (Carbone et al., 2001) and relative abundance of tigers and their prey as measured by camera traps (O'Brien et al., 2003). Therefore, the development of sophisticated statistical models for estimating population densities of unmarked or non-recognizable species is severely lacking given the wealth of individual and group level data being collected on a huge range of animal populations. This thesis aims to address a fundamental question in ecology — what is the estimated density of species from camera trap data that cannot be identified to the level of individuals? We develop an integrated model, which estimates population densities of unmarked animals and which accounts for fundamental correlation between data sets. This method extends on the Random Encounter Model (REM) developed by Rowcliffe et al. (2008) for non-recognizable animals from camera trap data. We start with a review of camera trapping with marked animals in Section 1.1. In Section 1.2 we discuss camera trapping with unmarked animals. The Chapter continues with the thesis motivation and aims in Section 1.3, and general thesis methodology in Section 1.4. Finally, the Chapter concludes with the thesis structure in Section 1.5.

(a) agouti

(b) roe deer

(c) mara

(d) collared Peccary

Figure 1.0.1: Examples of the diverse wildlife species surveyed with camera traps at Barro Colorado Island, Panama. Methods for density estimation of unmarked species such as a) agouti *Dasyprocta punctata*, b) roe deer *Capreolus capreolus*, c) mara *Dolichotis patagonum* and d) collared peccary *Tayassu tajacu* are less well established than methods for species with identification marks.

1.1 Camera trapping with marked animals.

For the past two decades camera trapping has been used as a noninvasive technique for the sampling of animal populations, particularly, those that are elusive often occur in low densities (O'Connell et al., 2010). As the estimation of density and abundance became increasingly important to ecologists, methods that use camera trap data were developed to address this issue. Early work by Karanth (1995) and Karanth and Nichols (1998) was the landmark for using camera traps to estimate population abundance of recognizable species, particularly tigers *Panthera tigris*. To estimate abundance, Karanth (1995) and Karanth and Nichols (1998) adopted a capture-recapture (C-R) approach. In capture-recapture analysis, the main objective is to estimate the (possibly changing) size of a population of individuals (e.g. animals, birds) in their natural state. To this end, the observer captures individuals from a population on a number of successive occasions, and each time an individual is caught, a record is marked for it or on

it to show the occasion of that capture; the individual is then returned to the population. In Karanth (1995) and Karanth and Nichols (1998) the tigers could be identified unambiguously from the photographs, and estimates of capture probabilities and population size were determined. Population abundance was estimated on the assumption of geographic closure around the sample site, and demographic closure during the study period. Karanth (1995) explored the potential use of camera traps to answer questions such as can tiger populations be adequately sampled from camera traps, can capture-recapture analysis derive reasonable estimates of population densities from camera trap data, and under what conditions is this approach practical. Building on this work, Karanth et al. (2004) explored camera trapping as a scientific tool to estimate population abundance of tigers using capture-recapture modelling. Rigorous sample surveys were designed to answer questions such as why sample, what to sample and how to sample rare or elusive animals.

Capture-recapture analysis of camera trap data has since become a common method of estimating population density, particularly for terrestrial mammals in the tropics. Trolle and Kéry (2003), for example, adopted the method of Karanth and Nichols (1998) which combined camera trapping with mark-recapture models to estimate density of ocelots in the Brazilian Pantanal — the first published attempt to estimate population size of felids in South America. The study adopted natural variation technique in body markings of ocelots in closed population to estimate density. This method was later used in a larger-scale investigation by Trolle and Kéry (2005) on ocelot density in the northern Pantanal. The focus was on further development and evaluation of the camera trap methodology, and the effect of placement of traps on trails and roads on trapping rates. Silver et al. (2004) also adopted methods of camera trapping and density estimation by Karanth and Nichols (2002) to estimate ocelot density in eastern Bolivia. The initial focus of the study was jaguars (Maffei et al., 2004) but valuable data on ocelots were collected simultaneously during the pilot studies conducted. Silver et al. (2004) used systematic camera trap surveys, and capture-recapture sampling methods to estimate total abundance across five Bolivian dry-forest sites with different habitat types and annual rainfall regimes. Further studies by Trolle et al. (2007) used capture-recapture analysis of camera trap data to estimate maned wolf density in two Brazilian study

areas: Cerrado and the Pantanal. Such capture-recapture applications on individually identifiable species — typically large, patterned carnivores — continued to develop rapidly, particularly with the advent of mark-resight methods (Fuller et al., 2001; Watts et al., 2008), and more recently spatially explicit capture-recapture (SERC) models (e.g., Royle et al., 2009; Royle, 2011; Royle et al., 2013; Borchers et al., 2014).

While these studies have been shown to provide robust, unbiased density estimates, the methods adopted are restricted to animals being individually recognizable, based on the natural variation in patterns (rosette, stripes, etc.) and/or tags. However, the majority of wildlife species are not recognizable to individual level from photographs, rendering capture-recapture approaches for abundance estimation difficult. Also the methods of capture-recapture that use camera traps are sensitive to the spacing of cameras relative to the size of animal home ranges (Rowcliffe et al., 2008). Therefore, new methods that adopt camera trapping techniques are required to estimate population densities of these unmarked species. The next section (1.2) gives detailed information on methodologies developed to monitor wildlife and estimate densities of unmarked individuals from camera trap data.

1.2 Camera trapping with unmarked animals.

There have been several analytical approaches proposed for population surveys of unmarked species using camera traps. For example, Carbone et al. (2001) and O'Brien et al. (2003) used relative abundance indices, such as detection rates, as a measure of estimated densities. However, these methods have been criticized for their simplistic approaches, and their implicit assumption that detectability is constant across areas, time and species (Burton et al., 2015). Further, these methods have not considered a functional relationship between a camera index and animal density. Jennelle et al. (2002) agreed that a properly calibrated index may be useful in rapid conservation assessments but to use such an index as a substitute for direct estimation of density, one must: 1) demonstrate a functional relationship between the index and the density that is invariant over the desired scope of inference; 2) calibrate the functional relationship by obtaining independent measures of the index and animal density; and 3) evaluate

the precision of the estimates. A properly calibrated index must model the underlying process between animals and camera traps. In the absence of such an index and individual identification of species, detection rates confound abundance and detectability — they may reflect both the number and behaviour of animals, as well as nuisance factors related to sampling errors (Burton et al., 2015). Rowcliffe et al. (2008) developed a Random Encounter Model (REM) to eliminate the need for individual identification of animals by modelling the underlying process of contact between animals and cameras. REM considers a functional relationship between the camera index and animal density, which is derived on theoretical grounds of the "ideal gas" model. REM accounts for both animal movement and detectability while considering the sampling process underlying the collection of animal photographs. When tested on an enclosed animal park with known abundances REM has been found to provide reliable density estimates. In general, subject to unbiased camera placement, REM opens the possibility of estimating animal density where it has not been previously possible (Rowcliffe et al., 2008). However, REM remains to be thoroughly tested in broader camera trapping surveys. Another advancement in methods that use camera trap data for density estimation is the spatial capture-recapture (SCR) model (Chandler et al., 2013) for inference about density when individuals cannot be uniquely identified nor detected with certainty. The SCR has the advantage over capture-recapture methods in yielding explicit estimates of animal density without the need for additional data. The SCR models of Chandler et al. (2013) SCR require spatially correlated count data, which prove to be informative about encounter rate parameters and density. However, like REM the broader reliability of SCR models remain to be more thoroughly tested.

Other studies (Mazzolli, 2010; Trolle et al., 2008; Kelly et al., 2008; Noss et al., 2003) have used camera traps to estimate abundance of species that lack individually identifiable natural markings. Phenotypic and environment-induced characteristics were used as identifiers, for example, pumas ($Puma/concolor$) and tapirs ($Tapiris/terrestris$) have been identified by scars, parasites, torn ears, toenail markings or color, tail length and kinks, and dark or light body markings. However, these are not necessarily found on all individuals within the population, and when samples are sufficiently large identification becomes increasingly difficult (Harmsen, 2006) and abundance estimation via

C-R methods is unlikely to be reliable (Foster and Harmsen, 2012). Noss et al. (2003), for example, assume that an unidentified individual is the same individual previously photographed at that location but this assumption is only valid if the target species defends exclusive territories (Foster and Harmsen, 2012). On the other hand Mazzolli (2010) excluded these photos completely, while Trolle et al. (2008) failed to clearly report the proportion of photographs that were unidentifiable.

As an alternative approach, camera trap studies used mark-resight estimators (for example Bartmann et al., 1987; Bowden and Kufeld, 1995; McClintock et al., 2009) to deal with unmarked species. Abundance estimation of unmarked species is obtained by using the frequency of marked and unmarked individuals. The Mark-resight estimator is a more robust approach to C-R methods, which does not require all animals in the sample to be marked, therefore, estimation is possible when only some of the study animals have unique identifiers (see Foster and Harmsen, 2012). However, a limitation of this method is that it requires that the number of marked animals be known, and so a sample of the population must be captured and marked prior to camera trapping (Fuller et al., 2001; Matthews et al., 2008), which may be nearly impossible to do, particularly for elusive species in the tropics.

1.3 Thesis motivation and aims

This thesis primarily aims to develop new analytical methods for abundance estimation of unmarked animals from camera trap data. Biodiversity is constantly threatened by factors such as climate change, habitat loss and overharvesting, and new statistical methods are needed to model and estimate population abundance. The majority of studies are based on species that can be identified to individual level either by unique markings or tags, and the REM method for unmarked species is yet to be applied to a more extensive data set. We focus on extending the REM of Rowcliffe et al. (2008) and developing methods that will be broadly applicable to many species, particularly unmarked species for abundance estimation from camera trap data.

REM was tested on a small data set with known census at Whipsnade Wild Animal

Park (WWAP), Bedfordshire, south England during a June-July, 2005 survey. For the purpose of analyses Rowcliffe et al. (2008) focused on four habituated mammals at WWAP: red necked wallaby (*Macropus rufogriseus*), Chinese water deer (*Hydropotes inermis*), Revee's muntjac (*Muntiacus reevesi*), and mara (*Dolichotis patagonum*). But given the efforts of Rowcliffe et al. (2011), during a 2008-2010 survey, huge amounts of data are available for a community of terrestrial mammals from a tropical moist forest at Barro Colorado Island (BCI), Panama, hence statistical techniques are required to exploit the information fully. Using REM for estimating density of unmarked animals currently requires five steps to derive all the necessary components of the analysis. Consequently, this thesis develops a single analytical framework of the REM formula, allowing for more robust model-based inference. This has particular relevance for incorporating covariates such as habitat/land cover, and dealing with spatial autocorrelation.

In addition to obtaining a unified approach for density estimation in REM, we aim to develop models that incorporate multiple sources of variance. By considering all the information required in the analysis in REM we aim to describe the implications of these variances for unbiased estimation of the density, and model how sample sizes in the different sources affect overall precision. We aim to develop parametric approaches for abundance estimation in REM with the aim of producing estimates of meaningful and relevant parameters.

In this thesis we develop robust and flexible frameworks for modelling encounters between animals and camera traps for density estimation in REM, which can be modified according to the purpose of the particular study or application. In doing so, further application of the models may provide new insights relevant to density estimation of unmarked species from camera trap data.

The development of new models that are suitable for describing data between animals and camera traps requires knowledge about common models for count data and continuous data, and their potential relevance for land mammals. Hence, we additionally consider the performance of new models developed for dependent and independent data sets. Combining multiple sources of data into a single framework utilizes an integrated

population modelling approach for parameter estimation, which is a popular tool for estimating parameters with overlapping information, but which rely on the assumption of independence between data sets (see Newman et al., 2014, Ch. 9). We develop a variety of modelling approaches with the aim of introducing new models that are more efficient, comprehensive and more informative, and applicable to all species, marked or unmarked, with possible adaptation where required.

1.4 General thesis methodology

The work in this thesis is carried out within the maximum-likelihood inference framework. Models are implemented in R Core Team (2016). Maximum-likelihood estimates are obtained by numerical minimization of the negative log-likelihood function using the optimization routine optim, which implements the Nelder-Mead simplex search algorithm (Nelder & Mead, 1965) as the default method. The optimization routine optim is a general-purpose optimization routine, which bundles six different optimization methods but we use the default method Nelder-Mead, and Brent algorithms for minimization without derivatives (Brent, 1973) for one-dimensional unconstrained functions. The Brent method is implemented in optimize, which searches an interval for a minimum or maximum. As optim is designed primarily for unconstrained optimization (although it includes an option for box-constrained optimization), parameters are transformed to the log scale as appropriate prior to optimization. The log scale is used for parameters that are constrained to $[0, \infty)$, such as density, mean speed, and shape parameters. Standard errors are derived on the transformed log scale with the option *hessian*, which returns an approximation to the an observed information matrix. This function returns a symmetric matrix giving an estimate of the Hessian at the solution found. The asymptotic variance-covariance matrix of the maximum-likelihood estimators is obtained by taking the inverse of the Hessian. Standard errors on the original parameter scale are calculated from those obtained on the transformed log scale using the delta method approximation.

We study properties of estimators using simulations. We compute the Root Mean Square Error (RMSE) and Standard deviation (Sd) using the formulas outlined below. The sample standard deviation (Sd), which measures the spread of the of the parameter

estimates is computed by using the sd function in R, which has formula

$$\text{sd} = \sqrt{\sum_{i=1}^{N} (\hat{\theta}_i - \bar{\hat{\theta}})^2 \Big/ (N-1)},$$

where N represents the number of simulations, $\hat{\theta}_i$ is the ith $(i = 1, 2, ..., N)$ parameter estimate, and $\bar{\hat{\theta}}$ is the mean of the parameter estimates. The Root Mean Square Error (RMSE) is computed from the Mean Square Error (MSE), which is defined as

$$\text{MSE} = \text{Var} + (\bar{\hat{\theta}} - \theta)^2,$$

and

$$\text{RMSE} = \sqrt{\text{MSE}}.$$

In analysing real data we use the sample mean, in some cases, as an estimate of a parameter and an approximate standard error is computed. An approximate standard error of the sample mean is computed as $\sigma_{\text{mean}} = \sigma/\sqrt{\text{n}}$, where σ is the standard deviation of the data, and n is the number of observations.

We also carry model selection using information criteria of the form

$$I = -2\text{log}L + q,$$

where L is the likelihood evaluated at the maximum likelihood estimates of the parameters and q is the penalty function that involves the number of parameters (Buckland et al., 1997). One of the most commonly used information criteria is the Akaike information criterion (AIC) (Akaike, 1992) with

$$\text{AIC} = -2\text{log}L + 2p,$$

where p is the number of parameters. To allow for the non-optimal AIC values to be more easily interpreted ΔAIC values are calculated, where

$$\Delta\text{AIC} = \text{AIC} - \min(\text{AIC})$$

(Burnham and Anderson, 2004). The model with the smallest AIC is the best model (and of course has ΔAIC equal to zero).

1.5 Thesis structure

This thesis consists of seven core Chapters.

In Chapter 2 we introduce the main model - the Random Encounter Model (REM) on which this thesis is built, demonstrating how the model can be used to estimate density of unmarked species from camera trap data. We start the Chapter by presenting the theoretical "ideal gas" model, on which REM is based, exploring its properties and application in biological and ecological studies. We then give a derivation of the REM formula proving that the REM method could be derived by assuming a Poisson distribution for the encounters between animals and camera traps and obtain the maximum likelihood estimate of the encounter rate. We also demonstrate the flexibility of the maximum likelihood framework by including habitat-specific covariates, accounting for the additional variation in population abundance. We then conclude the Chapter by applying the maximum likelihood framework of REM with a Poisson model to the analysis of the WWAP data set.

In Chapter 3 we consider an integrated population modelling approach for the maximum likelihood framework of the REM formula, which is widely used to account for overlapping information in multiple data sets (Newman et al., 2014, Chpt. 9). We develop an integrated REM (iREM), which combines the encounter data and animal speed data in a unified framework for abundance estimation. iREM provides a new approach for abundance estimation of unmarked animals accounting for the sampling variability in the data sets and allowing for the accurate treatment of precision and correlation in the estimators. We also show that more variable models such as a negative binomial, and zero-inflated Poisson, or a zero-inflated negative binomial can be used in REM to model the encounter data for density estimation. We also consider the case in which animals move in family groups by integrating this information in iREM to estimate density of the individual. The methods in Chapter 3 are illustrated by a simulation study and

analysis of data for four animal species at WWAP. The integrated REM (iREM) fitted in Chapter 3 has links with the models in the later Chapters of the thesis.

In Chapter 4 we extend the iREM model proposed in Chapter 3 to the case in which there are spatial covariates such as habitat/land-cover variables and random effects covariates such as camera random effect which may describe additional variation in the density. We detail two iREM model approaches that use a log link function to incorporate habitat-specific covariates and camera random effects. In the latter part of Chapter 4 we describe an iREM model approach that simultaneously incorporates habitat specific covariates and camera random effect covariates which may explain some of the variation in the model parameters. We illustrate the utility of the models proposed by fitting them to the data of four species at WWAP.

Chapter 5 builds on the iREM model in Chapter 3 to provide a comprehensive model, accounting for the sampling variability of the animal speed data, and the distance and angle at which the camera sensor detects the animal data. The models developed in Chapter 5 use distance sampling methodology (Buckland et al., 2001) to estimate detection distance and angle of the camera trap, which are required in REM for density estimation. Assuming independence between the data sets, we show how accounting for the sampling variability in model parameters can lead to more precise estimation of the density. The models are illustrated by simulation studies and an analysis of the BCI, Panama data set.

Chapter 6 considers animal speed data collected from camera traps. Faster moving animals are more likely to encounter camera traps than slower moving animals (Hutchinson and Waser, 2007), which could introduce bias in the estimate of the speed required in REM. Size biased sampling (Patil and Rao, 1978) deals with this bias. In size biased sampling the probability of encounter between an animal with a given speed and camera traps is proportional to a weight function, which is equal to the animal's speed. We show that the probability of encountering an animal with a given speed from a Poisson model, a negative binomial model (NB), a zero-inflated Poisson (ZIP) or a zero-inflated

negative binomial model (ZINB) is approximately proportional to the speed of the animal. We test this approximation in a simulation study and illustrate its use by an analysis of the BCI data set. We also consider the true probability of encountering an animal from a Poisson model in a size biased sampling method, which we illustrate in a simulation study. The size biased models fitted in Chapter 6 have links with the model in Chapter 7 where the size biased models are used for the speed data model component in iREM.

In Chapter 7 we propose and evaluate an iREM model with size biased sampling for abundance estimation. This model is an adjustment of the model in Chapter 3 correcting for the bias in animal speed. In integrated population modelling it is required that the data sets be independent. In most camera trapping studies, however, the data sets are collected from the same camera trap source, which makes the data sets dependent. In this Chapter we start by investigating the independence assumption of the encounter data and the speed data in a simulation study. We show that the effect on estimated density is inconsequential if the assumption of independence is disregarded. We show the relevance of accounting for the bias in speed of faster moving animals in density estimation. We illustrate the comprehensive iREM model framework developed in Chapter 5 with size biased sampling with an analysis of the data set at BCI, Panama.

Chapter 2

Random Encounter Model (REM)

Prior to Rowcliffe et al. (2008) Random Encounter Model (REM), estimation of animal abundance using camera trapping analysis was usually accomplished by capture-recapture methods. These methods are generally hard to implement at large spatial scales, and they under-represent populations of animals that cannot be identified to the individual level. REM estimates animal abundance from camera trap data without the need to identify animals (either by unique natural markings or tags).

This Chapter introduces REM and demonstrates how REM can be used to estimate density using a data set from Whipsnade Wild Animal Park (WWAP), Bedfordshire, south England. REM is based on the "ideal gas model", which is a theoretical gas composed of many randomly moving point particles that do not interact, except when they collide elastically (Hutchinson and Waser, 2007). We discuss the ideal gas model and its application to modelling encounter rates with animals in Section 2.1. The Chapter then continues with Section 2.2, which provides an explanation of how the ideal gas model can be adapted to modelling camera trap encounters to give REM. It then goes on, in Section 2.3, to give the methods used to estimate density in REM, while Section 2.4 gives the methods of estimating the variance of density in REM. In Section 2.5 we provide methods for constructing appropriate confidence intervals of the density. Several key assumptions are required in REM to make valid inference concerning the density of the population sampled. These assumptions are discussed in detail in Section 2.6. The data set from WWAP used by Rowcliffe et al. (2008) to test REM is described in Section 2.7. To conclude the Chapter, Section 2.8 provides an illustration of the

application of methods used in REM to estimate density.

2.1 Ideal gas model

An ideal gas model is a model for collision rates among molecules. Maxwell (1860) worked out expected rates of collision among molecules in three dimensions given the concentration of molecules in the gas, their mean speed, and their size. The formula in Maxwell (1860) assumes molecules move with uniform velocity (speed) in straight lines striking against the vessel in which they are contained, and thus producing pressure. The movement of molecules are also assumed to be independent, and equally likely in all directions. According to Maxwell (1860) it is not necessary to assume each molecule to travel any great distance in the same straight line; for the effect in producing pressure will be the same if the molecules strike against each other, so that the straight line described may be very short. The assumptions of the ideal gas model have proven attractive to biologists, who have applied the ideal gas model more generally to predict collision rates. For example, biologists have used these models, and their two-dimensional analogues, to predict "encounter" rates and duration of encounters among animals and social groups that move randomly and independently given population density, velocity, and distance at which encounters occurs. In this thesis we use the two-dimensional ideal gas model to predict encounter rates. The next section describes the encounter rate formula in a two-dimensional ideal gas model.

2.1.1 Encounter Rate Formula in a Two-dimensional Ideal Gas Model

Classic two-dimensional ideal gas models assume a circular detection zone around molecules within which contact occurs. Figure 2.1.1 shows a circular detection zone, with radius r, and the directions in which molecules are moving. In Figure 2.1.1 molecules are assumed to move randomly and independently of each other, and movement is also assumed to be equally likely in all directions.

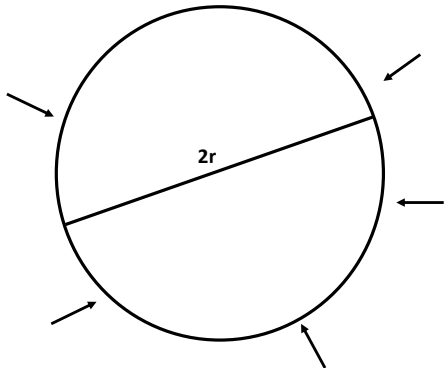


Figure 2.1.1: Circular detection zone where the arrows ($\longrightarrow$) denote the approach directions of molecules. The diameter of the circle is $2r$.

In a two-dimensional ideal gas model, the encounter rate formula links the number of times individuals come within a specified distance of one another to their speed and density. There are different versions of the formula depending on whether the speed of movement is constant or follows a Maxwell-Boltzmann distribution. A Maxwell-Boltzmann distribution is a probability distribution for describing the speed of individuals in idealized gas models. To use a Maxwell-Boltzmann distribution to describe the speed of movement, it is required that the temperature of the system in which individuals occupy and the mass of the individuals to be known.

For a stationary object with a circular detection zone within which contact occur (Figure 2.1.1), Hutchinson and Waser (2007) showed that the area covered by a moving individual can be easily defined as the product of the individuals' speed (v), the time period (t), and twice the radius of the detection zone (r). That is, the area covered by each individual is $2rtv$. If the total area within which the individuals are contained is defined by A, and the density of individuals is D, then the number of individuals present is given as DA and the area covered by moving individuals is therefore defined as $2rtvDA$. If the number of individuals in the total area, A, is assumed to follow a Poisson distribution, then the expected number of contacts is obtained by dividing the covered area, $2rtvDA$ by A. That is, the expected number of contacts

$$\lambda = \frac{2rvtDA}{A} = 2rvtD. \tag{2.1.1}$$

Here the speed of movement, v is assumed to be constant. To show this, let X_i (for $i =$

$1, 2, ..., N$) be N independent and identically distributed Bernoulli random variables defined as

$$X_i = \begin{cases} 1 & \text{if individual } i \text{ encounters the camera,} \\ 0 & \text{if individual } i \text{ does not encounter the cmaera.} \end{cases}$$

The probability of encounter, p is therefore the ratio of the area covered by an individual, $2rtv$, and the total area, A, that is $p = (2rtv/A)$. Hence,

$$X_i \sim \text{Bernoulli}\left(\frac{2rtv}{A}\right).$$

Let the number of individuals recorded by the camera be defined as $X = X_1 + X_2 + ... + X_N$. Then conditional on the number of individuals, N, the number of encounters has a binomial distribution, that is

$$X|(N = n) \sim \text{Binomial}\left(n,\ \frac{2rtv}{A}\right)$$

with expected value $np = n \times 2rtv/A$. If N has a Poisson distribution with mean DA, then the marginal distribution of X can be written as

$$\begin{aligned} P(X = x) &= \sum_{n=0}^{\infty} P(X = x \mid n) P(N = n) \\ &= \sum_{n=x}^{\infty} \left\{ \binom{n}{x} p^x (1-p)^{n-x} \right\} \left\{ \frac{e^{(-DA)}(DA)^n}{n!} \right\}, \end{aligned}$$

since the number of encounters conditional on N is Binomial, N is Poisson and the conditional probability is zero if $N = n < x$ (where, $p = 2rtv/A$). Simplifying this expression gives,

$$\begin{aligned} P(X = x) &= \frac{e^{(-DA)}(DAp)^x}{x!} \sum_{n=x}^{\infty} \frac{\{(1-p)DA\}^{n-x}}{(n-x)!} \\ &= \frac{e^{(-DA)}(DAp)^x}{x!} e^{(1-p)DA} \\ &= \frac{e^{(-DAp)}(DAp)^x}{x!}. \end{aligned}$$

Therefore, the number of encounters has the Poisson distribution with mean $DAp =$

$DA \times 2rtv/A$, that is, $X \sim \text{Pois}(2rtvD)$. Then the expected number of encounters is $\lambda = 2rtvD$.

If individuals vary their speed or if speed varies among individuals, Hutchinson and Waser (2007) state that the ideal gas formula must be amended. The term v in equation (2.1.1) must be replaced by a revised mean relative speed between the focal individual and other individuals in the population. To calculate this it is necessary to know the distribution of the absolute speeds, not just their means. But if the speed of the focal individual is known the revised encounter rate formula in two-dimension is given as $2rt\mu_v D$, where μ_v is the true population mean speed. To show this, let us suppose that individuals are moving at speeds v_i for $i = 1, 2, ..., n$, and let X_i (for $i = 1, 2, ..., n$) be n independently distributed Bernoulli random variables

$$X_i = \begin{cases} 1 & \text{if individual } i \text{ encounters the camera,} \\ 0 & \text{if individual } i \text{ does not encounter the camera,} \end{cases}$$

each with a probability of success $p_i = 2rtv_i/A$, that is

$$X_i \sim \text{Bernoulli}\left(\frac{2rtv_i}{A}\right).$$

If we let $\mathbf{P} = (p_1, p_2, ..., p_n)$, then conditional on $N = n$, and v_i the sum of these independent non-identical Bernoulli random variables, $X = \sum_{i=1}^{n} X_i$ is distributed as a Poisson-Binomial random variable with parameter $\mathbf{P}$, that is $X|(N = n,\ v_i) \sim \text{PBD}(\mathbf{P})$. We can obtain the encounter rate using probability generating functions. The random variables X_i, conditional on v_i, are independent so therefore the probability generating function is defined to be

$$G_{X|N=n,v_i}(s) = G_{X_1|N=n,v_1}(s)G_{X_2|N=n,v_2}(s)...G_{X_n|N=n,v_n}(s), \tag{2.1.2}$$

for all $s \in \mathbb{R}$ for which the expected value exists (note that the probability generating function G of a random variable X is defined as follows, for all $s \in \mathbb{R}$ for which the expected value exists: $G_X(s) = \mathbb{E}(s^X)$). As

$$G_{X_i|N=n,v_i}(s) = 1 - p_i + p_i s \quad \text{for} \quad i = 1, 2, ..., n. \tag{2.1.3}$$

and $p_i = av_i$, where $a = 2rt/A$, then equation (2.1.2) becomes

$$G_{X|N=n,v_i}(s) = \prod_{i=1}^{n}(1 - av_i + av_i s).$$

To obtain the marginal distribution of the random variable X, the random variables v_i need to be integrated out. So we obtain,

$$G_{X|N=n}(s) = \int_{v_n}\int_{v_{n-1}} ... \int_{v_1} \prod_{i=1}^{n}(1 - av_i + av_i s)f(v_1, ..., v_n)dv_1...dv_n,$$

which is

$$\begin{aligned} G_{X|N=n}(s) &= \prod_{i=1}^{n}\int(1 - av_i + av_i s)f(v_i)dv_i. \\ &= \prod_{i=1}^{n}\left[\int f(v_i)dv_i - \int av_i f(v_i)dv_i + \int asv_i f(v_i)dv_i\right]. \\ &= \prod_{i=1}^{n}(1 - a\mu_v + a\mu_v s). \end{aligned}$$

This is the probability generating function of a binomial distribution and from previous $X \sim$ Poisson distribution. Thus,

$$P(X = x) = \binom{n}{x}(p^*)^x q^{*n-x}.$$

where $p^* = a\mu_v = 2rt\mu_v/A$. For a binomial distribution, the expected value of X is defined as $\mathbb{E}(X) = n \times p^* = n \times 2rt\mu_v/A$ where n, in this case, is the number of individuals and is defined as the product of the density (D) and the area (A), that is $n = D \times A$. Hence, the expected encounter rate in an ideal gas for individuals with varying speed is defined to be

$$\lambda = \frac{2rt\mu_v \times D \times A}{A} = 2rt\mu_v D, \tag{2.1.4}$$

where μ_v is the expected speed of movement of the population. This formula (2.1.4) still holds for the mean number of encounters if movement is not in straight lines. However, the number of contacts will no longer follow a Poisson distribution since encounters

will now be re-encounters with the same individuals, which leads to a more variable distribution of contacts (see Hutchinson and Waser, 2007). For example, let us assume that the number of encounters follows a negative binomial distribution with expected encounter rate λ, defined in equation (2.1.4). Then, if $x_i \mid \lambda \sim \text{Pois}(\lambda)$ with probability mass function

$$h_{pois}(x_i \mid \lambda) = \frac{e^{-\lambda}}{x_i!}(\lambda)^{x_i}, \tag{2.1.5}$$

where $\lambda \sim \text{Ga}(\kappa, \beta)$ with probability density function

$$g(\lambda) = \frac{\beta^{\kappa}}{\Gamma(\kappa)}\lambda^{\alpha-1}\text{exp}(-\lambda\beta), \tag{2.1.6}$$

we can obtain the density function of the encounters x_i by integrating out λ.

$$h(x_i) = \int_{\lambda=0}^{\infty} h(x_i \mid \lambda)g(\lambda)d\lambda = \int_0^{\infty} \frac{e^{-\lambda}}{x_i!}(\lambda)^{x_i}\frac{\beta^{\kappa}}{\Gamma(\kappa)}\lambda^{\kappa-1}\text{exp}(-\lambda\beta)d\lambda$$

$$= \frac{\beta^{\kappa}\Gamma(\kappa+x_i)}{x_i!\Gamma(\kappa)\beta)^{\kappa+x_i}}\int_0^{\infty}\frac{\beta^{\kappa+x_i}\lambda^{x_i+\kappa-1}\text{exp}(-\lambda\beta)}{\Gamma(\kappa+x_i)}d\lambda$$

$$= \frac{\beta^{\kappa}\Gamma(\kappa+x_i)}{x_i!\Gamma(\kappa)\beta^{\kappa+x_i}} = \frac{\Gamma(x_i+\kappa)}{\Gamma(x_i+1)\Gamma(\kappa)}\left(\frac{1}{1+\beta}\right)^{x_i}\left(\frac{\beta}{1+\beta}\right)^{\kappa},$$

where the probability of encounter, $p^* = \beta/(1+\beta)$ and $1-p^* = 1/(1+\beta)$. Therefore, $x_i \sim NB\left(\kappa,\ \beta/(1+\beta)\right)$ with mean κ/β. Given that the encounter rate is λ, where $\lambda = 2rt\mu_v D$, and the mean encounters from the Poisson-gamma mixture is κ/β, then in this case $\lambda = \kappa/\beta$ and $\beta = \kappa/\lambda$. Therefore, the probability density function of the encounters, x_i defined as

$$h(x_i; \lambda,\ \kappa) = \frac{\Gamma(\kappa+x_i)}{\Gamma(x_i+1)\Gamma(\kappa)}\left(\frac{\kappa}{\lambda+\kappa}\right)^{\kappa}\left(\frac{\lambda}{\lambda+\kappa}\right)^{x_i},$$

where $1/\kappa$ is the dispersion parameter. The negative binomial distribution has NB-2 form as given in Hilbe (2011).

2.2 Encounter Rate Formula in REM

Rowcliffe et al. (2008) derived the encounter rate formula in REM from a two dimensional ideal gas model, which we described in Section 2.1.1 above. In REM the encounter rate links the density with the speed of the individual and the detection zone dimensions of a camera trap. In the case of a camera trap, however, the detection zone is not circular but rather it has a sector-shaped zone as shown in Figure 2.2.1. For a sector-shaped detection zone, the width of the individuals' path would no longer be $2r$ but instead it would depend on the angle of approach.

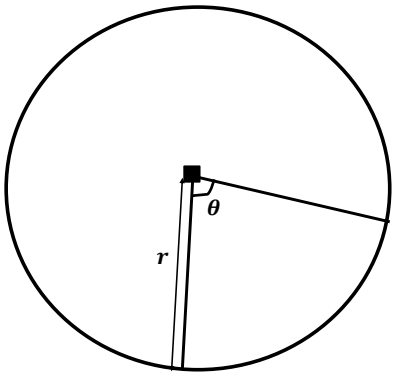


Figure 2.2.1: Representation of sensor detection width. The sector represents a sensor (e.g., camera); the detection zone of the sensor has width θ (radians), and distance r.

REM uses a stationary sector-shaped detection zone to represent the area seen by the camera trap as shown in Figure 2.2.1. The radius of the sector, r, is the detection distance and angle of the sector, θ, is the angle of detection. Now rather than a molecule (or particle) entering a detection zone, it is an animal being captured on film by a camera trap. If animals move within this sector-shaped detection zone they are captured with a probability of one; while outside this zone animals are captured with a probability of zero. As an animal can approach the sector-shaped sensor from any direction, the total path an animal can cover is no longer $2r$ but instead it changes with the approach angle between the sensor and the animal. The covered path within which an animal can be detected is called a profile, p. The path the animal takes is assumed to be equivalent to that of a molecule, however, as the contact zone is no longer circular the profile will

vary depending on the angle of approach. Figure 2.2.2 (page 25) shows 6 limiting cases of profiles, p, for π approach angles, where $\gamma_i = \{\gamma_1, ..., \gamma_7\}$ is the angle opposite the profile. The range of possible values γ_i can take is $0 \leq \gamma_i \leq \pi/2$ (see Rowcliffe et al., 2008).

Firstly, consider Figures (a) and (b) and possible approaches between the two figures. From Figure (b) the length of p can be determined using the trigonometric rule that $\sin(\gamma)$ is equal to opposite divided by the hypotenuse. The length of the hypotenuse can be determined from the triangle representing the camera, which is an isosceles triangle with angle, θ, and two sides of length r. The third side and hypotenuse has length $2r\sin(\theta/2)$; this is so because if we draw a perpendicular line from the vertex of the isosceles triangle at θ to the third side and we use the trigonometric rule of $\sin(\theta/2)$ is equal to opposite divided by hypotenuse, we can find p in terms of r and θ. Therefore,

$$\sin(\gamma) = \frac{p}{2r\sin\left(\frac{\theta}{2}\right)} \quad \text{or} \quad p = 2r\sin\left(\frac{\theta}{2}\right)\sin(\gamma). \tag{2.2.1}$$

In Figure (b) the angle γ_2 can also be found from the fact that the triangle is isosceles with angle θ, and two angles of γ_2. As the angles of a triangle add up to π we get, $\pi = \theta + 2\gamma_2$. Rearranging gives

$$\gamma_2 = \frac{\pi - \theta}{2}.$$

The profile arising from the transition between Figure (a) and Figure (b) is given by equation (2.2.1). The angle γ_1 is $\pi/2$. The contribution to the average for the directions between γ_1 and γ_2 is

$$\int_{\frac{\pi-\theta}{2}}^{\pi/2} 2r\sin\left(\frac{\theta}{2}\right)\sin(\gamma)d\gamma.$$

Secondly, consider Figures (b) and (c) and the possible approach directions between the two figures. To obtain the transition the angle of approach opposite the profile, γ_3 in Figure (c), is used. To get the profile, we start from Figure (b), and using the trigonometric rule, $\sin(\theta)$ is equal to opposite divided by the hypotenuse, we have

$$\sin(\theta) = \frac{p}{r} \quad \text{or} \quad p = r\sin(\theta). \tag{2.2.2}$$

In Figure (c), $p = r = r\sin(\pi/2)$. Between the two figures $p = r\sin(\gamma)$. Therefore, the contribution to the average for directions between Figure (b) and Figure (c) is

$$\int_{\theta}^{\pi/2} r\sin(\gamma)d\gamma.$$

Thirdly, for the transition between Figures (c) and (d), there is no angle opposite the profile, making the profile constant. Therefore, the profile, $p = r$. The angle $\gamma_4 = \theta$, and the contribution to the average for directions between Figures (c) and (d) is

$$\int_{0}^{\theta} r d\gamma.$$

Fourthly, for the transition between Figures (d) and (e), let us consider Figure (e). The angle of approach that is opposite the profile, p is γ_5. We can determine the profile p using the trigonometric rule $\sin(\theta)$ is opposite divided by hypotenuse, that is

$$\sin(\theta) = \frac{p}{r} \quad \text{or} \quad p = r\sin(\theta). \tag{2.2.3}$$

In Figure (e) $p = r = r\sin(\pi/2)$. Between the two figures $p = r\sin(\gamma)$. Therefore, the contribution to the average for directions between Figures (d) and (e) is

$$\int_{\theta}^{\pi/2} r\sin(\gamma)d\gamma.$$

Finally, consider Figures (e) and (f) and the possible approach directions between the two figures. Note that Figure (f) is similar to Figure (a) but with an opposite approach direction. We can determine p by using the trigonometric rule $\sin(\gamma)$ is equal to opposite divided by hypotenuse. The length of the hypotenuse can be determined from the triangle representing the camera, which is an isosceles with angle θ, and two sides of length r. The third side and the hypotenuse has length $2r\sin(\theta/2)$. Therefore, we obtain the equation

$$\sin(\gamma) = \frac{p}{2r\sin\left(\dfrac{\theta}{2}\right)} \quad \text{or} \quad p = 2r\sin\left(\frac{\theta}{2}\right)\sin(\gamma), \tag{2.2.4}$$

which is the exactly as equation (2.2.1). In Figure (e) $\gamma_5 = \theta$, and the angle γ_6 can be

found. Since angles of a triangle add up to π we get $\pi = \theta + 2\gamma_6$. Rearranging gives

$$\gamma_6 = \frac{\pi - \theta}{2}$$

The profile arising from the transition between Figures (e) and (f) is given by equation 2.2.4. The angle $\gamma_7 = \pi/2$. The contribution to the average directions between Figures (e) and (f) is

$$\int_{\frac{\pi-\theta}{2}}^{\pi/2} 2r \sin\left(\frac{\theta}{2}\right) \sin(\gamma) d\gamma.$$

Therefore, the widths of the profiles, p, and ranges of γ for each transition are given by: transitions from graph (a) to graph (b), and graph (e) to graph (f) $2r\sin(\theta/2)\sin(\gamma)$, $(\pi - \theta)/2 \leq \gamma \leq \pi/2$; transitions from graph (b) to graph (c) and graph (d) to graph (e), $r\sin(\gamma)$, $\theta \leq \gamma \leq \pi/2$; transition from graph (c) to graph (d), r for θ approach angles. REM is derived by integrating the sum of these profiles and averaging across π approach angles.

$$\frac{2\int_{(\pi-\theta)/2}^{\pi/2} 2r\sin\left(\frac{\theta}{2}\right)\sin(\gamma)d\gamma + 2\int_{\theta}^{\pi/2} r\sin(\gamma)d\gamma + \int_0^{\theta} r d\gamma}{\pi}$$
$$= r\frac{4\sin\left(\frac{\theta}{2}\right)\cos\left(\frac{\pi-\theta}{2}\right) + 2\cos(\theta) + \theta}{\pi} = r\frac{2+\theta}{\pi},$$

which is the total area an animal can cover for contact to occur in a sector-shaped detection zone. The expected number of contacts between animals and camera traps is therefore given by

$$\lambda = \frac{2+\theta}{\pi} r t \mu_v D. \tag{2.2.5}$$

Rearranging we can compute density, D as

$$D = \frac{\lambda}{t}\frac{\pi}{(2+\theta)r\mu_v}. \tag{2.2.6}$$

where λ/t is the encounter rate (see Rowcliffe et al., 2008).

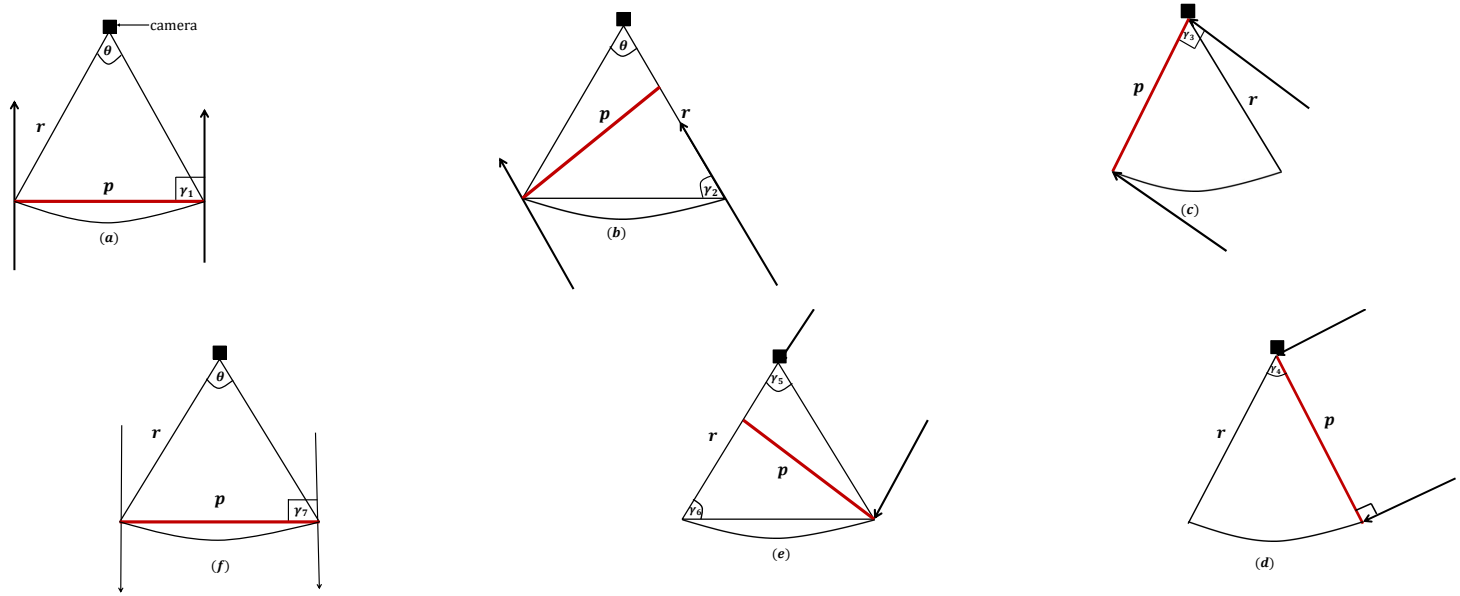


Figure 2.2.2: Graph showing approach directions of animals, indicated by the arrows, to a sector-shaped camera trap detection zone. The profiles are denoted by p, and γ_i is the angle of approach opposite the profile (Rowcliffe et al., 2008).

2.3 Estimating density in REM

In order to estimate density from camera trapping rate, independent estimates of the population expected speed of movement μ_v and average group size, g, which is discussed in Section 2.3.3 are required for each species. Rowcliffe et al. (2008) used day time photographs to derive an appropriate estimator of the expected speeds, μ_v. For a random sample of speeds $\{v_1, v_2, ..., v_n\}$ of size n, an appropriate estimator of the expected speed is the sample mean, denoted by $\bar{v}$. For the remainder of this Chapter the speed estimator $\bar{v}$ is used in the encounter rate formula for density estimation. The data and procedure for estimating the sample average speed and average group size are given in Section 2.7. From equation (2.2.5) the density in REM is therefore defined as

$$D = \frac{\lambda}{t}\frac{\pi}{(2+\theta)r\bar{v}}, \tag{2.3.1}$$

where r and θ are the dimensions of the camera, and t is duration of the time for which the camera was functioning. These parameters are all fixed. In Rowcliffe et al. (2008), density is estimated as

$$\hat{D} = \frac{\text{total no. encounters}}{\text{total no. camera trap days}} \times \frac{\pi}{(2+\theta)r\bar{v}}, \tag{2.3.2}$$

and the standard error is computed by bootstrapping the camera locations with replacement and taking the variance of a large number of resampled density estimates. REM is a nonparametric method, which could be derived by assuming a Poisson model for the number of encounters. The maximum likelihood estimator of the encounter rate, λ is the total number of encounters divided by total number of camera trap days. In Section 2.3.1 below we provide a maximum likelihood framework for REM.

2.3.1 Maximum likelihood estimation of density in REM

In this Section we provide a maximum likelihood framework for REM. In the two dimensional ideal gas model from which REM is derived, the number of contacts is assumed to follow a Poisson distribution, however, in Chapter 3 we relax this assumption. The maximum likelihood approach provides a framework for extending the model from a

fixed estimate of the speed (as shown in Chapter 3). The parametric framework allows for estimation of the variance of the density. Let the number of encounters

$$y_{ij} \sim \text{Poisson}(\lambda),$$

where $i = 1, 2, ..., c$, is the ith camera trap; $j = 1, 2, ..., n$ is the jth trap day; and λ is the expected encounter rate,

$$\lambda = \frac{2+\theta}{\pi} rt\bar{v}D. \tag{2.3.3}$$

The detection distance is denoted by r; the detection angle is θ, and $\bar{v}$ is the sample mean speed. The probability mass function of the encounters is

$$h_{pois}(y_{ij} \mid \lambda) = \frac{e^{-\lambda}}{y_{ij}!}\lambda^{y_{ij}}. \tag{2.3.4}$$

From equation (2.3.4) the likelihood can be constructed. We assume the encounters, y_{ij} are independent for all i and all j. That is, all animals are captured independently of each other, and independently at each camera trap. The assumptions in REM are based on the assumptions from the ideal gas model. A key assumption in the ideal gas model is that the movement of particles are independent, random and equally likely in all directions. In REM camera traps are placed randomly and the traps are not baited nor are the animals lured to the traps in any way. Therefore, the likelihood is,

$$L(y_{ij}; \lambda) = \prod_{i=1}^{c}\prod_{j=1}^{n} \frac{e^{-\lambda}\lambda^{y_{ij}}}{y_{ij}!}.$$

The log-likelihood is given as

$$\ell(y_{ij}; \lambda) = \text{constant} - cn\lambda + \sum_{i=1}^{c}\sum_{j=1}^{n} y_{ij}\log(\lambda).$$

Differentiating with respect to λ and setting equal to zero, we have an estimator of λ such that

$$\hat{\lambda} = \frac{\sum_{i=1}^{c}\sum_{j=1}^{n} y_{ij}}{cn}. \tag{2.3.5}$$

Using equation (2.3.5) an estimator of the density, D, conditional on the mean speed,

$\bar{v}$ in equation (2.3.1) is therefore

$$\hat{D} = \frac{\hat{\lambda}}{t} \times \frac{1}{K} = \frac{\sum_{i=1}^{c} \sum_{j=1}^{n} y_{ij}}{cn} \times \frac{1}{K},$$

where $K = \{(2+\theta)/\pi\}\, r\bar{v}$, and $\hat{D}$ is the maximum likelihood estimate conditional on $\bar{v}$. An estimate of the standard error is computed from the inverse Hessian matrix. An example of the R-code for the log-likelihood function above used to estimate density is given in appendix A.1.

The encounter data at WWAP were collected on a daily basis, that is for a fixed camera trap time period, $t = 1$ (day). Hence, we have organised the encounter data as y_{ij} ($i = 1, 2, ...c,\ j = 1, 2, ..., n$), the number of encounters per camera per day to facilitate later modelling that includes habitat and camera location, which are discussed in Chapter 4. However, regardless of the time scale used to aggregate the data the likelihood would be the same. This is true since the sum of independent Poisson random variables is a Poisson random variable with parameter equal to the sum of the individual Poisson parameter means. Therefore, the maximum likelihood estimate and its variance do not depend on the time scale used. We demonstrate below in Section 2.3.2 that the time scale used to aggregate the data has no effect on estimated density and its variance.

2.3.2 Varying the time scale used to aggregate encounter records

The data could be aggregated on a daily basis, over two days or a week, depending on the observers' choice, or alternatively, the data could be aggregated per camera as in Rowcliffe et al. (2008). Regardless of the way in which the data is aggregated the estimates of the density and its variance will be identical. To see this, let y_{ij} be the number of encounters recorded on the ith camera trap ($i = 1, 2, ..., c$) and jth ($= 1, 2, ..., n$) camera trap time period, which could be in hours, in days or in weeks. Here we let the time scale the data is collected be defined as t_1 (for example $t_1 = 1$ so that j is per unit time). Let x_{ij} be the number of encounters recorded on the ith camera trap ($i = 1, 2, ..., c$) and collected under a different time scale t_2 (for example, $t_2 = 2$ days, so that j is per two-days). Let Y_i be the number of encounters recorded in total for the ith camera, then let t_c be the number of days the cameras are out. The total

number of encounters is $S = \sum_{i=1}^{c} \sum_{j=1}^{n_1} y_{ij} = \sum_{i=1}^{c} \sum_{j=1}^{n_2} x_{ij} = \sum_{i=1}^{c} Y_i$, where n_1 is the total number of t_1 time periods, for example the number of days; and n_2 is the total number of t_2 time periods, for example the number of two-day period. We assumed the number of encounters are Poisson distributed with means, $\hat{\lambda}_1 = \sum_{i=1}^{c} \sum_{j=1}^{n_1} y_{ij}/cn_1$, $\hat{\lambda_2} = \sum_{i=1}^{c} \sum_{j=1}^{n_2} x_{ij}/cn_2$, and $\hat{\lambda}_c = \sum_{i=1}^{c} Y_i/c$, respectively. The total length of camera times used to record the y_{ij}, x_{ij} and Y_i encounters are $T = ct_1n_1$, $T = ct_2n_2$, and $T = ct_c$, respectively. Estimates of density, $\hat{D}_1$, $\hat{D}_2$, and $\hat{D}_c$ can be computed as

$$\hat{D}_1 = \frac{\sum_{i=1}^{c} \sum_{j=1}^{n_1} y_{ij}}{cn_1t_1} \times \frac{1}{K} = \frac{S}{T} \times \frac{1}{K},$$

$$\hat{D}_2 = \frac{\sum_{i=1}^{c} \sum_{j=1}^{n_2} x_{ij}}{cn_2t_2} \times \frac{1}{K} = \frac{S}{T} \times \frac{1}{K},$$

and,

$$\hat{D}_c = \frac{\sum_{i=1}^{c} Y_i}{ct_c} \times \frac{1}{K} = \frac{S}{T} \times \frac{1}{K},$$

respectively. These results are a standard for Poisson random variables, where the sum of two independent Poisson random variables is a Poisson random variable with mean equal the sum of the parameters of the individual Poisson random variables. Also, since λ and $\hat{D}$ depend on the total number of encounters the results, $\hat{D}_1 = \hat{D}_2 = \hat{D}_c$ follow immediately. Here $\hat{D}_c$ is exactly equal to equation (2.3.2) $K = \{(2+\theta)r\bar{v}\}/\pi$. Since $\hat{D}_1 = \hat{D}_2 = \hat{D}_c$, then the variances are equal and can be computed directly. Let y_{ij} be the number of encounters recorded on the ith camera for a time period of $t = 1$ day. The variance of estimated density $\hat{D}$ can be computed as follows:

$$\hat{\lambda} = \sum_{i=1}^{c} \sum_{j=1}^{n} \frac{y_{ij}}{cn} = \frac{S}{cn},$$

$\sum_{i=1}^{c} \sum_{j=1}^{n} y_{ij} = S$. The density $\hat{D}$ is

$$\hat{D} = \frac{\hat{\lambda}}{t} \times \frac{1}{K} = \frac{S}{cntK} = \frac{S}{TK},$$

where $T = cnt$. The variance of $\hat{\lambda}$ is

$$\begin{aligned}
\mathrm{Var}(\hat{\lambda}) &= \frac{\hat{\lambda}}{cn}. \\
\mathrm{Var}(\hat{D}) &= \mathrm{Var}\left(\frac{\hat{\lambda}}{Kt}\right) = \frac{1}{(tK)^2}\mathrm{Var}(\hat{\lambda}) \\
&= \frac{\hat{\lambda}}{t^2K^2cn} = \frac{S}{t^2K^2c^2n^2} = \frac{S}{T^2K^2}.
\end{aligned}$$

2.3.3 Estimating density for animals found in groups

To estimate density in REM an independent estimate of average group size for each species is required, if animals are found to be moving in groups. However, there are inherent difficulties with the use of camera traps in assessing group sizes. Buckland et al., (2001, Ch. 1), state that the detection probability of groups is dependent on both the distance from the point (in this case the camera trap) and the group size. And, if the centre of the group is inside the detection zone, then the count of the size of the group must include all individuals in the group, and distances should be measured from the point to the geometric centre of the group. But in camera trapping analysis the distance from the camera to the centre of the group is unobservable, and it becomes impossible to quantify the detection zone, particularly for larger, more dispersed groups and estimation of the density becomes difficult. However, if a count of the number of individuals in each observed group can be made, one can estimate the average group size in the population, and, in which case the density of individuals can be expressed as a product of density of the groups times the average group size; the simplest estimate of the average group size is the mean size of number of detected groups but detection may be a function of group size (see Buckland et al., (2001, ch. 1, page 13)).

Rowcliffe et al. (2008) tested REM using a small test data set, which we describe in Section 2.7, where group size was independently observable and was small enough to record accurately. In this case a count of the number of individuals was made, and average group size was estimable. Rowcliffe et al. (2008) suggest that the formula in equation (2.3.1) must, therefore, be modified for animals that move in groups, in which case the independent unit recorded by the camera is the group rather than the

individual. The parameter $\hat{D}$ in equation (2.3.1) is the density of the group (here we will denote density of the group as $\hat{D}_G$) and must be multiplied by an unbiased independent estimate of average group size, $\hat{g}$, to give the density of individual animals (as in Buckland et al. (2001)). As discussed in Section 2.7, average group sizes were estimated by conducting 10 watches (for each species), which were distributed evenly between 08:00 to 18:00 to control for any variation in movement patterns over the day, and systematically recording the numbers of individuals encountered along transects through the study area. With an estimate of the density of the group and group size, the density of the individual, $\hat{D}$ is estimated as

$$\hat{D} = \hat{D}_G \times \hat{g}.$$

2.3.4 Estimating animal density split by habitat

Animals generally respond to each other and their physical environment, so they may spend more time in some habitats than others and be restricted to partially overlapping home ranges (Hutchinson and Waser, 2007). As such the density of animals can vary with habitat type. In REM the encounter rate depends on the density and since REM is a nonparametric method Rowcliffe et al. (2008) estimated the density in each habitat separately. Here we provide a covariate framework for incorporating habitat into the model, which is also straightforward and can allow for the incorporation of other covariates such as camera random effects (as is done in Chapter 4), and other continuous covariates such as climatic conditions, which can include temperature and rainfall, elevations among others. Since the encounter rate cannot be negative, we use a log-link function and the covariate (habitat) enters the model through the encounter rate, λ.

Let the habitats be denoted by p such that $p = 1, 2, ..., H$. The model structure in equation 2.3.3 can be modified to incorporate habitat into the model in the following way

$$\boldsymbol{\lambda} = \left(\frac{2+\theta}{\pi}\bar{v}rt\right)\exp(\mathbf{X}\boldsymbol{\beta}), \tag{2.3.6}$$

where $\boldsymbol{\lambda}$ $(= \lambda_{n\times 1})$ is a vector of expected encounter rates, and n is the number of camera trap days; r and θ are the distance of detection and angle of detection, respectively;

$t = 1$ day is the camera trap time period across habitats, and $\bar{v}$ is the mean speed of the animals, which is assumed to be same in each habitat. These parameters are all fixed. The matrix containing the habitat-specific covariates

$$\mathbf{X} = \begin{bmatrix} 1 & x_{12} & x_{13} & \dots & x_{1H} \\ 1 & x_{22} & x_{23} & \dots & x_{2H} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_{n2} & x_{n3} & \dots & x_{nH} \end{bmatrix}_{n \times H}$$

where

$$x_{jp} = \begin{cases} 1 & \text{observation } j \text{ is from habitat } p. \\ 0 & \text{observation } j \text{ is not from habitat } p. \end{cases}$$

and

$$D_p = \begin{cases} \exp(\beta_1) & \text{for } p = 1. \\ \exp(\beta_1 + \beta_p) & \text{for } p = 2, 3, ..., H. \end{cases}$$

The parameter $\boldsymbol{\beta}_{1 \times H} = (\beta_1, \beta_2, ..., \beta_H)^T$ is the vector of regression coefficients. The mean density for a given species, D_T can be computed as

$$D_T = \sum_{p=1}^{H} \frac{A_p \exp(\beta_p)}{A_T} = \frac{\sum_{p=1}^{H} A_p D_p}{A_T}$$

where A_p $(p = 1, 2, ..., H)$ is the area of habitat p in the vicinity of the traps; and $A_T = A_1 + A_2 + ... + A_H$ is the total area surveyed. For known variance of D_p the variance of D_T can be computed directly as

$$\begin{aligned} \text{Var}(D_T) &= \text{Var}\left(\frac{\sum_{p=1}^{H} A_p D_p}{A_T}\right) \\ &= \sum_{p=1}^{H} \frac{A_p^2 \text{Var}(D_p)}{A_T^2}. \end{aligned}$$

2.4 Estimating variance

In this section we provide the different ways in which the variance of the density can be estimated. Rowcliffe et al. (2008) used a nonparametric bootstrap approach to approximate the variance of estimated density. We also adopt a nonparametric bootstrap

approach for estimating the variance of estimated density, which is discussed in Section 2.4.2. In Section 2.3.2 above we showed that the variance of the density can be estimated directly. However, this method cannot be extended to later Chapters. As such we use the inverse Hessian matrix from the optimization routine optim in R to approximate the variance. In Section 2.4.1 we also provide an adjusted variance method where the expected animal speed, $\bar{v}$, is not fixed.

2.4.1 Adjusted variance

The average animal speed of movement, $\bar{v}$, is an estimated constant from the given data. REM assumes $\bar{v}$ to be a fixed constant. Hence, the variance of estimated density is computed directly from equation (2.3.1),

$$\text{Var}(\hat{D}) = \text{Var}\left(\frac{\hat{\lambda}}{t}\frac{\pi}{(2+\theta)r\bar{v}}\right) = \left(\frac{\pi}{t(2+\theta)r\bar{v}}\right)^2 \text{Var}(\hat{\lambda}), \tag{2.4.1}$$

where the variance of $\hat{\lambda}$ is known. The variability of animal speed is not considered, which may cause the variance of estimated density to be underestimated. To correct for this underestimation in the variance, we assume that average animal speed, $\bar{v}$, is not a constant. Using Taylor series expansion we can find an approximate variance of estimated density such that

$$\text{Var}(\hat{D}) = \left(\frac{\pi}{t(2+\theta)r}\right)^2 \text{var}\left(\frac{\hat{\lambda}}{\bar{v}}\right), \tag{2.4.2}$$

where

$$\begin{aligned}
\text{Var}\left(\frac{\hat{\lambda}}{\bar{v}}\right) &\approx \frac{1}{\mathbb{E}^2(\bar{v})}\text{Var}(\hat{\lambda}) + 2\frac{-\mathbb{E}(\hat{\lambda})}{\mathbb{E}^3(\bar{v})}\text{Cov}(\hat{\lambda},\bar{v}) + \frac{\mathbb{E}^2(\hat{\lambda})}{\mathbb{E}^4(\bar{v})}\text{Var}(\bar{v}) \\
&= \frac{\mathbb{E}^2(\hat{\lambda})}{\mathbb{E}^2(\bar{v})}\left[\frac{\text{Var}(\hat{\lambda})}{\mathbb{E}^2(\hat{\lambda})} - 2\frac{\text{Cov}(\hat{\lambda},\bar{v})}{\mathbb{E}(\hat{\lambda})\mathbb{E}(\bar{v})} + \frac{\text{Var}(\bar{v})}{\mathbb{E}^2(\bar{v})}\right] \qquad (2.4.3) \\
&= \frac{\hat{\lambda}^2}{\bar{v}^2}\left[\frac{\text{Var}(\hat{\lambda})}{\hat{\lambda}^2} - 2\frac{\text{Cov}(\hat{\lambda},\bar{v})}{\hat{\lambda}\bar{v}} + \frac{\text{Var}(\bar{v})}{\bar{v}^2}\right].
\end{aligned}$$

(see Van Kempen and Van Vliet, 2000). In equation (2.4.3) the covariance $\mathrm{Cov}(\hat{\lambda}, \bar{v})$ is zero. This is so because the expected encounter rate, $\hat{\lambda}$, and average animal speed, $\bar{v}$, are assumed to be independent of each other. This is true for the WWAP data set discussed in Section 2.7 and in Rowcliffe et al. (2008), an underlying assumption of REM is that the data sets are independent of each other. To find an approximate variance of the sample mean animal speed, the formula given below is used

$$\mathrm{Var}(\bar{v}) = \frac{{\sigma_v}^2}{m}, \tag{2.4.4}$$

where σ_v is the sample standard deviation of animal speed; and m is the sample size of animal speed data.

2.4.2 Nonparametric Bootstrap Method

It is also possible to use bootstrapping to estimate the variance of the estimated density. Rowcliffe et al. (2008) used a nonparametric bootstrap approach to estimate the variance of the estimated density. This was done by resampling camera locations with replacement and taking the variance of a large number of resampled density estimates. Note that the current REM formula lacks the potential to account for the sampling variability in the speed of movement. So, to allow for some variability in the animal speed, which would improve estimation precision we use an alternative nonparametric bootstrap approach, resampling both the speed data and the encounter data. Nonparametric resampling allows us to estimate the sampling distribution of a statistic empirically without making assumptions concerning the distribution of, or model for, the data. Here we give the general algorithm for a nonparametric bootstrap. Suppose we have a vector $\mathbf{x}$ of m independent observations, and we are interested in estimating a parameter $\hat{\theta}(\mathbf{x})$ and its variance. The general nonparametric bootstrap algorithm is as follows:

1. Sample m observations randomly with replacement from $\mathbf{x}$ to obtain a bootstrap data set, denoted by $\mathbf{x}^*$.

2. Calculate the bootstrap version of the statistic of interest, $\theta^* = \hat{\theta}(\mathbf{x}^*)$.

3. Repeat steps 1 and 2 a large number of times, say B, to obtain an estimate of the

bootstrap distribution

4. calculate the average of the bootstrapped statistics, $\sum_{b=1}^{B} \theta^*{}_{(b)}/B$

5. compute the variance of the estimator $\hat{\theta}(\mathbf{x})$ through the variance of the set $\theta^*{}_{(b)}$, $b = 1, 2, ..., B$, given by

$$\frac{\sum_{b=1}^{B} \left(\theta^*{}_{(b)} - \theta^*{}_{(\cdot)}\right)^2}{(B-1)}$$

where $\theta^*{}_{(\cdot)} = \sum_{b=1}^{B} \theta^*{}_{(b)}/B$ (see Carpenter and Bithell, 2000).

2.5 Constructing Confidence Intervals for Density

A traditional approach to statistical inference is to make assumptions about the structure of the population, for example an assumption of normality, and to use these assumptions to derive the sampling distribution of a statistic, for example population density, on which classical inference is based. If the assumptions of the population are wrong, then the corresponding sampling distribution of the statistic may be seriously inaccurate. On the other hand, if the asymptotic results are relied upon, these may not hold to the required level of accuracy in a relatively small sample (see Fox, 2002). Since the density, D cannot be negative we consider approaches other than the $normal - theory\ interval$ to construct appropriate 95% confidence intervals. We consider logarithmic transformations and the percentile method to provide appropriate confidence intervals of the density.

2.5.1 Natural Logarithm Transformation

For small to moderate amounts of data, the distribution of a statistic may be asymmetrically skewed toward large values. To counteract the skewness, we can set the confidence limits on the log scale (that is, after a logarithmic transformation). For the density we can use the following equation to determine a 95% confidence interval.

$$\{\ln(\hat{D}_L), \ln(\hat{D}_U)\} = \ln(\hat{D}) \pm 1.96 \times \mathrm{SE}\left\{\ln(\hat{D})\right\}, \tag{2.5.1}$$

where the variance is approximated by the inverse Hessian matrix from optim in R, and hence, the standard error of the logarithm of the density, SE $\left\{\ln(\hat{D})\right\}$ is computed. The term ln() is the natural logarithm transformation, $\hat{D}_L$ is the lower confidence limit and $\hat{D}_U$ is the upper confidence limit. Because this equation (2.5.1) gives the confidence limits on the log scale, the limits need to be converted back to the $\hat{D}$ scale after they are calculated, by reversing the transformation, which involves taking antilogarithms. Therefore, for a 95% confidence interval of the density we have

$$\{\hat{D}_L, \hat{D}_U\} = e^{\left(\ln(\hat{D}) \pm 1.96 \times \mathrm{SE}\left\{\ln(\hat{D})\right\}\right)}. \tag{2.5.2}$$

(Rothman, 2012, Ch. 8, pages 161-162).

2.5.2 *Percentile Method*

The percentile method is a method for approximating bootstrap confidence intervals. The percentile method makes no assumption about the distribution of the observations. The procedure for finding the percentile confidence interval is outlined below.

- Suppose we draw a sample $\mathbf{x} = \{X_1, X_2, ..., X_m\}$ of size m from a population $\mathbf{P} = \{x_1, x_2, ..., x_M\}$, sampling with replacement. Call the resulting bootstrap sample $\mathbf{x_1}^* = \{X_{11}{}^*, X_{12}{}^*, ..., X_{1m}{}^*\}$. And, suppose we are interested in some statistic $T = t(\mathbf{x})$ as an estimate of the corresponding population parameter $\theta = t(\mathbf{P})$. We repeat this procedure a large number of times, B, selecting many bootstrap samples; the bth such bootstrap samples is denoted by $\mathbf{x_b}^* = \{X_{b1}{}^*, X_{b2}{}^*, ..., X_{bm}{}^*\}$, and computing the statistic T for each bootstrap sample, that is $T_b{}^* = t(\mathbf{x_b}^*)$. Then the distribution of $T_b{}^*$ around the original estimate T is analogous to the sampling distribution of the estimator T around the population parameter θ. Therefore, the average of the bootstrap statistics,

$$\bar{T}^* = \frac{\sum_{b=1}^{B} T_b{}^*}{B}$$

- Using the percentiles $\alpha/2$ and $1 - \alpha/2$, where α is the significance level, compute the interval. To do this sort the estimates $T^*{}_b$ as $T^*{}_1 \leq T^*{}_2 \leq ... \leq T^*{}_B$ and use $T^*{}_{25}$ to estimate the 2.5th percentile of $\bar{T}^*$ and use $T^*{}_{975}$ to estimate the 97.5th

percentile of $\bar{T}^*$, so that the desired 95% confidence interval is

$$T^*{}_{25} < \theta < T^*{}_{975}$$

To obtain sufficiently accurate 95% bootstrap percentile confidence intervals, the number of bootstrap samples, B, should be on the order of 1000 or more. The percentile method implicitly assumes the sampling distribution of $T = t(\mathbf{x})$ is symmetric, but not necessarily normal, and centred at $\theta = t(\mathbf{P})$ (unbiased). However, the coverage error is often substantial if the distribution of $\hat{\theta}$ is not nearly symmetric, however, other methods such as the Bias-Corrected and Bias-Corrected Accelerated can be used to address this issue, but these are not considered here (see Fox, 2002; Davison and Hinkley, 1997; Carpenter and Bithell, 2000).

2.6 Assumptions in REM

The assumptions in REM are based on the assumptions from the ideal gas model discussed in Section 2.1. A key assumption in the ideal gas model is that the movement of particles is independent, and equally likely in all directions. In REM the movement of animals is assumed to be random and independent relative to one another. Clearly, this is unrealistic in natural settings as animals respond to one another and their physical environment. However, even though animals respond to each other and their natural settings the results from the Whipsnade Wild Animal Park field test by Rowcliffe et al. (2008) provided an adequate approximation of the detection process with no apparent bias in density estimates for three of the four species. This suggests that the REM method is reasonably robust to typical behaviour patterns that may violate underlying model assumptions. Rowcliffe et al. (2003) applied an equivalent method to model rates of capture by snares and found it to be successful across a wide range of species. Some animals are also found in pairs or family groups, which may violate model assumptions. This evidence suggests that REM is robust to typical behaviour patterns that may violate model assumptions.

Another important assumption of REM is that animals move independently of cameras.

For example, if animals are trap-shy, that is, they avoid the camera units themselves or the flash then this assumption may be violated. Some animals may also be attracted to the camera traps and may consistently return to a camera trap unit. The assumption will also be violated in this case. Note that some camera trapping methods use baits to attract animals, which may significantly increase capture rates (see for example du Preez et al., 2014; Thorn et al., 2009). However, baiting and luring of animals to camera traps violates the underpinning assumption of REM. To solve the problems where animals are trap shy or trap-attracted Rowcliffe et al. (2008) suggested using infra-red imaging instead of flash photography or, if the animal of interest is at least partly diurnal, by disabling the flash and relying on natural light and day-time photographs only. Also, the camera traps can be set up in a such a way that they are protected by tree covers to avoid them being too conspicuous. REM does not allow cameras to target focal species, instead REM requires randomized placement of camera traps. However, too rigid randomization of camera placements will be unrealistic in many situations as some species are highly elusive and rare; the probability of detection is minimal and the data collected will be too sparse to be of any use. Rowcliffe et al. (2008) suggest that a camera traps must be carefully randomized relative to the distribution, giving each camera a clear view to provide a reasonable chance of detecting animals.

Rowcliffe et al. (2008) state that the data collected must be in the form of numbers of independent contacts between animal (individual or group) and camera. To obtain independent contacts between animals and camera traps, it is required that an animal leaves the camera detection zone after a contact, and that the same or a different animal later re-enters in order to give a second independent contact. However, there is the problem of having several photographs from a single effective contact because of an animal lingering in the detection zone, or a large group passing the trap zone, which Rowcliffe et al. (2008) addressed by limiting the amount of film in the camera. This is done by setting the camera to be inactive for 2 minutes after each photograph. Rowcliffe et al. (2008) pointed out that a long latency period runs the risk of missing independent contacts occurring in quick succession, and they have suggested little or no camera latency, and that further work on camera latency would be useful.

2.7 The data

In Rowcliffe et al. (2008) REM was tested using data from a survey carried out over a 6 weeks period from 13 June to the 24 July 2005 at Whipsnade Wild Animal Park (WWAP), located in Bedfordshire, south England. The advantage of this data set is that it is from a zoo with a known number of animals. There was a total of $n = 42$ camera trap days. The park houses several free-ranging species but only four of these species were considered for the purposes of the analysis:

- red necked wallaby (*Macropus rufogriseus*)
- Chinese water deer (*Hydropotes inermis*)
- Revee's muntjac (*Muntiacus reevesi*), and
- mara (*Dolichotis patagonum*).

Note that for brevity, the common names: wallaby, water deer, mara and muntjac will be used in the rest of this thesis. The park was divided into four areas of contrasting habitat. The first two areas are open grasslands with scattered scrub:

1. *Downs* open grassland with scattered scrub and a steep scarp slope;
2. *Institute Paddock* with gentle slopes;
3. *Old Farm*, which is an area of rough grassland and thicket on largely level but highly broken ground; and
4. *Central Park*, which is an area of mixed lawns, roads, buildings and enclosures housing large animals with scattered trees.

2.7.1 Estimating Speed and Group Size

The calculation of density requires independent estimates of speeds, which could only be obtained during the day, and average group size. Rowcliffe et al. (2008) used day range to derive an approximate estimator of speed. To derive estimators of speed and average group size Rowcliffe et al. (2008) conducted 10 focal watches, for each species,

distributed evenly between 08:00 and 18:00 in order to control for any variation in movement patterns over the day. For the speed estimator, Rowcliffe et al. (2008) arbitrarily select individuals following each individual for 30 minutes during each focal watch and recording the total distance travelled during that time as the sum of all straight-line movements. Day range for each species was then calculated as the mean across all focal watches for that species. So the distribution of the speed data over a day was used as the distribution of speeds over the trapping period. It was observed that some animals were not moving during the period they were watched, hence, a zero speed was recorded. While the probability of observing a zero speed would be sensitive to observation timescale, tending to zero as timescale increases, the aim was to derive an estimate of average speed, which is insensitive to timescale, and not to obtain an unbiased estimate of the probability of observing a zero speed.

Average group sizes were estimated by systematically recording the numbers of individuals in groups encountered along transects through the study area. In Table 2.7.1 we give a summary of the fixed parameters required to estimate the density. Note that the angle of detection, θ in Table 2.7.1 is quite narrow since the camera traps used was a model called DeerCam, which has this limitation. These results are taken from Rowcliffe et al. (2008).

We also provide the coefficient of variation (C_v) of the speed estimators for each species. The coefficient of variation is a measure of dispersion of the data relative to the mean. It is defined as the sample standard deviation, σ divided by the mean, that is $C_v = \sigma/\mu$, where μ is the mean speed. Often the C_v is expressed as a percentage, which corresponds to the following formula

$$C_v\% = 100(\sigma/\mu). \tag{2.7.1}$$

The C_v is useful for comparing the variability of two or more samples of data from different variables or from the same variables when the means are very different (Brown, 1998). In Table 2.7.1 the $C_v\%$ is quite large for the encounter rate and the speed estimators of the four species suggesting that the speed of movement variation among

animals and the variability in the encounters are substantial.

Table 2.7.1: Summary of fixed parameters required to estimate density (standard error in parentheses), which is taken from Rowcliffe et al. (2008).

	mara	muntjac	wallaby	water deer
Mean encounter ($\hat{\lambda}$)	0.06 (0.02)	0.35 (0.07)	4.78 (0.47)	1.51 (0.19)
C_v% of encounter	389%	239%	117%	154%
Average day range ($\bar{v}$, in km day^{-1})	2.56 (1.08)	8.27 (1.92)	0.71 (0.36)	1.17 (0.49)
C_v% of day range	133%	73%	160%	127%
Average group size ($\hat{g}$)	1.8 (0.63)	1.5 (0.53)	1 (0)	1 (0)
Sample size of speed (m)	10	10	10	10
Detection arc (θ, radians)	0.175			
Detection distance (r, km)	0.012			

2.7.2 Animal Census

A census, counting the number of animals, was carried out at the end of the camera trapping period between 09:30 and 14:00 by a team of 12 counters. A systematic co-ordinated line approach (to avoid double counting), was adopted to cover areas in three of the four habitats (Downs, Institute Paddock, Old Farm) in a single sweep, while in the Central Park area small teams systematically counted central areas, co-ordinating movements to ensure complete coverage without double counting (see Rowcliffe et al., 2008). Since the census was taken during a specific time period, and animals moved among habitats there were observed zero census count in some habitats for some species, for example, the census of the mara species in Old Farm. However, during the camera trapping period there were records of the mara species in Old Farm (see Table 2.7.2). Table 2.7.2 gives the summary of the census data split by habitats for the four species at WWAP.

Table 2.7.2: Summary of the census data split by habitat, which is taken from Rowcliffe et al. (2008).

	Habitat				
	Downs	Institute Paddock	Old Farm	Central Park	Total
Area (km^2)	0.49	0.28	0.23	1.26	2.26
Camera hours (day time)	898	440	543	317	2198
Census count					
mara	15	2	0	136	153
muntjac	3	0	5	22	30
wallaby	544	213	185	120	1062
water deer	36	10	133	91	270
Density (animals km^2**)**					
mara	30	7	0	108	68
muntjac	6	0	22	18	13
wallaby	1101	760	803	96	468
water deer	73	36	577	72	119
Day-time camera trap photos					
mara	3	1	2	3	9
muntjac	10	1	4	23	38
wallaby	225	195	78	38	536
water deer	32	3	89	23	147

2.8 Application of REM to field data at WWAP

This section gives the results of REM applied to the Whipsnade Wild Animal Park data set. Section 2.8.1 gives the results of Rowcliffe et al. (2008). In Section 2.8.2 the results of estimated density using maximum likelihood estimation are given. The model was fitted using the optim function in the R software package (R version 3.2.4 or earlier) using the Brent algorithm. Density estimates split by habitat are given in Section 2.8.3.

2.8.1 Results of Rowcliffe et al. (2008)

Rowcliffe et al. (2008) observed that animals generally did not move about much, particularly the wallabies, and like the water deers were not seen in cohesive groups. The muntjacs, however, were considerably more mobile than the other species, and like the maras were frequently found in pairs or family groups. REM has shown to perform well in extracting absolute densities from encounter rate data. Rowcliffe et al. (2008) found that the estimated densities for muntjac, wallaby and water deer did not differ

significantly from, and were within 22% of the census results. The problem with the mara estimates is a result of poor survey design. The mara like to inhabit open areas, and during the WWAP survey the mara were mostly seen in Central Park, which is an open area most frequented by people. So in order to avoid crowding camera traps with too many human photographs, camera placements were located away from these crowded areas, and where the maras did not graze frequently. Hence, limited capture data were recorded, leading to the severe underestimation of the density.

2.8.2 Results from maximum likelihood method

Table 2.8.1 below compares the density from the census with estimated density for the four species at WWAP. Here we give the estimated standard error of density from the direct method, inverse of the negative Hessian matrix, the adjusted variance method, and bootstrap method where 1000 bootstrap samples of animal speeds and encounter data are used in the estimation process. We assume the encounters are Poisson distributed (with mean and variance being equal).

The results (Table 2.8.1) show that estimated standard errors from the adjusted variance method and bootstrap method are substantial compared with the estimates from the direct method and the Hessian matrix. In particular, the bootstrap estimates, which are considerably larger than those from the adjusted variance method, are skewed as can be seen from the average of the estimated densities and the estimate of the median density value. However, when compared on the log-scale, the similarity of the standard errors are clear, as expected, since both methods are attempting to account for the variability that comes from using an estimate of the mean speed. As can be seen in Table 2.7.1 above, the variability in the speed data and encounter data is high, as shown by the huge values of the coefficient of variation $C_v\%$. As such, we expect the estimated standard errors from the adjusted variance method to be high, particularly for the wallaby and water deer species (see Table 2.8.1). For the direct method and the Hessian method, the estimated standard errors are the same, as expected, and considerably smaller that those from the other methods. It is worth noting that the inverse of the negative Hessian is an estimator of the asymptotic covariance matrix and, the

standard error estimates based on the Hessian of the log-likelihood function are guaranteed to be valid asymptotically, under general regularity conditions, as sample sizes become infinite. However, for finite sample sizes the standard error estimates based on the Hessian may be biased. So the difference between the standard error estimates from the Hessian and the standard error estimates from the adjusted method and the bootstrap method is due to finite sample sizes, as well as possible non-normality of the sample. The data set may have come from a distribution with heavier tails than that of the corresponding normal distribution.

We also provide 95% confidence intervals of the density from the four methods. We use logarithm transformations described in Section 2.5.1 to obtain confidence intervals from the direct variance method, adjusted variance method and the inverse of the negative Hessian matrix. The percentile method, discussed in Section 2.5.2 is used to obtain bootstrap confidence intervals of the density. The mara and muntjac species were found to be moving in family groups during the WWAP survey, as such REM is modified, using equation (2.3.3) to obtain estimated density of the individual. The 95% confidence intervals from the adjusted variance method and the bootstrap method are comparable and are wide, and the density from the census for three of the four species are captured within the interval for the bootstrap method; and for two of the four species within the interval for the adjusted variance method. Again, the confidence interval from the Hessian is biased given that it relies on asymptotic theory. When sample size is not large enough bootstrap standard errors and confidence intervals may be more appropriate (Hall, 2013). As shown in Table 2.8.1 the density from the census, for all species, is not captured within an approximate 95% confidence interval based on the Hessian and the direct method. Finally, the estimated density for each species is different from the mean density from census because the Poisson REM developed in this Chapter does not account for variability in the density in heterogeneous habitats. Therefore, in Section 2.8.3, we incorporate habitat in the model to determine its effect on the density.

Table 2.8.1: Density from census data compared with estimated density. Standard errors from the direct method, the inverse Hessian matrix, and errors using the adjusted variance method are compared with bootstrap standard errors.

Species	census (km^2)	Estimate	Bootstrap estimates		Estimated speed	Direct method	Hessian method	Adjusted method		Bootstrap method	
	D	$\hat{D}$	(average) $\hat{D}$	(median) $\hat{D}$	$\bar{\hat{v}}$ $(Se\{\bar{\hat{v}}\})$	$Se(\hat{D})$	$Se(\hat{D})$	$Se(\hat{D})$	$Se\{\log(\hat{D})\}$	$Se(\hat{D})$	$Se\{\log(\hat{D})\}$
mara	68	5	7	5	2.52 (1.02)	1.76	1.76	1.80	0.34	11.22	0.63
muntjac	13	8	8	8	8.24 (3.27)	1.07	1.07	1.46	0.19	2.50	0.31
wallaby	468	816	821	1020	0.72 (0.11)	31.10	31.10	414.19	0.51	635.09	0.52
water deer	119	155	192	158	1.15 (0.19)	10.52	10.52	63.33	0.41	118.70	0.46

Table 2.8.2: A 95% confidence interval (CI) of the density from the four methods used to estimate variance is given for each species.

Species	census (D, km^2)	**Direct method CI**	**Hessian method CI**	**Adjusted method CI**	**Bootstrap method CI**
mara	68	(2.75, 10.17)	(2.75, 10.17)	(2.71, 10.33)	(1.96, 22.46)
muntjac	13	(5.74, 10.00)	(5.74, 10.00)	(5.19, 11.06)	(4.02, 13.79)
wallaby	468	(757.53, 879.54)	(757.53, 879.54)	(301.43, 2206.75)	(373.26, 2733.97)
water deer	119	(136.06, 177.44)	(136.06, 177.44)	(69.90, 345.39)	(81.57, 500.84)

2.8.3 *Density estimates split by habitat*

This section gives the results of density split by habitat. Estimated standard error from the inverse Hessian matrix is compared with estimated standard error from the bootstrap method. We use 1000 bootstrap samples of animal speed and encounter data to compute the standard error, and an approximate 95% confidence interval is given. Logarithm transformation is used to compute confidence intervals based on the Hessian of the log-likelihood function and the percentile method is used to compute bootstrap confidence intervals of the density. As noted in Section 2.7.2 the census data was collected at the end of the trapping period and since animals moved among habitats there were records of zero census counts in some habitats for some species, even though there were encounter records during the trapping period (see Table 2.7.1).

Table 2.8.3 gives the results for the mara species from including habitat as a covariate in the model. We provide estimates of the density using the standard method and bootstrap method, comparing the average density estimate with median density estimates. The estimated standard errors from the bootstrap method are substantial compared to estimates from the Hessian because the bootstrap estimates are very skewed (huge differences between the average density estimates and the median density estimates). We therefore compare standard errors on the log-scale. Table 2.8.3 shows that the mean estimated density is 8, an increase of 6 when habitat is incorporated in the model (estimated density from a Poisson REM without habitat is 2; see Table 2.8.1 in Section 2.8.2). However, this estimate is still nowhere near the density from the census. Also, in Old Farm where the observed census is 0, an estimated density of 5 is obtained due to the fact that encounters were recorded during the trapping period (see Table 2.7.2), and the fixed term, $C = \{(2+\theta)/\pi\}rt\bar{v}$ in the encounter rate estimator in equation (2.3.6). As expected, the bootstrap standard error is larger than the error obtained from the Hessian since some of the variability in the speed data is included in the model. Also, given that the standard error estimates from the Hessian relies on asymptotic theory, these may be biased for finite sample sizes. Even though the bootstrap confidence intervals are wider than the confidence intervals based on the Hessian, for both methods the density from the census in only one habitat is captured within an approximate 95%

confidence interval.

Similar results are obtained for the muntjac species in Table 2.8.4. The standard error estimates and confidence intervals from the bootstrap method are larger and wider, respectively, compared with those based on the Hessian of the log-likelihood function since bootstrapping the data accounts for the variability in the data, and, given that the the standard error estimates from Hessian rely on asymptotic theory, these might be biased for finite (and particularly small) sample sizes. For the muntjac species there was a zero census record in Institute Paddock but there were records of encounter during the trapping period (see Table 2.7.2 above). This is because animals can move between habitats, and because of this we have a non-zero estimate of density in this habitat. This is also due to the fixed term in the encounter rate estimator in equation (2.3.6) above. Incorporating habitat in REM has improved the mean density estimate across habitats, and density from the census in three of the four habitats and the mean density from the census are captured within an approximate 95% confidence interval based on the Hessian matrix and a 95% confidence interval based on the bootstrap method.

For the wallaby species (Table A.2.1, appendix A.2) incorporating habitat in the model improved the mean density across habitats as this is closer to the density from the census. However, the confidence intervals based on the Hessian does not capture the mean density from the census within an approximate 95% confidence interval. Also, the standard error estimates and confidence intervals based on the Hessian are smaller and more narrow, respectively, compared with those from the bootstrap method. This is due to the fact that the estimates from the Hessian rely on the assumption that sample size approaches infinity, which would be biased for finite (small) sample sizes. Only the density from the census in Old Farm is captured within an approximate 95% confidence interval based on the Hessian. For the bootstrap method the density from the census in three of the four habitats, and the mean density from the census are captured within a 95% confidence interval.

Like the wallaby species in Table A.2.1, appendix A.2 the mean density from the census

of the water deer species (Table A.2.2, appendix A.2) is not captured within an approximate 95% confidence interval based on the estimates from the Hessian. As expected, wider confidence intervals are obtained from the bootstrap method, and density from the census for two of the four habitats, and the mean density from the census are captured within a 95% confidence interval. These results suggest that accounting for the variability in habitats is important in estimating the density. The mean density across habitats are much closer to the density from the census compared with the estimated density when habitat is not included in the model; see Table 2.8.1 in Section 2.8.2. Also, the standard errors from the bootstrap method are always larger than those based on the Hessian, which are typically quite inaccurate when sample sizes are not large enough. As such, bootstrap standard errors and confidence intervals may be more appropriate when sample sizes are small.

Table 2.8.3: Density from the census of the mara species compared with estimated density. The estimated standard error of the density, and estimated standard error on the log-scale are given. We give 95% Confidence Intervals (CI) from the inverse of the negative Hessian matrix and Bootstrap method, using the standard errors on the log-scale are given.

Habitat	Census Density D (in km^2)	Estimate $\hat{D}$	Bootstrap estimates (average) $\hat{D}$	(median) $\hat{D}$	Hessian errors $Se(\hat{D})$	$Se\{\log(\hat{D})\}$	Bootstrap errors $Se(\hat{D})$	$Se\{\log(\hat{D})\}$	Hessian CI	Bootstrap CI
Downs	30	5	6	4	2.67	0.58	6.50	0.96	(1.49, 14.32)	(0.24, 18.58)
Institute Paddock	7	3	4	3	2.82	1.00	5.24	1.47	(0.40, 20.04)	(0.24, 15.30)
Old Farm	0	5	6	5	3.42	0.71	7.74	1.26	(1.21, 19.35)	(0.24, 22.38)
Central Park	108	11	13	10	6.11	0.58	16.35	1.01	(3.41, 32.82)	(0.24, 39.37)
Total	68	8	10	7	24.43	5.32	66.23	8.90	(2.23, 29.35)	(1.59, 58.80)

Table 2.8.4: Density from the census of the muntjac species compared with estimated density. The estimated standard error of the density, and estimated standard error on the log-scale are given. We give 95% Confidence Intervals (CI) from the inverse of the negative Hessian matrix and Bootstrap method, using the standard errors on the log-scale are given.

Habitat	Census Density D (in km^2)	Estimate $\hat{D}$	Bootstrap estimates (average) $\hat{D}$	(median) $\hat{D}$	Hessian errors $Se(\hat{D})$	$Se\{\log(\hat{D})\}$	Bootstrap errors $Se(\hat{D})$	$Se\{\log(\hat{D})\}$	Hessian CI	Bootstrap CI
Downs	6	6	7	6	1.59	0.25	2.53	0.41	(3.89, 10.36)	(2.43, 12.23)
Institute Paddock	0	1	1	1	0.73	1.00	0.78	0.89	(0.10, 5.16)	(0.20, 3.05)
Old Farm	22	4	4	4	1.53	0.41	1.96	0.57	(1.68, 8.33)	(0.86, 8.61)
Central Park	18	25	26	24	4.73	0.19	8.94	0.34	(16.84, 35.80)	(11.69, 47.30)
Total	13	16	17	15	25.53	4.21	43.18	7.10	(9.35, 26.61)	(7.13, 38.37)

2.9 Discussion

In this Chapter we introduced the Random Encounter Model developed by Rowcliffe et al. (2008). The main focus of the Chapter has been on the development of a maximum likelihood framework, which can later be extended, to estimate density of unmarked animals from camera trap data in REM.

We started by explaining the ideal gas model for which REM is based, and gave the derivation of REM by Rowcliffe et al. (2008). The ideal gas model is a model for collision rates or encounters, which depends on three things: the size and speed of the particles and their density (Yapp, 1956). REM uses this concept to determine the rate of encounter between animals and camera traps. The encounter rate in REM is a function of the dimensions of the camera, the density and an estimate of animal speed of movement, which is treated as a fixed value. In this Chapter we have illustrated the maximum likelihood REM formula using a small data set from Whispsnade Wild Animal Park, and the results were compared with the results from Rowcliffe et al. (2008).

Rowcliffe et al. (2008) used the total encounters divided by the total number of camera trapping days and multiplied by a constant term, which includes the average speed and the dimensions of the camera trap. This formula gives the equivalent of a maximum likelihood estimate with an underlying assumption of a Poisson model, which we have shown. To estimate the variance of the estimated density Rowcliffe et al. (2008) used bootstrapping by resampling camera locations with replacement and taking the variance of a large number of resampled density estimates. Our approach is different from Rowcliffe et al. (2008) in that we modelled the encounter data using a Poisson REM while assuming a fixed value of animal speed as in Rowcliffe et al. (2008). We have estimated the variance of estimated density using four methods: inverse of the negative Hessian matrix, which comes from the optim function for minimizing the negative log-likelihood in R; an adjusted variance method, which uses Taylor expansion to approximate the variance; bootstrap method, which involves resampling the animal speed data and encounter data, and the direct standard error method.

Testing REM using the WWAP data Rowcliffe et al. (2008) found that the estimated densities for three of the four species (wallaby, water deer and muntjac) did not differ significantly from, and were within 22% of the census results. The underestimation of the mara species was a result of nonrandom placements of camera traps in areas where the mara frequented during the survey period. The estimates of the density from the maximum likelihood framework and the density from the census differ substantially since the model does not account for the variability in the density in heterogeneous habitats. Also, the current REM formula does not account for the variability in animal speed, which would have an effect on standard error estimates. Estimates of the standard error based on the direct method and the Hessian of the log-likelihood function are the same, as expected, and are guaranteed to be valid asymptotically, as the sample size becomes infinite. However, the standard error estimates are biased since the sample size is small, as well as possible non-normality of the sample. The confidence intervals based on the Hessian are also narrow and do not capture the density from the census within an approximate confidence for all species. The standard error estimates based on the adjusted variance method and the bootstrap method are comparable, and the density from the census is captured within an approximate 95% confidence interval. The variability in the speed is huge, and the standard error estimates of the estimated density from the adjusted variance method is dominated by this uncertainty, which resulted in wider confidence intervals. The bootstrap method also allows for the variability in the speed as well as the variability in the encounter data. The advantage of using the adjusted variance method and the nonparametric bootstrap method for estimation precision is that they do not assume any distribution for the data. However, it is worth noting that the adjusted variance method and the direct method are limited in their use and cannot be used to estimate standard errors of the density for more complex models. For example, it may be possible to use the adjusted variance method for the Poisson REM with habitat, but it may not be easy to do so.

We have considered including habitat as a covariate in REM to determine whether it would have an effect on estimated density. We found that for the mara species the estimated mean density across habitats increased but the increase is nowhere near the

density from the census. For the muntjac and wallaby species the estimated mean density improved substantially when habitat is considered as it is closer to the density from the census, and the density from the census for the muntjac species is captured within a approximate 95% confidence interval based on the Hessian of the log-likelihood function and the bootstrap method. For the water deer species the mean density is overestimated and the bias is larger compared with the bias when habitat is not included in the model. The mean density from the census is also not captured within an approximate 95% confidence interval based on the estimates from the Hessian and the bootstrap method. It is worth noting that the adjusted variance method is limited in its flexibility. While it may be possible to estimate variance within and across habitats using the adjusted variance method, it may not be easy to do so.

The method adopted by Rowcliffe et al. (2008) for estimating density of individuals moving in groups, from camera trap data, is limited to the case that group sizes are small and counts of the number of individuals in each group is possible. There is an inherent difficulty with camera traps in that it is not possible to measure true group size. Also, if the group size is dependent on the detection distance, there would be difficulties in obtaining an unbiased estimate of the expected group size. This dependence arises because large groups are more likely to be detected further away from the camera, while small groups might remain undetected. Buckland et al. (2001) suggest that this phenomenon would cause an overestimation of the expected group size because too few small groups are detected (that is, they are under-represented in the sample). Another complication is that large groups near the camera would be detected but it is possible that their centres would lie outside of the detection zone; for instance, all animals in the group might not be detected given the narrow width of the detection zone near the camera. According to Buckland et al. (2001), if the centre of the group is located inside the detection zone then the count of the size of the group must include all individuals in the group, even if some animals are beyond the detection zone; and if the centre of the group is outside the detection zone, then no observation is recorded. The difficulty with this in camera trapping is that even if groups detected have their centres within the detection zone the distance from the camera to the centre of the group is unobservable. Buckland et al. (2001) suggest that a possible approach would be to replace a group of

a given size by objects with the same size of the group at the same distance. In this case there would be no need to estimate mean group size. But the issue of violation of the independence assumption arises, invalidating analytical variance estimation and model selection procedures. However, if robust inference methods for variance estimation are adopted this difficulty can be avoided, but the issue of model selection remains. This approach would require exploration, and remains an avenue for future research in camera trapping analysis.

We have also shown that varying the time scale in aggregating the data from camera traps (that is over 1 day, 2 days or a week) has no effect on estimated density and its standard error. In this Chapter we have organized the encounters data in the form per camera per day to facilitate later modelling which would include covariates such as habitat and the random location of camera traps.

We have concluded that REM is relatively accurate but precision is dependent upon the method of variance estimation used. Therefore, we consider an integrated likelihood approach (iREM), which is comprehensive as it combines all of the data sets in one coherent framework, accounting for the variability in the observed data. This iREM approach is a more general and flexible approach in estimating density and its standard error. We could easily include covariates such as habitat or weather, and other independent variables required to estimate density into the model. iREM is discussed in Chapter 3. This approach corrects the precision, and allows for the accurate treatment of correlation in the estimators.

Chapter 3

integrated Random Encounter Model (iREM)

In this Chapter we develop an integrated likelihood method to estimate animal density. The integrated Random Encounter Model (iREM) builds on the Random Encounter Model (REM) discussed in Chapter 2. Rather than using a fixed estimate of animal speed as REM does, iREM simultaneously models the encounter data and animal speed data in one coherent framework. iREM utilizes an integrated population modelling (IPM) approach to estimate animal density, which is discussed in Section 3.1. The Chapter then continues with Section 3.2, which provides a description of the model. In Section 3.3 and Section 3.4 descriptions of the parametric distributions used to model animal speed of movement and the encounters, respectively, are given. Some species generally move around in pairs or family groups, so we also show how iREM can be extended to include group data in Section 3.2, and the model for the group data is given in Section 3.5. Examples of the likelihood function are given in Section 3.6. The performance of the models is tested via simulations in Section 3.7. To conclude the Chapter, Section 3.8 provides an illustration of the application of iREM using the Whipsnade Wild Animal Park (WWAP) data set.

3.1 Integrated Population Modelling (IPM)

Demographic and survey information at the population and individual levels are often simultaneously available when monitoring wildlife populations. The information collected is often analysed separately or in isolation, and separate results are presented.

For example, a survey may be designed to provide information on population abundance, while another survey may be designed to provide information about survival for a specific life stage. These surveys, however, may provide overlapping information about demographic processes or abundances; and analyses that utilize that overlap would be more powerful, and would provide more information, than multiple piecemeal analyses. One such approach which utilizes overlapping information is an integrated population model (IPM) (Newman et al., 2014, Chpt. 9; McCrea and Morgan, 2014, Chpt.12). Integrated population modelling provides a single, coherent analysis framework for a range of data sets collected from different surveys, all relating to the same species. By combining all sources of information in a single analysis, integrated population models simultaneously describe all the data, and consequently generally result in more precise parameter estimators (McCrea and Morgan, 2014, Chpt. 12). Early work on integrated population modelling was done by Fournier and Archibald (1982), who presented the idea for fishery data. Fournier and Archibald (1982) developed a flexible model to include extra information regarding the aging procedure of a fishery. Since then there have been several developments in that area. In fisheries research, the approach is termed integrated analysis (McCrea and Morgan, 2014, Chpt. 12).

Besbeas et al. (2002) developed an integrated analysis of different types of census and demographic data on animals of the same species. They devised a state-space model forming a combined likelihood for census data and data on survival from ring-recovery, under the assumption that the data sets are independent. The likelihood is formed by means of the Kalman filter, using appropriate normal variables to approximate Poisson and binomial random variables. By maximizing the combined likelihood the parameters estimated provided a simultaneous description of both data sets, and parameters such as productivity, which could not be estimated from the data sets separately were estimable under the combined likelihood framework. But to overcome a potential deficiency in combining likelihoods that are formed using specialist computer programs, which is an obstacle to the joint analysis, Besbeas et al. (2003) suggested a multivariate normal approximation, which was evaluated on data sets of two birds species, lapwings and herons, and which has been shown to be efficient and accurate. Extending this work Besbeas et al. (2005) adopted a multivariate normal approximation for the form

of the likelihood of the survival data, making use of parameter estimates and corresponding estimates of error obtained from analysing the survival data alone. In this case, the particular computer programs or packages need only be run once, to obtain maximum-likelihood estimates of the relevant parameters, and of their standard error and correlation. Schaub et al. (2007) argued that these integrated population models, however, have been applied to species without the lack of demographic data, so Schaub et al. (2007) have demonstrated the flexibility of integrated population models to estimate demographic parameters from sparse data, with a relictual colony of greater horseshoe bats (*Rhinolopbus ferrumequinum*). Schaub et al. (2007) applied a Bayesian integrated population modelling approach to the data and found that if the data were analysed separately, they would not have been able to estimate fecundity, the estimates of survival would have been less precise, and the estimate of population growth would have been biased.

As described above integrated population models have the advantage over the piecemeal approach of estimating parameters that are otherwise inestimable, and obtaining more precise parameter estimates. Cole and McCrea (2016) for example, demonstrated that aside from the natural advantages of improved precision of parameter estimates and reduced correlation, integrated population models have the additional advantage of making it possible to estimate some parameters that were not estimable from modelling the data individually.

There are some problems, however, with the use of integrated population models. For instance, as in the case of Besbeas et al. (2002), the specialist computer programs or packages in which the separate component likelihoods are constructed and combined would be an obstacle to the joint analysis. Also, integrated population models rely on the assumption that different data sets are independent, which is frequently violated in practice (Abadi et al., 2010). Besbeas et al. (2009) for example, showed that the danger of combining recovery information with dependent census data is increased root mean square errors compared with the case of combining with independent census data. But Abadi et al. (2010) used simulation methods to assess how the violation of the assumption of independence affects the statistical properties of the parameter estimators. They

found that this violation had only minor consequences on the precision and accuracy of the parameter estimates.

This thesis is particularly interested in estimating animal density and its variability correctly. The integrated Random Encounter Model (iREM) developed in this Chapter uses an integrated population modelling approach, which combines the speed data and the encounter data in a single framework. The iREM method is advantageous over piecemeal approaches as it accounts for the sampling variability of the estimator of animal speed of movement. iREM also allows for accurate treatment of precision and correlation in the estimators. The encounter data and animal speed data at WWAP were collected from separate sources, therefore, they are considered statistically independent, so their contributions to the likelihood could be multiplied. The next section describes the iREM method used to estimate animal density.

3.2 The Model

This Section describes the integrated REM for estimating animal density. iREM is a combined likelihood that simultaneously models the encounter data, y_{ij}, and the animal speed observations, $x = \{x_1, ..., x_m\}$, where $i = 1, 2, ..., c$ is the camera traps; $j = 1, 2, ..., n$ is the number of camera trap days. The assumptions of the iREM are

- encounters between animals and camera are independent
- animals move randomly and encounter camera traps independently of each other
- speeds are independent and identically distributed

In an integrated population framework it is assumed that the separate data sets are independent. In the case of the WWAP data discussed in Chapter 2, Section 2.7, the encounter data y_{ij} for $i = 1, 2, ..., c$ camera traps, and $j = 1, 2, ..., n$ camera trap days; and the speed observations $x = \{x_1, x_2, ..., x_m\}$ are independent as they were collected from separate sources (see Rowcliffe et al., 2008). In Section 2.1.1 we derived the encounter rate formula (as in Hutchinson and Waser, 2007) for animals with variable speeds, $x = \{x_1, x_2, ..., x_m\}$, showing that the fixed speed (constant speed) is replaced

by the expected speed, and that the encounter rate formula in REM can be expressed as

$$\lambda = \frac{(2+\theta)}{\pi} rt\mu_x D, \tag{3.2.1}$$

where μ_x is the population mean speed, which is an unknown quantity; t is the duration of time for which the camera was functioning; θ is the detection angle of the camera trap; r is the detection distance; and D is the density (see equation (2.2.5) in Section 2.2). As discussed in Section 2.2, Rowcliffe et al. (2008) used an appropriate estimator, $\bar{x}$ (which is treated as a fixed value) for μ_x. In this Chapter, however, we model the speed observations, $x = \{x_1, x_2, ..., x_m\}$ and the encounter observations y_{ij}, obtaining estimates of the density, D and expected speed, μ_x.

So suppose we have observed m independent speed observations, i.e., i.i.d, $x = \{x_1, x_2, ..., x_m\}$ from the same density $f(x_l \mid \mu_x,\ \nu)$, where $l = 1, 2, ..., m$, μ_x is the expected speed and ν represents any additional parameters in the model. And suppose the encounter observations have probability mass function $h(y_{ij} \mid \lambda,\ \tau)$ where λ is the encounter rate formula defined in equation 3.2.1 above, and τ represents any additional parameter in the model. Then the joint log-likelihood can be constructed as follows

$$\ell = \sum_{i=1}^{c} \sum_{j=1}^{n} \log\, h(y_{ij} \mid \lambda,\ \tau) + \sum_{l=1}^{m} \log\, f(x_l \mid \mu_x,\ \nu), \tag{3.2.2}$$

Possible models for the speed data, $f(\cdot)$, are given in Section 3.3. We also provide possible models for the encounter data, $h(\cdot)$, in Section 3.4. At WWAP some animal species were found to be moving in pairs or family groups. But as discussed in Chapter 2 (Sections 2.3.3, and 2.9) it can be problematic to estimate average group sizes. It is worth noting that WWAP study data was a special case, where the group sizes were independently observable using the cameras (see Section 2.7.1). Group sizes were small family units (on average 1.80) and it was possible to make counts of the number of individuals in the group. Rowcliffe et al. (2008) obtained an estimate of average group size by computing the mean size of the number of detected groups, a method also recommended by Buckland et al. (2001). In this section we assume study groups are small, given that the probability of detection depends on the size of the group, and the

distance. Note that small groups (and their centres) are more likely to be detected within the detection zone. We assume the group size data follows some discrete distribution whose support is the set of positive integers. The data for animals moving in groups is s_j where $j = 1, 2, ..., S$ is the groups. Suppose the group size data has probability mass function $k(s_j \mid \phi)$, where ϕ represents the parameters of the model. Then the joint likelihood becomes

$$\ell = \sum_{i=1}^{c} \sum_{j=1}^{n} \log\ h(y_{ij} \mid \lambda,\ \tau) + \sum_{l=1}^{m} \log\ f(x_l \mid \mu_x,\ \nu) + \sum_{j=1}^{S} \log\ k(s_j \mid \phi), \tag{3.2.3}$$

One possible model for the probability mass function of group data, $k(\cdot)$, is given in Section 3.5.

3.3 Models for animal speed of movement data

We propose four non-negative probability density functions to model animal speed data. These are as follows:

A gamma distribution with probability density function

$$f_{ga}(x \mid \alpha,\ \nu) = \frac{\nu^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\nu x} \qquad x,\ \nu,\ \alpha > 0, \tag{3.3.1}$$

with expected speed of movement, $\mu_x = \alpha/\nu$, and shape parameter, ν.

A lognormal model with probability density function

$$f_{Ln}(x \mid \mu_x,\ \nu) = \frac{1}{x\nu\sqrt{(2\pi)}} \exp\left(-\frac{[\log(x) - \epsilon]^2}{2\nu^2}\right) \qquad x > 0, \tag{3.3.2}$$

and with mean $\mu_x = \exp(\epsilon + \frac{1}{2}\nu^2)$, and ν is the scale parameter;

A Weibull distribution with probability density function

$$f_{weibull}(x \mid \nu,\ \beta) = \frac{\nu}{\beta}\left(\frac{x}{\beta}\right)^{\nu-1} e^{-\left(\frac{x}{\beta}\right)^{\nu}}, \tag{3.3.3}$$

and with mean $\mu_x = \beta\Gamma\left(1 + \frac{1}{\nu}\right)$, and ν is the shape parameter;

and finally, a zero-adjusted gamma distribution (ZAGA). Zero-adjusted distributions are suitable for data where there are animals that are observed to not move, such as in the WWAP data discussed in Chapter 2, Section 2.7. The probability density function of the zero-adjusted gamma is

$$f_{zaga}(x \mid \omega,\ \alpha,\ \nu) = \begin{cases} \omega, & \text{for } x = 0. \\ (1-\omega) \times \frac{\nu^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\nu x}, & \text{for } x > 0, \end{cases} \tag{3.3.4}$$

where ν is the shape parameter, and ω is the probability that an animal does not move. The mean of the ZAGA model is given as $\mu = (1 - w) \times \nu$. Whilst the possibility of forming a zero-adjusted distribution is not limited to a gamma distribution, as a demonstration we have only considered the gamma distribution. A zero-adjusted Weibull model or zero-adjusted lognormal model could also be used. Note that zero-adjusted distributions on zero and the positive real line are a special case of mixed distributions where the component at zero is a point mass. These distributions are appropriate when the response variable X takes values from zero to infinity including zero, that is, $[0, \infty)$.

3.4 Models for encounter data

The underlying assumption of REM used in Chapter 2 is that encounters between animals and camera traps follow a Poisson distribution. The Poisson distribution assumes that the mean and variance are the same but in most cases the variance would be greater than the mean, and therefore, the distribution would no longer be Poisson. This is known as overdispersion. For instance, the detection distances of animals generally vary over the environment since visibility depends on the vegetation. And, according to Hutchinson and Waser (2007) using a mean detection distance in the encounter-rate formula for two dimensions would still be valid but this would result in a higher variance in the encounter rate, and the distribution would no longer be Poisson. For example, in 2.1.1 we showed that it is possible to use a negative binomial distribution in REM to estimate the density. Cox (1983) also confirmed that the effects of overdispersion

include larger variance of summary statistics, and the possible loss of efficiency in using statistics appropriate for the single-parameter family. If there is overdispersion an appropriate model is a negative binomial (NB) distribution. The negative binomial model has one parameter more than the Poisson model that adjusts the variance independently from the mean.

Sometimes the data set may contain more zeros than the expected number from a Poisson distribution. For instance some animals may be less mobile than others or the placement strategies of the camera traps, such as in the WWAP survey and data set discussed in Chapter 2, Section 2.7 would result in records of more zeros than is expected for a Poisson model. In this case a zero-inflated Poisson (ZIP) distribution would be more appropriate to model such data. In other cases, the data may contain more zeros than usual and also show additional signs of overdispersion. In such situations a zero-inflated negative binomial (ZINB) distribution would be more appropriate given its flexibility in modelling both the zeros and the non-zero counts.

Suppose the encounter data, y_{ij}, has probability mass function $h(y_{ij} \mid \lambda,\ \tau)$. The four probability mass functions, $h(\cdot)$, proposed for the encounter data are given below. In each case the parameter λ is replaced by equation (3.2.1) to link to the parameters for density (D) and animal speed (μ_x).

Poisson REM

A Poisson model with mean λ, has probability mass function

$$h_{pois}(y_{ij} \mid \lambda) = \frac{e^{-\lambda}}{y_{ij}!}\lambda^{y_{ij}}; \tag{3.4.1}$$

with variance, $\text{Var}(y_{ij}) = \lambda$.

Negative binomial REM (NB REM)

For the negative binomial model case, we take the NB-2 form (Ismail and Zamani, 2013) with mean λ, and probability mass function

$$h(y_{ij} \mid \lambda,\ \kappa) = \frac{\Gamma\left(\kappa + y_{ij}\right)}{\Gamma(y_{ij}+1)\Gamma\left(\kappa\right)}\left(\frac{\kappa}{\lambda+\kappa}\right)^{\kappa}\left(\frac{\lambda}{\lambda+\kappa}\right)^{y_{ij}}, \tag{3.4.2}$$

where $1/\kappa$ denotes the dispersion parameter. The variance $\text{Var}(y_{ij}) = \lambda + \frac{1}{\kappa}\lambda^2$. The NB REM model reduces to a Poisson REM in the limit as $1/\kappa \to 0$, and displays overdispersion when $1/\kappa \to \infty$.

Zero-inflated Poisson REM (ZIP REM)

A zero-inflated Poisson (ZIP) model with mean $\lambda(1-\rho)$, has probability mass function

$$f_{zip}(y_{ij} \mid \rho,\ \lambda) = \begin{cases} \rho + (1-\rho)e^{-\lambda}, & \text{for } y_{ij} = 0, \\ \\ \dfrac{(1-\rho)e^{-\lambda}\lambda^{y_{ij}}}{y_{ij}!}, & \text{for } y_{ij} > 0, \end{cases} \tag{3.4.3}$$

where $0 \leq \rho \leq 1$ is the probability which inflates the zero response category. The ZIP REM has variance, $\text{Var}(y_{ij}) = (1-\rho)(\lambda + \rho\lambda^2)$. The ZIP REM reduces to a Poisson REM when $\rho = 0$ and displays overdispersion when $\rho > 0$.

Zero-inflated negative binomial REM (ZINB REM)

A zero-inflated negative binomial REM (ZINB REM) with mean $\lambda(1-\rho)$, has probability mass function

$$f(y_{ij} \mid \kappa,\ \lambda,\ \rho) = \begin{cases} \rho + (1-\rho)\ \left(\frac{\kappa}{\lambda+\kappa}\right)^{\kappa}, & \text{for } y_{ij} = 0, \\ \\ (1-\rho)\dfrac{\Gamma\left(\kappa + y_{ij}\right)}{\Gamma(y_{ij}+1)\Gamma\left(\kappa\right)}\ \left(\dfrac{\kappa}{\lambda+\kappa}\right)^{\kappa}\left(\dfrac{\lambda}{\lambda+\kappa}\right)^{y_{ij}}, & \text{for } y_{ij} > 0, \end{cases} \tag{3.4.4}$$

where $0 \leq \rho \leq 1$ is the probability which inflates the zero response category; $1/\kappa > 0$ is the dispersion parameter. The ZINB REM has variance, $\text{Var}(y_{ij}) = (1-\rho)\lambda + (1-\rho)(\rho + 1/\kappa)\lambda^2$. When $\rho = 0$ the ZINB REM reduces to a NB REM, and reduces to a ZIP REM when $\kappa \to \infty$.

3.5 Model for group size data

For the data to be considered group data, there must be at least one observed individual, as in the case of the WWAP data set, discussed in Chapter 2, Section 2.7. Here we consider a zero-truncated Poisson model (ZTP) to model the group size data whose support is the set of positive integers. A zero-truncated Poisson (ZTP) distribution which has mean, $g = \dfrac{\phi}{(1 - e^{-\phi})}$, is defined as

$$f_{ztp}(s_j \mid \phi) = \frac{\phi^{s_j} e^{-\phi}}{s_j!(1 - e^{\phi})}, \qquad s_j > 0, \tag{3.5.1}$$

where ϕ is the mean before truncation. Whilst the possibility of forming a zero-truncated distribution is not limited to a Poisson distribution, as a demonstration, we have used a zero-truncated Poisson distribution to model group size data. A zero-truncated negative binomial distribution could also be used to model group size but this is not considered in thesis.

3.6 Examples of log-likelihood functions

Below we give an example of the R-codes for the iREM log-likelihood function without group data and with the group data.

Example 1

In this example we give the log-likelihood function for a poisson iREM for the encounter data and a gamma model for the speed data. The parameters to estimate density are held fixed: $r = 0.012$ (km), $\theta = 0.175$ (radians), and $t = 1$ (day). The function estimates the average speed and shape parameter from a gamma model and the density. The scale parameter is estimated using the estimated average speed and estimated shape parameter. Using the fixed parameters required for density estimation, estimated average speed parameter and estimated density parameter, the encounter rate is computed. The function "gamma.fits" can be optimized using an optimizer such as optim, which minimizes the negative joint log-likelihood.

```
#loglikelihood function to estimate density. Fitting a gamma model
#to animal speed data, x, and a Poisson model to encounter data,y.
 theta = 0.175     #fixed detection angle
 r     = 0.012     #fixed detection distance
 t     = 1         #camera trap time period

#gamma.fits returns the negative loglikelihood
gamma.fits <- function(param,x,y){
 mean.speed = exp(param[1])
 nu         = exp(param[2])
 Density    = exp(param[3])
 alpha      = mean.speed*nu
 lambda     = ((2+theta)/pi)*r*t*mean.speed*Density
-sum(dpois(y,lambda,log=TRUE))-sum(dgamma(x,alpha,nu,log=TRUE))}
```

Example 2

In this example we add the log-likelihood function for group data to the log-likelihood functions of the encounter data and the speed data given in example 1 above. The function "gamma.group" returns the negative joint log-likelihood function. Using the function optim in R, we can minimize the negative log-likelihood, estimating average speed, the shape parameter, ν, from a gamma model, the mean from the Poisson model before truncation, ϕ, which is used to estimate the average group size, g, and the density.

```
#loglikelihood function to estimate density for animals
#moving in groups. Fitting a ZTP to group data, s, a gamma model
#to animal speed data, x, and a Poisson model to encounter data,y.
gamma.group <- function(param,x,y, s){
 mu_x = exp(param[1])
 nu          = exp(param[2])
 phi         = exp(params[3])
 g           = phi/(1-exp(-phi))
 Density     = exp(param[4])/g
 alpha       = mu_x*nu; lambda = ((2+theta)/pi)*r*t*mu_x*Density
 -sum(dpois(y, lambda,log=TRUE))-sum(dgamma(x, alpha, nu, log=TRUE))-
 (s*phi - sum(s)*log(phi)+(s*log(1-exp(-phi))))}
```

3.7 Simulation Study

This simulation study explores the performance of the models, looking at the six following cases:

(i) Firstly, we verify that iREM can be used in realistic settings by fitting a Poisson iREM and a Poisson REM to encounter data drawn from a Poisson REM (Section 3.7.1).

(ii) Secondly, we explore how important it is to account for overdispersion in encounter data. To do this we fit a Poisson iREM and a Poisson REM with encounters following a NB REM. We also fit a NB iREM and a NB REM to the same encounters (Section 3.7.2).

(iii) Thirdly, we explore how important it is to account for zero-inflation in encounter data. To do this we fit a Poisson iREM and a Poisson REM with encounters following a ZIP REM. We also fit a ZIP iREM and a ZIP REM to the same encounters (Section 3.7.3).

(iv) Fourthly, we explore how important it is to account for both zero-inflation and extra variation in encounter data. To do this we fit a Poisson iREM and a Poisson REM to encounters simulated from a ZINB REM. We also fit a NB iREM and a NB

REM to the same encounter data. We compare these models with a ZINB iREM and a ZINB REM. These results are shown in Section (3.7.4), and in appendix B.1.3.

(v) Fifthly, we verify that iREM can be used in realistic settings to estimate group size by fitting a Poisson iREM to encounters simulated from a Poisson REM and we fit a ZTP to data simulated from the relevant model (Section 3.7.5).

In these five cases we fit a gamma model, a lognormal model or a Weibull model to animal speed data simulated from the relevant models.

(vi) Sixthly, we verify that iREM can be used in realistic settings where animals are observed to not move by fitting a Poisson iREM to encounters simulated from a Poisson REM. As discussed in Section 3.3, some animals are generally less mobile than others, and therefore, a ZAGA model would be appropriate for data where there are animals that are observed to not move, as in the case of the WWAP data discussed in Chapter 2, Section 2.7 (Section 3.7.6).

For the simulations we generate scenarios in which the true density and true animal speed are plausible ecologically for our motivating WWAP data set. The joint likelihoods were maximized using the optim function in the R software package (version 3.2.4) with the default Nelder-Mead algorithm. For each simulation case, the average parameter estimates (standard error in parentheses), Standard deviation (Sd) and Root Mean Square Error are computed from 200 simulation runs. Under all scenarios, estimates from iREM are compared with the maximum likelihood REM framework proposed in Chapter 2. We use the formula in equation (2.5.2) proposed in Chapter 2 Section 2.5.1 to computer approximate 95% confidence intervals. That is,

$$(\hat{D}_L, \hat{D}_U) = e^{\left(\ln(\hat{D}) \pm 1.96 \times \mathrm{SE}\left\{\ln(\hat{D})\right\}\right)}. \tag{3.7.1}$$

where the variance of estimated density is approximated from the inverse Hessian matrix in optim in R and, therefore the standard error of the logarithm of density $\left(\mathrm{SE}\left\{\ln(\hat{D})\right\}\right)$ can be computed.

3.7.1 Testing the performance of the Poisson iREM

The true values and sample sizes used in the simulation process are given in Table 3.7.1 .

Table 3.7.1: Sample sizes and true values used to simulate data.

	Sample sizes		**True values**	
	trap days (n)	speed (m)	D	μ_x (km/day^{-1})
1.	40	40	20	0.60
2.	100	100	20	0.60
3.	40	40	100	0.60
4.	100	100	100	0.60
5.	40	40	20	4.60
6.	100	100	20	4.60
7.	40	40	100	4.60
8.	100	100	100	4.60

Table 3.7.2 illustrates the performance of the Poisson iREM under different parameter settings and sample size conditions. The simulation results show that the Poisson iREM works under different parameter settings and with small and large sample sizes. When the sample size is small there can be a small positive bias but, as expected, with larger sample sizes this bias is minimal and estimation precision improves. REM also provides similar estimates of the density compared with iREM, however, the estimated standard errors are smaller in REM than in iREM. The variability in animal speed is not considered in REM, therefore, as expected, the standard error is underestimated. To evaluate the performance of iREM the RMSE is calculated, and the results show that the RMSEs are broadly similar but in most cases the RMSEs from iREM are lower than REM. For large expected animal speed, the density is accurately estimated with lower estimates of the standard error compared with a lower value of expected animal speed. The simulations also show that there is no real difference between the three speed data models. In Table B.1.1, in appendix B.1, we give the results of estimated density, expected animal speed and the additional parameters from the speed data models. The simulation results show that all parameters are estimated well, for example all the true values are captured within an approximate 95% confidence interval.

Table 3.7.2: Average density estimates from iREM compared with REM for $\mu_x = 0.60$ (km/day^{-1}) or $\mu_x = 4.60$ (km/day^{-1}) (average standard errors in parentheses). The Standard deviation (Sd) and Root Mean Square Error (RMSE) are given.

	gamma		lognormal		Weibull	
	iREM $\hat{D}$	REM $\hat{D}$	iREM $\hat{D}$	REM $\hat{D}$	iREM $\hat{D}$	REM $\hat{D}$
$\mu_x = 0.60$						
$D = 20$; $n = m = 40$						
estimate	23 (13.66)	24 (12.58)	23 (13.27)	23 (12.31)	23 (14.52)	24 (12.94)
Sd	9.44	14.00	11.85	12.11	11.91	15.82
RMSE	9.79	14.53	12.18	12.55	12.20	16.41
$D = 20$; $n = m = 100$						
estimate	20 (7.94)	20 (6.68)	20 (7.37)	20 (6.72)	21 (8.36)	20 (6.86)
Sd	6.07	7.75	7.15	7.36	7.34	8.46
RMSE	6.09	7.75	7.16	7.36	7.36	8.51
$D = 100$; $n = m = 40$						
estimate	106 (42.28)	106 (24.32)	105 (36.56)	106 (25.58)	109 (47.97)	112 (25.80)
Sd	45.54	45.18	37.30	38.66	50.15	52.16
RMSE	45.92	45.53	37.55	39.17	50.87	53.34
$D = 100$; $n = m = 100$						
estimate	102 (25.55)	102 (14.62)	100 (21.68)	102 (14.66)	102 (28.26)	104 (14.90)
Sd	26.41	26.41	21.17	22.47	26.94	29.07
RMSE	26.47	26.46	21.17	22.57	27.01	29.29
$\mu_x = 4.60$						
$D = 20$; $n = m = 40$						
estimate	20 (3.84)	20 (3.70)	22 (8.25)	23 (4.11)	20 (3.82)	20 (3.68)
Sd	3.75	3.76	8.13	8.85	3.73	3.74
RMSE	3.76	3.77	8.38	9.21	3.74	3.74
$D = 20$; $n = m = 100$						
estimate	20 (2.38)	20 (2.30)	20 (4.90)	21 (2.41)	20 (2.38)	20 (2.29)
Sd	2.33	2.33	4.90	5.73	2.44	2.44
RMSE	2.33	2.33	4.91	5.79	2.44	2.44
$D = 100$; $n = m = 40$						
estimate	100 (8.23)	100 (8.10)	101 (10.68)	101 (8.15)	100 (8.23)	100 (8.11)
Sd	7.95	7.96	10.28	10.28	7.79	7.79
RMSE	7.95	7.96	10.29	10.30	7.79	7.79
$D = 100$; $n = m = 100$						
estimate	100 (5.21)	100 (5.13)	100 (6.79)	101 (5.14)	100 (5.21)	100 (5.13)
Sd	5.33	5.31	6.88	6.85	5.23	5.23
RMSE	5.35	5.32	6.90	6.87	5.24	5.24

3.7.2 *Investigating the importance of accounting for overdispersion in encounter data*

The performance of the Poisson iREM is also investigated when there is extra variation in the data. Encounter data are simulated from a NB REM and we fit a Poisson iREM and a NB iREM to the encounters, for $n = 40$ or $n = 100$ camera trapping days and $m = 40$ or $m = 100$ animal speed observations, where for illustration the parameter values used were based upon reasonable values that might be applicable for real species. We set $\kappa = 0.25$, to illustrate large variance in the encounters, and for expected animal speed, we set $\mu_x = 2.60$ (km/day^{-1}) or $\mu_x = 8.60$ (km/day^{-1}). The encounter rate is a function of expected speed, therefore, the smaller the value of μ, the smaller the encounter rate, particularly, for small density values. Therefore, appropriate values of μ_x were chosen to generate encounters plausible from a NB REM. The density was set to $D = 20$ or $D = 100$ to represent small and large densities of animals.

Table 3.7.3 gives the estimated density from fitting a Poisson iREM and a NB iREM to encounter data. Estimated densities from iREM are also compared with estimates from REM proposed in Chapter 2. The simulation results reveal that induced-variation in the Poisson iREM introduced a negative bias for large density values and an underestimation of the standard error of the density that is at least 2.5 times lower than the estimated standard error from the NB iREM. For small density values, however, the Poisson iREM gave similar estimates of the density to the NB iREM but the standard errors are underestimated, however, this underestimation is not as severe (at least 1.5 times lower) as in the case of large density values. In all scenarios, REM gave similar estimates of the density from iREM but REM gave smaller estimates of the standard error of the density. As expected the NB iREM gave smaller RMSEs, and, as sample sizes increase, the RMSE reduces.

We also investigated the performance of the Poisson iREM for a larger expected speed of movement. The simulation results are given in Table 3.7.4, which show smaller estimated standard error of the density compared with the results from a smaller expected speed, μ_x in Table 3.7.3. As in the case of the lower expected speed ($\mu_x = 2.60$ km/day^{-1}),

under large density conditions the estimated standard error of the density from fitting a Poisson iREM to encounter data is at least 2.5 times lower than the estimated standard error from fitting a NB iREM to the same encounter data, and at least 2 times lower under small density values. As expected, REM gave relatively accurate estimates of the density but obtained smaller estimates of the standard error, and in all cases, the RMSEs from the NB iREM is lower that those from the Poisson iREM or REM. Also, increasing sample size improves estimation and its precision.

In appendix B.1.1 we give the estimates of the density, the dispersion parameter, expected animal speed and the additional parameters from the speed data models. Table B.1.2 gives the parameter estimates from a Poisson iREM, while Tables B.1.3 and B.1.4 give the parameter estimates from a NB iREM. The simulation results show that all parameters are estimated with reasonable accuracy, for example all the true values are captured within an approximate 95% confidence interval.

Finally, under all parameter settings, the simulation results reveal that there is no real difference between the three speed data models.

Table 3.7.3: Average density estimates from fitting a Poisson iREM and a NB iREM to encounters simulated from a NB REM. The estimates from the iREMs are compared with the estimates from REM proposed in Chapter 2 (average standard errors in parentheses). The parameters are set to $\mu_x = 2.60$ (km/day^{-1}); and $\kappa = 0.25$. The Standard deviation (Sd) and Root Mean Square Error (RMSE) are given.

	gamma model				**lognormal model**				**Weibull model**			
	Poisson		NB		Poisson		NB		Poisson		NB	
	iREM $\hat{D}$	REM $\hat{D}$	iREM $\hat{D}$	REM $\hat{D}$	iREM $\hat{D}$	REM $\hat{D}$	iREM $\hat{D}$	REM $\hat{D}$	iREM $\hat{D}$	REM $\hat{D}$	iREM $\hat{D}$	REM $\hat{D}$
$D = 20$; $n = m = 40$												
estimate	21 (5.35)	20 (5.26)	21 (8.51)	21 (8.48)	21 (5.29)	21 (5.27)	21 (8.57)	21 (8.48)	21 (5.42)	20 (5.30)	21 (8.65)	21 (8.53)
Sd	8.31	8.61	8.05	8.61	8.58	8.72	8.39	8.72	8.79	8.83	8.55	8.83
RMSE	8.35	8.65	8.10	8.65	8.64	8.76	8.46	8.76	8.86	8.88	8.64	8.88
$D = 20$; $n = m = 100$												
estimate	20 (3.21)	20 (3.14)	21 (5.25)	20 (5.14)	21 (3.76)	21 (3.19)	21 (5.66)	21 (5.23)	20 (3.20)	20 (3.14)	21 (5.22)	20 (5.14)
Sd	5.23	5.34	5.13	5.34	5.75	5.89	5.70	5.89	5.25	5.31	5.20	5.31
RMSE	5.24	5.35	5.16	5.35	5.79	5.93	5.77	5.93	5.26	5.32	5.22	5.32
$D = 100$; $n = m = 40$												
estimate	98 (12.09)	97 (11.06)	99 (33.12)	97 (32.76)	98 (18.73)	98 (11.15)	99 (36.36)	98 (33.04)	96 (12.03)	97 (11.00)	99 (32.88)	97 (30.43)
Sd	31.29	31.75	31.28	31.75	35.53	35.75	34.86	35.75	30.34	30.42	30.61	30.43
RMSE	31.39	31.85	31.29	31.86	35.57	35.79	34.87	35.79	30.65	30.59	30.64	30.59
$D = 100$; $n = m = 100$												
estimate	98 (7.57)	98 (6.87)	100 (21.54)	98 (20.79)	98 (11.74)	99 (6.90)	100 (23.45)	99 (20.90)	99 (7.61)	99 (6.91)	101 (21.76)	99 (20.93)
Sd	22.20	22.39	22.26	22.39	24.07	24.71	23.81	24.72	22.72	22.53	22.17	22.54
RMSE	22.27	22.47	22.26	22.47	24.16	24.76	23.81	24.76	22.75	22.57	22.21	23.76

Table 3.7.4: Average density estimates from fitting a Poisson iREM and a NB iREM to encounters simulated from a NB REM. The estimates from the iREMs are compared with the estimates from Rowcliffe et al. (2008) REM (average standard errors in parentheses). The parameters are set to $\mu_x = 8.60$ (km/day^{-1}); and $\kappa = 0.25$. The Standard deviation (Sd) and Root Mean Square Error (RMSE) are given.

	gamma model				**lognormal model**				**Weibull model**			
	Poisson		NB		Poisson		NB		Poisson		NB	
	iREM $\hat{D}$	REM $\hat{D}$	iREM $\hat{D}$	REM $\hat{D}$	iREM $\hat{D}$	REM $\hat{D}$	iREM $\hat{D}$	REM $\hat{D}$	iREM $\hat{D}$	REM $\hat{D}$	iREM $\hat{D}$	REM $\hat{D}$
$D = 20$; $n = m = 40$												
estimate	20 (3.16)	20 (2.77)	21 (7.30)	20 (6.97)	20 (3.59)	20 (2.78)	21 (7.55)	20 (6.97)	20 (3.15)	20 (2.77)	21 (7.35)	20 (6.98)
Sd	6.69	6.71	6.46	6.71	6.57	6.57	6.52	6.57	6.66	6.79	6.30	6.79
RMSE	6.70	6.71	6.51	6.71	6.57	6.57	6.58	6.57	6.66	6.79	6.33	6.79
$D = 20$; $n = m = 100$												
estimate	20 (1.95)	20 (1.70)	20 (4.52)	20 (4.33)	20 (1.91)	20 (1.70)	20 (4.47)	20 (4.34)	20 (1.94)	20 (1.69)	20 (4.48)	20 (4.31)
Sd	4.32	4.36	4.17	4.36	4.29	4.39	4.15	4.38	4.33	4.37	4.22	4.37
RMSE	4.32	4.36	4.19	4.36	4.29	4.39	4.17	4.39	4.34	4.37	4.23	4.37
$D = 100$; $n = m = 40$												
estimate	98 (9.62)	98 (6.09)	100 (33.36)	98 (31.59)	98 (12.79)	99 (6.15)	102 (35.13)	99 (31.88)	99 (9.54)	98 (6.12)	101 (35.51)	98 (31.72)
Sd	33.92	34.68	33.65	34.68	37.57	37.41	36.00	37.41	34.24	33.64	32.56	33.64
RMSE	33.99	34.74	33.65	34.74	37.60	37.42	36.03	37.42	34.29	33.68	32.56	33.67
$D = 100$; $n = m = 100$												
estimate	98 (6.02)	98 (3.74)	100 (20.99)	98 (19.90)	98 (5.67)	98 (3.76)	100 (20.78)	98 (19.99)	97 (5.93)	98 (3.75)	99 (20.85)	98 (18.80)
Sd	18.37	18.40	18.28	18.54	19.15	18.80	18.52	18.71	19.33	18.80	18.52	18.80
RMSE	18.47	18.54	18.39	18.54	19.24	18.80	18.52	18.80	19.53	18.91	18.53	18.91

3.7.3 *Investigating the importance of accounting for zero-inflation in encounter data*

To be able to test the performance of the Poisson iREM in the case of zero-inflation, we fit a Poisson iREM to encounter data simulated from a ZIP REM. For the probability of the zero-response category, we set $\rho = 0.30$, and for the expected animal speed, we set $\mu_x = 4.60$ (km/day^{-1}). The survey effort is $n = 40$ or $n = 100$ camera trapping days, and $m = 40$ or $m = 100$ animal speed observations.

Table 3.7.5 illustrates the performance of the Poisson iREM when there is zero-inflation in the encounter data. The simulation results (Table 3.7.5) show that induced-zero-inflation introduced a strong negative bias in estimated density from the Poisson iREM, and an underestimation of the standard error of the density under all scenarios. The bias introduced in REM is similar to iREM but the estimated standard errors of the density are substantially smaller. The simulation results also reveal that there is no real difference between the three speed data models.

Table B.1.5 (in appendix B.1.2) gives the simulation results of the estimated parameters from fitting a Poisson iREM to encounters simulated from a ZIP REM. These include average estimates of the density, expected animal speed and the additional parameters from the three speed data models. While Table B.1.6 (in appendix B.1.2) gives the estimated parameters from a ZIP iREM, which includes the density, the probability of the zero-response category, ρ, expected animal speed, μ_x, and the additional parameters from the three speed data models. The results show that all parameters, except in the case of the density estimated from the Poisson iREM, are estimated well and are captured within an approximate 95% confidence interval.

Table 3.7.5: Average density estimates from fitting a Poisson iREM and a ZIP iREM to encounters simulated from a ZIP REM. The average estimates are also compared with average estimates from Rowcliffe et al. (2008) REM (average standard errors in parentheses). The parameters are set to $\mu_x = 4.60$ (km/day^{-1}), and $\rho = 0.30$. The Standard deviation (Sd) and Root Mean Square Error (RMSE) are given.

	gamma model				**lognormal model**				**Weibull model**			
	Poisson		ZIP		Poisson		ZIP		Poisson		ZIP	
	iREM	REM	iREM	REM	iREM	REM	iREM	REM	iREM	REM	iREM	REM
	$\hat{D}$	$\hat{D}$	$\hat{D}$	$\hat{D}$	$\hat{D}$	$\hat{D}$	$\hat{D}$	$\hat{D}$	$\hat{D}$	$\hat{D}$	$\hat{D}$	$\hat{D}$
$D = 20$; $n = m = 40$												
estimate	14 (3.41)	14 (3.09)	20 (7.07)	20 (4.97)	14 (3.26)	14 (3.07)	20 (6.78)	20 (4.76)	14 (3.39)	14 (3.08)	20 (7.06)	20 (4.78)
Sd	3.35	3.35	6.64	4.75	3.15	3.18	6.53	4.51	3.50	3.50	6.98	4.98
RMSE	6.92	6.92	6.65	4.76	6.92	6.92	6.54	4.52	7.02	7.02	6.98	4.99
$D = 20$; $n = m = 100$												
estimate	14 (2.13)	14 (1.93)	20 (4.80)	20 (3.01)	14 (1.97)	14 (1.93)	20 (4.56)	20 (3.02)	14 (2.14)	14 (1.94)	20 (4.83)	20 (3.04)
Sd	2.06	2.06	4.50	2.95	1.90	1.90	4.33	2.72	2.05	2.04	4.39	2.90
RMSE	6.35	6.35	4.50	2.95	6.29	6.30	4.33	2.72	6.22	6.22	4.39	2.90
$D = 100$; $n = m = 40$												
estimate	70 (9.89)	70 (6.85)	100 (14.56)	100 (10.31)	69 (8.72)	70 (6.81)	100 (12.91)	100 (10.24)	70 (9.76)	70 (6.83)	100 (14.36)	100 (10.28)
Sd	11.78	11.78	13.73	13.74	10.76	10.85	12.48	12.61	12.29	12.27	14.73	14.70
RMSE	32.30	32.30	13.74	13.75	32.36	32.36	12.48	12.61	32.65	32.67	14.74	14.70
$D = 100$; $n = m = 100$												
estimate	70 (6.25)	70 (4.30)	99 (9.05)	100 (6.35)	70 (4.80)	70 (4.30)	99 (7.06)	100 (6.35)	70 (6.27)	70 (4.34)	100 (9.09)	100 (6.41)
Sd	7.20	7.20	8.68	8.62	5.95	5.95	6.88	6.78	7.19	7.17	8.55	8.49
RMSE	30.64	30.64	8.70	8.64	30.36	30.37	6.90	6.80	30.01	30.01	8.56	8.50

3.7.4 *Investigating the importance of accounting for zero-inflation and overdispersion in the encounter data*

We also test the performance of the Poisson iREM and the NB iREM by fitting these models to data simulated from a ZINB REM. Simulations are conducted for survey efforts of $n = 40$ or $n = 100$ camera trapping days, and $m = 40$ or $m = 100$ animal speed observations. We set $\kappa = 0.10$, and the probability of the zero-response category $\rho = 0.10$. For the density, and expected speed we set $D = 20$ or $D = 100$, and $\mu_x = 6.60$ (km/day^{-1}), respectively.

The simulation results are given in Table B.1.7 and Table B.1.8, in appendix B.1.3. The simulation results reveal that induced-variation and induced-zero-inflation in a Poisson iREM introduced a slight negative bias and an underestimation of the standard error that is at least 3 times smaller than that from the ZINB iREM model for small density values. But for larger density values, the bias is much larger and the underestimation of the standard error is even worse, the estimated standard error of the density from a Poisson iREM is at least 6 times lower than that from the ZINB iREM. REM gave similar estimates of the density as the iREMs but it underestimates the standard error of the density.

The NB iREM and ZINB iREM gave similar estimates of the density and standard error for small density values, but for larger density values a negative bias is introduced from the NB iREM and the standard error of the density is underestimated.

Based on these results it can be argued that the use of the underlying Poisson model in REM should be taken with care as larger variance and zero-inflation can induce estimation bias and underestimation of variance for scenarios that can be considered ecologically realistic. In particular, when zero-inflation is allowed in the Poisson iREM, the true density is not captured within an approximate 95% confidence interval. If the variance of the encounters is large the Poisson iREM would give stable estimates of the density but the standard error would be underestimated. Therefore, it is advisable to take this into account when deciding on a suitable distribution to model counts observed

on camera traps, rather than assuming that the underlying Poisson model in REM is generally the best model for encounter records. A similar conclusion can be drawn for the use of REM, which gave stable estimates of the density but gave smaller estimates of the standard errors compared with iREM. It is worth iterating that REM does not take into account the sampling variability of speed, therefore, we expect an underestimation of the standard error in REM.

3.7.5 Testing the performance of the Poisson iREM where animals move in groups

The simulation results for animals moving in groups are given in this section. Group size data were simulated from a zero-truncated Poisson (ZTP), and as a demonstration we fit a Poisson REM to encounter data, and we fit a gamma model to the animal speed data. For group data, $s_j (j = 1, 2, ..., n)$, the mean from a ZTP model is given as

$$\mathbb{E}(s_j) = g = \frac{\phi}{1 - \exp(-\phi)},$$

where ϕ is the mean from the Poisson model before truncation. In the simulations, estimations are obtained for the mean, ϕ, of the Poisson model before truncation and this is used to compute average group size, g. For the mean before truncation parameter, we set $\phi = 2.65$, hence, average group size, $g = 2.85$. For the density, we set $D = 100$ or $D = 20$, and the sampling design has $n = 40$ or $n = 100$ camera trapping days, $m = 40$ or $m = 100$ animal speed observations, and group data sample size $S = 10$ or $S = 50$. The parameters used to estimate density are $r = 0.012$ (km), $\theta = 0.175$, $t = 1$ (day), and expected speed, $\mu_x = 0.70$ (km/day^{-1}).

Table 3.7.6 gives the simulation results for a small sample of $S = 10$ group sizes. The simulations show that the model works well under different scenarios for small sample sizes of groups. Increasing the sample size improves estimation and precision, as expected. When the sample sizes of groups is large ($S = 50$) estimated standard errors are smaller compared with smaller sample sizes of groups ($S = 10$). These results are given in Table 3.7.7, with the estimated standard error (in parentheses), Standard deviation (Sd) and Root Mean Square Errors (RMSE).

Table 3.7.6: Average parameter estimates for animals moving in pairs or family groups (average standard errors in parentheses). The following are the true values $\mu_x = 0.70$ (km/day^{-1}), $g = 2.85$, and sample size of group size data $S = 10$. The Standard deviation (Sd) and Root Mean Square Error (RMSE) are also given.

	gamma			lgnormal			Weibull		
	$\hat{D}$	$\hat{\mu}_x$	$\hat{g}$	$\hat{D}$	$\hat{\mu}_x$	$\hat{g}$	$\hat{D}$	$\hat{\mu}_x$	$\hat{g}$
$D = 100,\ n = m = 40$									
estimate	104 (38.17)	0.73 (0.18)	2.90 (0.76)	103 (36.03)	0.72 (0.16)	2.90 (0.69)	102 (38.75)	0.73 (0.19)	2.90 (0.74)
Sd	39.61	0.19	0.33	32.21	0.16	0.33	32.02	0.19	0.33
RMSE	39.82	0.19	0.34	32.31	0.16	0.34	32.09	0.19	0.34
$D = 100,\ n = m = 100$									
Estimate	103 (27.23)	0.71 (0.11)	2.88 (0.48)	101 (25.89)	0.71 (0.10)	2.88 (0.37)	102 (27.88)	0.71 (0.12)	2.88 (0.50)
Sd	22.91	0.11	0.37	18.39	0.09	0.37	21.55	0.12	0.37
RMSE	23.07	0.11	0.38	18.40	0.09	0.38	21.68	0.12	0.38
$D = 20,\ n = m = 40$									
estimate	21 (12.46)	0.73 (0.18)	2.90 (0.76)	21 (12.14)	0.72 (0.16)	2.90 (0.69)	21 (12.49)	0.73 (0.19)	2.90 (0.74)
Sd	11.34	0.19	0.33	10.37	0.16	0.33	10.53	0.19	0.33
RMSE	11.40	0.19	0.34	10.42	0.16	0.34	10.58	0.19	0.34
$D = 20\ n = m = 100$									
Estimate	21 (7.97)	0.71 (0.11)	2.88 (0.48)	20 (7.65)	0.71 (0.10)	2.88 (0.37)	21 (8.04)	0.71 (0.12)	2.88 (0.38)
Sd	7.44	0.11	0.37	6.19	0.09	0.37	7.23	0.11	0.37
RMSE	7.50	0.11	0.37	6.21	0.09	0.37	7.29	0.12	0.37

Table 3.7.7: Average parameter estimates for animals moving in pairs or family groups (average standard errors in parentheses). The following are the true values $\mu_x = 0.70$ (km/day^{-1}), $g = 2.85$, and sample size for group size $S = 50$. The Standard deviation (Sd) and Root Mean Square Error (RMSE) are also given.

	gamma			lgnormal			Weibull		
	$\hat{D}$	$\hat{\mu}_x$	$\hat{g}$	$\hat{D}$	$\hat{\mu}_x$	$\hat{g}$	$\hat{D}$	$\hat{\mu}_x$	$\hat{g}$
$D = 100,\ n = m = 40$									
estimate	103 (35.49)	0.73 (0.19)	2.90 (0.77)	103 (32.57)	0.72 (0.16)	2.91 (0.70)	104 (36.62)	0.73 (0.20)	2.90 (0.75)
Sd	37.58	0.21	0.16	30.06	0.16	0.15	34.49	0.20	0.15
RMSE	37.71	0.21	0.18	30.20	0.16	0.16	34.67	0.19	0.16
$D = 100,\ n = m = 100$									
Estimate	103 (22.60)	0.71 (0.11)	2.91 (0.48)	101 (21.03)	0.71 (0.10)	2.91 (0.53)	101 (23.06)	0.71 (0.12)	2.91 (0.42)
Sd	22.25	0.12	0.15	17.88	0.10	0.18	21.71	0.13	0.15
RMSE	22.49	0.12	0.16	17.89	0.10	0.16	21.73	0.13	0.16
$D = 20,\ n = m = 40$									
estimate	21 (12.02)	0.70 (0.18)	2.91 (0.78)	21 (11.40)	0.71 (0.16)	2.91 (0.73)	21 (12.11)	0.72 (0.19)	2.91 (0.77)
Sd	11.40	0.19	0.15	11.03	0.16	0.15	11.64	0.20	0.15
RMSE	11.43	0.19	0.16	11.05	0.16	0.16	11.71	0.20	0.16
$D = 20\ n = m = 100$									
Estimate	20 (7.08)	0.71 (0.11)	2.91 (0.48)	20 (6.73)	0.71 (0.10)	2.91 (0.53)	19 (6.99)	0.72 (0.12)	2.91 (0.42)
Sd	6.88	0.12	0.15	5.84	0.10	0.15	6.22	0.13	0.15
RMSE	6.88	0.12	0.16	5.88	0.10	0.16	6.25	0.13	0.16

3.7.6 *Testing the performance of the Poisson iREM for animals with observed zero speed of movement*

As discussed in Section 2.7.1 Rowcliffe et al. (2008) used day range to derive an approximate estimator of speed. Ten (10) focal watches, for each species, distributed evenly between 08:00-18:00 were conducted to monitor movement, and control for variation in movement patterns. Arbitrary animals were selected and followed around for 30 minutes during each focal watch, recording the total distance travelled during that time as the sum of all straight-line movements. Day range was then calculated as the mean across all focal watches for that species. Hence, the distribution of speeds over this short-time period was then used as the distribution of speeds over the entire trapping period. It was noted that during the period the animals were followed around, some of them were observed to not move, hence, there are records of zero speeds. We expect that these observed zero speeds would have an effect on estimated density. Note that in REM density is estimated as

$$\hat{D} = \frac{\lambda}{t}\frac{\pi}{(2+\theta)r\bar{v}}, \tag{3.7.2}$$

where r and θ are the detection distance and detection angle of the camera trap, respectively, and t is the camera trap time period. These parameters are held fixed. The parameter $\bar{v}$ is the mean speed. So including the observed zero speeds would result in a smaller estimate in the average speed, and hence, a larger estimate of the density compared with the density estimated density from excluding the observed zero speeds. For example, suppose we have $m = 4$ speed observations such that $v = \{0.453, 0, 1.865, 0\}$. Let $\bar{v} = 0.5795$ (km/day^{-1}), the mean speed including the observed zeros, and $\bar{v}_1 = 1.159$ (km/day^{-1}), the mean speed excluding the observed zeros, therefore, we would expect the density $\hat{D}$, estimated using $\bar{v}$ to be larger compared with the density $\hat{D}_1$ estimated using $\bar{v}_1$.

In this section we explore how a zero-adjusted gamma model works in estimating the density. As a demonstration we fit a Poisson iREM model to the encounter data. We also examine the fit of the gamma model with the observed speeds, excluding the zeros, and we compare estimates from iREM with estimates from REM. First, we test the ZAGA

model using large and small values of the density; $D = 100$ (km^2) or $D = 20$ (km^2) with a probability of the zero-response category set to $w = 0.30$, and sampling effort of the camera trap days and speed observations set to $n = 40;\ m = 40$ or $n = 100;\ m = 100$. The number of speed observations, m, is considered to be reasonably large, that is, greater than 30. The results are given in Table 3.7.8. The simulation results (3.7.8) indicate that the model works well in estimating the density for the given set of parameter values and sample sizes. Also, estimation precision improves when sample sizes increase. REM and iREM gave similar estimates of the density but REM gave smaller estimates of the standard errors and larger RMSEs.

Second, we examine the fit of the ZAGA model and the gamma for parameter values and sample sizes that reflect our WWAP data set. In Table 3.7.9 we give the results from fitting iREM and REM where animal speed is assumed to follow a ZAGA model or a gamma model. The density $D = 468$ (km^2) or $D = 119$ (km^2); the expected speed is $\mu_x = 0.71$ (km/day^{-1}) or $\mu_x = 1.17$ $\left(\text{km/day}^{-1}\right)$, respectively, and the probability of the zero-response category is $w = 0.21$ or $w = 0.40$, respectively. For each scenario, we set the number of the speed observations to $m = 10$, which is the sample size of the observed speeds for the WWAP data. We compare estimates of the density for larger sample sizes of the speed observations: $m = 40$ or $m = 70$. The number of camera trapping days, $n = 42$, which is the number of camera trapping days for the WWAP survey, is held fixed. The results (Table 3.7.9) show that for the ZAGA model, a positive bias is introduced for all scenarios and this bias is substantial for smaller sample sizes of the speed observations, in particular, when the density is large ($D = 468$ km^2). But for the gamma model the bias introduced when the sample size is small is much smaller, and the model appears to provide better estimates of the density compared with the ZAGA model. It is worth reiterating that including the observed zeros would give smaller estimates of the mean speed and, hence, inflate the density. This implies that the estimated mean speed from excluding the zeros would be larger, and hence, provide a smaller estimate of density. We give the sample sizes of the speed observations, excluding the zeros, for the first 6 simulation runs in Table 3.7.11. For example when $D = 119$ (km^2) (Table 3.7.9), the first sample size simulated for the speed observations, excluding the zeros, and used in the gamma model is 4 (Table 3.7.11, Case

(4)), which suggests that 6 out of the 10 simulated speeds are zeros. A sample size of 4 is a rather small sample size, therefore, we would expect a larger estimate of the mean speed, hence, a smaller estimate of the density, compared with the estimated mean speed for this same scenario from a ZAGA model which would provide a larger estimated density. In Table 3.7.9 average estimated densities with the true sample size, $m = 10$ used in the simulation process are: $\hat{D} = 166.08$ (0.27) (km^2) for the ZAGA model and $\hat{D} = 92.68$ (39.23) (km^2) for the gamma (average estimated standard errors in parentheses). On the other hand, the evidence shows that the ZAGA model provides reasonably accurate estimates of the density when m increases but a negative bias is introduced from the gamma model, which increases with sample size (see Table 3.7.9).

Finally, we examine the fit of the ZAGA model and the gamma model for a small fixed sample size of the speed observations $m = 10$ and increasing probabilities of the zero-response category. We set $D = 68$ (km^2) or $D = 468$ (km^2), which reflects the density from the census for two species at WWAP. The probability of the zero-response category is set to $w = 10$, $w = 20$, or $w = 30$. The number of camera trap days is set to $n = 42$. The results are given in Table 3.7.10. The simulation results show that for small m and small w ($= 0.10$) the bias introduced from a ZAGA model when D ($= 68$ km^2) is much smaller (and positive) compared with larger values of the density ($D = 468$ km^2). Also, the bias increases with increasing values of w. For the gamma model, when the density is small a negative bias is introduced, which increases with increasing values of w, and the bias is larger compared with the bias from a ZAGA model. On the other hand, when the density is large the bias introduced from the gamma model is smaller (and positive) compared with the bias produced by the ZAGA model, in particular, when $w = 0.20$. Note that we expect to have more non-zero speed observations in the sample for reasonably small values of w (see for e.g., sample Case (11) in Table 3.7.11).

Based on these simulation results we can conclude that the size of the density, the sample size of the speed observations and the size of the probability of the zero-response category would have an effect on estimated density. If there are observed zeros in the speed data, it is important to model these but we would recommend, where possible to increase survey effort, particularly for the speed observations and, if there are more zeros than

expected. One important point to note is that the WWAP survey is a special case where independent estimates of the speed and average group size were obtained by following the animals around, and during this time period some animals were observed to not move. As discussed in Section 2.1 the speed of movement of an animal is assumed to be constant but this speed may vary between animals. However, whether an animal varies it speeds or move at a fixed speed (or stop altogether) is not as important as the total distance covered by the individual over the trapping period, since this is what determines the probability of encounter. And since we are interested in the mean across animals of the mean speed over time for each individual, we need to include the observed zero speeds in estimating that mean, to avoid bias. But as shown, at least 60 observations are required to provide reasonably accurate estimates of the density when the probability of the zero-response category is high, and when the density is large. But for small density values, and if the probability of observing zero speeds is low then smaller sample sizes can provide reasonably accurate estimates of the density.

Table 3.7.8: Average parameter estimates for animals with observed zero speed of movement (average standard errors in parentheses). The following are the true values; $\mu_x = 0.71$ (km/day^{-1}) and $w = 0.30$. The Standard deviation (Sd) and Root Mean Square Error (RMSE) are also given.

	Poisson iREM			**Poisson REM**
	$\hat{D}$	$\hat{\mu}_x$	$\hat{w}$	$\hat{D}$
$D = 100,\ n = m = 40$				
estimate	103 (31.79)	0.73 (0.17)	0.29 (0.10)	105 (21.79)
Sd	29.27	0.16	0.07	32.47
RMSE	866.81	0.03	0.005	1079.09
$D = 100,\ n = m = 100$				
estimate	101 (19.61)	0.72 (0.10)	0.30 (0.07)	102 (13.28)
Sd	17.12	0.09	0.04	18.43
RMSE	294.20	0.01	0.02	342.58
$D = 20,\ n = m = 40$				
estimate	21 (10.96)	0.74 (0.17)	0.29 (0.10)	22 (10.71)
Sd	8.07	0.17	0.06	11.28
RMSE	65.77	0.03	0.004	129.76
$D = 20\ n = m = 100$				
estimate	20 (6.65)	0.72 (0.10)	0.30 (0.07)	20 (6.11)
Sd	5.33	0.10	0.04	6.12
RMSE	28.37	0.01	0.002	37.54

Table 3.7.9: Average parameter estimates from fitting a ZAGA model and a gamma model to the speed data where the density is set to $D = 468$ (km^2) or $D = 119$ (km^2) (average standard errors in parentheses), and where sample size of the speed data, m increases. The number of camera trapping days is $n = 42$. The chosen parameter values and sample sizes reflect the WWAP data set. The Standard deviation (Sd) and Root Mean Square Error (RMSE) are also given.

	ZAGA model				gamma model		
	Poisson iREM			**Poisson REM**	**Poisson iREM**		**Poisson REM**
	$\hat{D}$	$\hat{\mu}_x$	$\hat{w}$	$\hat{D}$	$\hat{D}$	$\hat{\mu}_x$	$\hat{D}$
$D = 468,\ m = 10,\ w = 0.21, \mu_x = 0.71$							
estimate	608.93 (283.27)	0.69 (0.68)	0.21 (0.16)	608.94 (56.98)	472.20 (204.13)	0.88 (0.37)	472.20 (44.19)
Sd	380.06	0.31	0.13	380.98	306.00	0.38	306.00
RMSE	405.35	0.10	0.13	405.35	306.02	0.42	306.02
$D = 468,\ m = 30,\ w = 0.21, \mu_x = 0.71$							
estimate	512.63 (147.64)	0.71 (0.53)	0.21 (0.10)	512.63 (47.54)	400.20 (108.71)	0.90 (0.23)	400.20 (37.12)
Sd	173.46	0.19	0.07	125.21	142.39	0.30	125.21
RMSE	179.11	0.04	0.07	179.11	142.39	0.29	142.39
$D = 468,\ m = 60,\ w = 0.21, \mu_x = 0.71$							
estimate	479.54 (102.67)	0.72 (0.41)	0.21 (0.07)	479.53 (44.73)	376.92 (76.63)	0.91 (0.16)	376.92 (35.16)
Sd	105.45	0.15	0.05	105.45	78.43	0.17	78.43
RMSE	106.08	0.02	0.05	106.08	120.19	0.26	120.19
$D = 119,\ m = 10,\ w = 0.40,\ \mu_x = 1.17$							
estimate	166.08 (0.27)	1.16 (0.72)	0.39 (0.27)	166.08 (24.08)	92.68 (39.23)	1.89 (0.75)	92.68 (13.44)
Sd	123.73	0.62	0.14	123.73	54.00	0.87	54.00
RMSE	132.39	0.38	0.14	132.39	60.07	1.13	60.07
$D = 119,\ m = 30,\ w = 0.40,\ \mu_x = 1.17$							
estimate	133.42 (44.10)	1.15 (0.76)	0.40 (0.15)	133.42 (19.21)	77.96 (22.83)	1.92 (0.49)	77.96 (11.23)
Sd	43.73	0.35	0.09	133.42	23.43	0.49	23.43
RMSE	46.05	0.13	0.09	46.05	47.26	0.90	47.26
$D = 119,\ m = 60,\ w = 0.40,\ \mu_x = 1.17$							
estimate	122.26 (31.45)	1.17 (0.64)	0.40 (0.11)	122.26 (17.82)	72.00 (16.83)	1.97 (0.36)	72.00 (10.49)
Sd	33.22	0.24	0.06	33.22	16.63	0.35	16.63
RMSE	33.38	0.06	0.06	33.38	49.86	0.87	49.86

Table 3.7.10: Average parameter estimates from fitting a ZAGA model and a gamma model to the speed data where the probability of the zero-response category, w increases (average standard errors in parentheses). The chosen parameter values and sample sizes reflect the WWAP data set. Here we set the density, $D = 68$ (km^2) or $D = 468$ (km^2). The Standard deviation (Sd) and Root Mean Square Error (RMSE) are also given.

	ZAGA model				gamma model		
	Poisson iREM			**Poisson REM**	**Poisson iREM**		**Poisson REM**
	$\hat{D}$	$\hat{\mu}_x$	$\hat{w}$	$\hat{D}$	$\hat{D}$	$\hat{\mu}_x$	$\hat{D}$
$D = 68,\ m = 10,\ w = 0.10, \mu_x = 2.56$							
estimate	71.78 (24.28)	2.62 (0.31)	0.10 (0.06)	71.77 (9.35)	63.79 (20.42)	2.91 (0.85)	63.79 (8.31)
Sd	24.27	0.31	0.06	24.83	20.59	0.38	20.59
RMSE	25.11	0.64	0.09	25.11	21.02	0.92	21.02
$D = 68,\ m = 10,\ w = 0.20, \mu_x = 2.56$							
estimate	75.64 (26.94)	2.62 (0.36)	0.19 (0.14)	75.64 (9.72)	59.34 (18.84)	3.24 (0.93)	59.34 (7.63)
Sd	29.86	0.90	0.12	29.86	20.53	0.98	20.53
RMSE	179.11	0.04	0.07	179.11	142.39	0.29	142.39
$D = 68,\ m = 10,\ w = 0.30, \mu_x = 2.56$							
estimate	77.52 (28.26)	2.66 (0.41)	0.28 (0.20)	77.52 (9.72)	53.11 (15.83)	3.69 (0.98)	53.11 (6.84)
Sd	39.98	0.97	0.14	39.98	20.53	0.98	21.92
RMSE	179.11	0.04	0.07	179.11	142.39	0.29	142.39
$D = 468,\ m = 10,\ w = 0.10, \mu_x = 0.71$							
estimate	575.23 (245.16)	0.74 (0.66)	0.09 (0.05)	575.23 (53.68)	512.39 (211.85)	0.81 (0.33)	512.19 (47.80)
Sd	412.21	0.33	0.09	412.21	322.91	0.36	321.29
RMSE	405.35	0.10	0.13	405.35	306.02	0.42	306.02
$D = 468,\ m = 10,\ w = 0.20, \mu_x = 0.71$							
estimate	600.77 (269.94)	0.74 (0.65)	0.19 (0.13)	600.77 (56.06)	455.28 (190.12)	0.92 (0.37)	455.28 (42.48)
Sd	514.74	0.35	0.14	514.74	278.88	0.40	514.74
RMSE	405.35	0.10	0.13	405.35	306.02	0.42	306.02
$D = 468,\ m = 10,\ w = 0.30, \mu_x = 0.71$							
estimate	636.37 (302.74)	0.72 (0.68)	0.30 (0.22)	636.37 (59.28)	416.46 (174.92)	1.05 (0.43)	416.46 (38.79)
Sd	522.94	0.35	0.14	522.94	263.22	0.52	263.22
RMSE	405.35	0.10	0.13	405.35	306.02	0.42	306.02

Table 3.7.11: Sample sizes of speed observations, excluding the zero speeds, for the first 6 simulation runs for the simulation scenarios examined.

Sample sizes for scenarios in Table 3.7.9							**Sample sizes for scenarios in Table 3.7.10**						
	$D = 468,\ m = 10,\ w = 0.21,\ \mu_x = 0.71$							$D = 68,\ m = 10,\ w = 0.10,\ \mu_x = 2.56$					
Case (1)	8	8	5	9	9	10	Case (7)	8	10	10	10	8	9
	$D = 468,\ m = 30,\ w = 0.21,\ \mu_x = 0.71$							$D = 68,\ m = 10,\ w = 0.10,\ \mu_x = 2.56$					
Case (2)	23	24	25	22	21	22	Case (8)	10	8	7	8	7	9
	$D = 468,\ m = 60,\ w = 0.21,\ \mu_x = 0.71$							$D = 68,\ m = 10,\ w = 0.10,\ \mu_x = 2.56$					
Case (3)	46	49	53	51	50	47	Case (9)	6	5	5	7	9	5
	$D = 119,\ m = 10,\ w = 0.40, \mu_x = 1.17$							$D = 468,\ m = 10,\ w = 0.10,\ \mu_x = 2.56$					
Case (4)	4	6	6	6	4	7	Case (10)	10	9	9	9	8	9
	$D = 119,\ m = 30,\ w = 0.40, \mu_x = 1.17$							$D = 468,\ m = 10,\ w = 0.20,\ \mu_x = 2.56$					
Case (5)	20	18	14	21	18	17	Case (11)	10	7	7	9	8	8
	$D = 119,\ m = 60,\ w = 0.40,\ \mu_x = 1.17$							$D = 468,\ m = 10,\ w = 0.30,\ \mu_x = 2.56$					
Case (6)	36	39	38	31	36	33	Case (12)	4	4	7	5	7	6

3.8 Application of iREM to WWAP data set

Here we illustrate the application of iREM using the WWAP data set for the four species: wallaby, which is discussed in Section 3.8.1; water deer, which is discussed in Section 3.8.2; muntjac, which is discussed in Section 3.8.3; and mara, which is discussed in Section 3.8.4. Estimates of the density for animals with observed zero speed of movement are given in Section 3.8.5, while estimates of density for species moving in family groups are given in Section 3.8.6. Given that the underlying distribution in REM is the Poisson model, estimates from a Poisson REM are compared with estimates from iREM where encounters are assumed to follow a Poisson model. These results are given in Section 3.8.7. We use the formula provided in equation (3.7.1) in Section 3.7 above to compute approximate 95% confidence intervals of the density.

As discussed in Chapter 2, Section 2.7.1 and Section 3.7.6 above Rowcliffe et al. (2008) conducted 10 focal watches between 08:00 and 18:00 during the WWAP survey to monitor and record movement patterns of animals to estimate average speed and average group size. During each focal watch, arbitrary animals were selected and followed around for 30 minutes. It was observed that some animals did not move during this time period, hence, zero speeds were recorded. We, therefore, illustrate iREM using the data with the observed zeros and without the observed zeros. Appropriate models are used to account for the observed zeros, and where the zeros are excluded from the model. A summary of the speed data, excluding the observed zero speed from Rowcliffe et al. (2008) is given in Table 3.8.1. The number of animal speed data for each species is $m = 10$. This data set is used in the analysis where animal speed is assumed to follow a ZAGA model. The speed data, excluding the observed zero speeds, is used in the analysis where animal speed is assumed to follow a gamma model, a lognormal model or a Weibull model. The coefficient of variation, $C_v\%$ expressed as a percentage, for each species is also given. The $C_v\%$ for each species is huge, and for both cases where the observed zero speeds are included or excluded from the analysis. This suggests that the variability in the speed data is substantial.

Table 3.8.1: Summary of the mean speeds, mean encounters and sample size of speeds (standard error in parentheses). The coefficient of variation ($C_v\%$) expressed as a percentage is also given

Species	mean speed without zeros, $\bar{v}$, in km/day^{-1}	$C_v\%$	Sample size without zero (m)	mean speed with zeros, $\bar{v}$, in km/day^{-1}	$C_v\%$	Average group size, g
mara	3.66 (1.35)	97%	7	2.56 (1.08)	133%	1.80 (0.40)
muntjac	8.27 (1.92)	73%	10	8.27 (1.92)	73%	1.50 (0.28)
wallaby	0.88 (0.43)	138%	8	0.71 (0.36)	160%	1 (0)
water deer	1.95 (0.60)	75%	6	1.17 (0.47)	127%	1 (0)

3.8.1 Estimated density of the wallaby species

Table 3.8.2 shows the parameter estimates of the wallaby species and the ΔAIC values. The three speed data models gave similar estimates of the parameters but the results suggest that a lognormal model is the best model for the speed data of the wallaby species (Table 3.8.2). The results also show that a NB iREM model fits the wallaby data set substantially better than all models, except for the ZINB iREM model which fits closely to the NB iREM model. NB iREM and ZINB iREM gave similar estimates to two decimal places. The results also show that a Poisson iREM, NB iREM and ZINB iREM gave similar estimates of the density but a Poisson iREM slightly underestimates the standard error. The results show that there is overdispersion in the data, which may have caused the underestimation of the standard error in a Poisson iREM. Note that as κ increases the variance of the encounter data decreases and the model tends to a Poisson iREM. While the estimated densities are not close to the density from the census, the estimated standard errors are large enough to capture the density from the census within an approximate 95% confidence interval. The ΔAIC values show no support for allowing for zero-inflation.

Table 3.8.2: Parameter estimates of the wallaby species from the different iREM being fitted to the data (standard error in parentheses). The ΔAIC values for the different iREM are also given.

	Count data models			
	Poisson	NB	ZIP	ZINB
Speed data models		**Census (D) = 468**		
gamma				
$\hat{D}$	653 (243.20)	653 (249.55)	763 (283.85)	653 (249.51)
$\hat{\mu}_x$	0.88 (0.65)	0.88 (0.66)	0.88 (0.65)	0.88 (0.66)
$\hat{\nu}$	1.03 (0.59)	1.03 (0.59)	1.03 (0.59)	1.03 (0.59)
$\hat{k}$	-	0.95 (0.14)	-	1.06 (0.15)
$\hat{\rho}$	-	-	0.14 (0.03)	3e-07 (0.0001)
lognormal				
$\hat{D}$	633 (336.29)	634 (340.70)	739 (392.60)	633 (340.57)
$\hat{\mu}_x$	0.91 (0.50)	0.91 (0.52)	0.91 (0.50)	0.91 (0.52)
$\hat{\nu}$	1.16 (0.29)	1.16 (0.29)	1.16 (0.29)	1.16 (0.29)
$\hat{k}$	-	0.95 (0.14)	-	1.06 (0.15)
$\hat{\rho}$	-	-	0.14 (0.03)	4.26e-06 (0.0003)
Weibull				
$\hat{D}$	659 (261.70)	659 (267.72)	768 (305.42)	662 (268.92)
$\hat{\mu}_x$	0.87 (0.04)	0.87 (0.10)	0.87 (0.04)	0.87 (0.10)
$\hat{\nu}$	0.89 (0.23)	0.89 (0.23)	0.89 (0.23)	0.89 (0.23)
$\hat{k}$	-	0.95 (0.14)	-	1.06 (0.15)
$\hat{\rho}$	-	-	0.14 (0.03)	9.5e-05 (0.002)
		ΔAIC values by iREM models		
gamma	407.46	1.16	310.96	3.16
lognormal	406.30	0	309.80	2.00
Weibull	407.30	1.00	310.82	3.54

3.8.2 Estimated density of the water deer species

The parameter estimates and the ΔAIC values for the water deer species are given in Table 3.8.3. The ZINB iREM gave estimates that are reasonably close to the density from the census ($D = 119$), in particular when the speed of movement is assumed to follow a gamma model or a Weibull model. When the speed of movement is assumed to follow a lognormal model the difference in estimated density and the density from the census is quite large. The results show strong support for allowing for overdispersion in the model but no support for allowing for zero-inflation.

According to the ΔAIC values the NB iREM where animal speed is assumed to follow a Weibull model has been selected as the best model that explains the water deer data set. But the NB iREM where animal speed is assumed to follow a gamma model and ZINB iREM where animal speed is assumed to follow a Weibull model fit closely to the NB iREM where the speed data is assumed to follow a Weibull model, suggesting that there is no real difference between these models. For all models, the density from the census is captured within an approximate 95% confidence interval.

Table 3.8.3: Parameter estimates of the water deer species from the different iREM being fitted to the data (standard error in parentheses). The ΔAIC values for the different iREM are also given.

	Count data models			
	Poisson	NB	ZIP	ZINB
Speed data models		**Census (D) = 119**		
gamma				
$\hat{D}$	91 (36.76)	91 (38.29)	174 (70.35)	115 (55.95)
$\hat{\mu}_x$	1.95 (2.54)	1.95 (2.58)	1.95 (2.55)	1.95 (2.57)
$\hat{\nu}$	0.54 (0.35)	0.54 (0.35)	0.54 (0.35)	0.54 (0.35)
$\hat{k}$	-	0.51 (0.10)	-	1.21 (0.75)
$\hat{\rho}$	-	-	0.48 (0.09)	0.27 (0.32)
lognormal				
$\hat{D}$	59 (47.60)	60 (48.19)	113 (91.00)	75 (63.60)
$\hat{\mu}_x$	3.00 (0.70)	2.99 (0.72)	2.99 (0.70)	3.00 (0.71)
$\hat{\nu}$	1.39 (0.40)	1.39 (0.40)	1.39 (0.40)	1.39 (0.40)
$\hat{k}$	-	0.51 (0.10)	-	1.20 (0.74)
$\hat{\rho}$	-	-	0.48 (0.09)	0.28 (0.32)
Weibull				
$\hat{D}$	92 (33.16)	92 (34.89)	176 (63.56)	117 (52.50)
$\hat{\mu}_x$	1.94 (0.10)	1.94 (0.14)	1.93 (0.10)	1.93 (0.14)
$\hat{\nu}$	1.15 (0.41)	1.15 (0.41)	1.15 (0.41)	1.15 (0.41)
$\hat{k}$	-	0.51 (0.10)	-	1.21 (0.75)
$\hat{\rho}$	-	-	0.48 (0.09)	0.27 (0.32)
		ΔAIC values by iREM models		
gamma	143.96	0.14	36.38	1.60
lognormal	146.44	2.62	38.86	4.08
Weibull	143.82	0	36.24	1.46

3.8.3 *Estimated density of the muntjac species*

The muntjacs were observed to be moving in pairs or family groups, and the density estimated in REM is that of the group instead of the individual density. To obtain the individual density, Rowcliffe et al. (2008) multiplied the density of the group by an independent estimate of average group size, $g = 1.50$ (Table 3.8.1). Here we multiplied the estimated density by average group, g, to give the individual density, and the results and ΔAIC values are given in Table 3.8.4.

The results from fitting a Poisson iREM, a NB iREM, a ZIP iREM and a ZINB iREM (Table 3.8.4) show that the ZINB iREM gave better estimates of the density than the other models with the estimated density from the ZINB iREM being closer to the density from the census. The results also indicate that the ZINB iREM is the best model that explains the muntjac data (Table 3.8.4). There is some support for allowing zero-inflation in the model. For the ZINB iREM and the ZIP iREM the density from the census is captured within an approximate 95% confidence interval but the Poisson iREM and the NB iREM fail to do so.

Table 3.8.4: Parameter estimates of the muntjac species from fitting iREM (standard error in parentheses). The ΔAIC values for the different iREM are also given.

	Count data models			
	Poisson	NB	ZIP	ZINB
Speed data models		**Census (D) = 13**		
gamma				
$\hat{D}$	8 (1.94)	8 (2.19)	23 (6.92)	16 (10.65)
$\hat{\mu}_x$	8.27 (14.70)	8.27 (14.80)	8.27 (14.77)	8.28 (14.98)
$\hat{\nu}$	0.27 (0.12)	0.27 (0.12)	0.27 (0.12)	0.26 (0.12)
$\hat{k}$	-	0.37 (0.15)	-	0.63 (1.17)
$\hat{\rho}$	-	-	0.67 (0.20)	0.52 (0.62)
lognormal				
$\hat{D}$	7 (2.21)	7 (2.43)	22 (7.55)	15 (10.68)
$\hat{\mu}_x$	8.52 (0.37)	8.52 (0.43)	8.52 (0.40)	8.52 (0.42)
$\hat{\nu}$	0.74 (0.17)	0.74 (0.17)	0.74 (0.17)	0.74 (0.17)
$\hat{k}$	-	0.37 (0.15)	-	0.64 (1.19)
$\hat{\rho}$	-	-	0.67 (0.20)	0.52 (0.63)
Weibull				
$\hat{D}$	8 (1.92)	8 (2.18)	23 (6.86)	16 (10.72)
$\hat{\mu}_x$	8.32 (0.15)	8.32 (0.16)	8.32 (0.16)	8.32 (0.16)
$\hat{\nu}$	1.53 (0.36)	1.53 (0.36)	1.53 (0.36)	1.53 (0.30)
$\hat{k}$	-	0.37 (0.15)	-	0.64 (1.20)
$\hat{\rho}$	-	-	0.67 (0.20)	0.52 (0.63)
		ΔAIC values by iREM models		
gamma	21.24	0.08	1.00	0
lognormal	21.62	0.44	1.36	2.00
Weibull	21.60	0.44	1.34	1.98

3.8.4 Estimated density of the mara species

The mara species was also observed to be moving in pairs and family groups, and the estimate of the density from iREM is the density of the group. The individual density is obtained by multiplying estimated density of the group by an independent estimate of average group size, $g = 1.8$, given in Table 3.8.1 above. At WWAP, camera traps were not randomized nor were they set on trails in the area where the mara inhabited. These traps were placed nonrandomly and away from the areas where the mara frequented, which resulted in low trap rates, and which would explain the dramatic underestimation of the mara abundance. Table 3.8.5 gives the parameter estimates of the mara species. The results show that $\hat{\kappa}$ is substantially large, suggesting a Poisson iREM would be

more appropriate to fit the data than other models such as a negative binomial model, and as indicated by the ΔAIC values. The NB REM proposed in this Chapter for the encounter observations, y_{ij} has variance, $\text{Var}(y_{ij}) = \hat{\lambda}+(1/\hat{\kappa})\hat{\lambda}^2$, and as $1/\hat{\kappa} \to 0$ the NB iREM tends to a Poisson iREM. It is worth noting that the sample size of the encounter data for mara species is rather limited (see Table 2.7.1 in Section 2.7.1), which may be a reason for the huge estimates of κ. While the ΔAIC values indicate that the Poisson iREM where animal speed is assumed to follow a gamma model is the best model that explains the data set of the mara species, there is not much difference in ΔAIC between the other speed data models.

Table 3.8.5: Parameter estimates of the mara species from fitting various iREMs (standard error in parentheses). The ΔAICfor the different iREMs are also given.

	Count data models			
	Poisson	NB	ZIP	ZINB
Speed data models		**Census (D) = 68**		
gamma				
$\hat{D}$	4 (2.00)	4 (2.00)	4 (2.00)	4 (2.14)
$\hat{\mu}_x$	3.65 (11.26)	3.66 (11.29)	3.66 (11.30)	3.61 (10.97)
$\hat{\nu}$	0.22 (0.14)	0.22 (0.14)	0.22 (0.14)	0.22 (0.14)
$\hat{k}$	-	1.09e+07 (2.39e+08)	-	0.0001 (0.02)
$\hat{\rho}$	-	-	4e-05 (0.01)	0.01 (0.18)
lognormal				
$\hat{D}$	3 (2.32)	3 (2.32)	3 (2.52)	3 (2.36)
$\hat{\mu}_x$	5.03 (0.78)	5.03 (0.78)	5.08 (0.78)	5.01 (0.78)
$\hat{\nu}$	1.46 (1.29)	1.46 (0.39)	1.47 (0.39)	1.47 (0.39)
$\hat{k}$	-	6.50e+07 (8.27e+08)	-	0.0002 (0.03)
$\hat{\rho}$	-	-	0.03 (0.30)	0.002 (0.09)
Weibull				
$\hat{D}$	4 (2.02)	4 (2.12)	4 (2.02)	4 (2.02)
$\hat{\mu}_x$	3.68 (0.65)	3.68 (0.65)	3.68 (0.65)	3.69 (0.66)
$\hat{\nu}$	0.89 (0.28)	0.88 (0.28)	0.88 (0.28)	0.88 (0.28)
$\hat{k}$	-	4.05e+08 (1.79e+09)	-	0.0004 (0.04)
$\hat{\rho}$	-	-	5e-05 (0.01)	0.001 (0.05)
		ΔAIC values by iREM models		
gamma	0	2.00	2.00	4.00
lognormal	0.95	2.95	2.95	4.95
Weibull	0.09	2.20	2.20	4.20

3.8.5 *Estimated density of animals with observed zero speed of movement*

Table 3.8.6 gives the estimates of the density of species with observed zero animal speed. The estimates associated with animals that did not move about much, and therefore, have a zero record of speed of movement suggest that these zeros do have an effect on estimated density. As shown by the simulation results in Section 3.7.6 for large density values and small sample size of the speed observations, estimated density is inflated because of a smaller estimate of the expected speed. For the wallaby and water species the density from the census is large and the probabilities of observing a zero speed are also reasonably large, so we expect the estimated density to be inflated. Note that the sample size for the speed observations is rather small ($m = 10$) and the evidence in the simulation study (Section 3.7.6) showed that for such small sample sizes and, in particular, for larger values of the density, a positive bias would be introduced, and this bias would be substantial. When we examine the estimates of the density where the observed zero speeds are excluded from the model (Sections, 3.8.1, 3.8.2 and 3.8.4 above) we see that the bias introduced is much smaller compared with the bias from fitting a ZAGA model. These results are expected, and are similar to the results given in the simulation study (see Section 3.7.6). This evidence suggests that the sample sizes are too small, and the ZAGA model would provide better estimates of the density if we sample more speed observations. For the mara species, we expect the estimated density to be small because of the low trapping rate. Therefore, a Poisson iREM may be more appropriate to model encounter data with limited data as indicated by the huge estimate of κ. Note that the muntjacs were considerably more mobile than the other species, and therefore, no record of zero speed of movement. The results also show that for species with higher trapping rates (wallaby and water deer) there is more support for models that allow for overdispersion in the data, while the Poisson iREM is more appropriate for animals with low trap rates such as the mara species (Table 3.8.7).

Table 3.8.6: Parameter estimates from fitting different iREMs to speed observations distributed by a ZAGA model (standard error in parentheses).

Count data models	**Species**		
	wallaby	water deer	mara
Poisson			
$\hat{D}$	816 (330.23)	152 (72.73)	5 (3.06)
$\hat{\mu}_x$	0.66 (0.76)	0.65 (1.28)	2.56 (2.89)
$\hat{\nu}$	0.88 (0.33)	1.95 (0.78)	3.66 (1.55)
$\hat{w}$	0.25 (0.20)	0.67 (0.43)	0.30 (0.21)
NB			
$\hat{D}$	816 (337.58)	152 (74.89)	5 (3.06)
$\hat{\mu}_x$	0.71 (0.80)	1.17 (1.33)	2.56 (2.89)
$\hat{\nu}$	0.88 (0.33)	1.95 (0.78)	3.66 (1.55)
$\hat{k}$	0.95 (0.14)	0.51 (0.10)	2.00e+05 (3.91e+08)
$\hat{w}$	0.20 (0.16)	0.40 (0.26)	0.30 (0.21)
ZIP			
$\hat{D}$	952 (385.18)	310 (145.41)	5 (3.62)
$\hat{\mu}_x$	0.66 (0.76)	0.62 (1.19)	2.55 (2.88)
$\hat{\nu}$	0.88 (0.33)	1.81 (0.70)	3.63 (1.54)
$\hat{w}$	0.25 (0.20)	0.87 (0.16)	0.30 (0.21)
$\hat{\rho}$	0.17 (0.04)	0.66 (0.43)	0.03 (0.32)
ZINB			
$\hat{D}$	882 (249.55)	192 (105.55)	5 (3.13)
$\hat{\mu}_x$	0.88 (0.66)	1.17 (2.40)	2.54 (131.62)
$\hat{\nu}$	1.03 (0.59)	1.96 (0.78)	3.63 (1.54)
$\hat{k}$	1.06 (0.15)	1.21 (0.75)	1.62e-8 (7.04e-23)
$\hat{w}$	0.24 (0.18)	0.21 (0.25)	0.30 (0.21)
$\hat{\rho}$	0.001 (0.01)	0.40 (0.26)	0.003 (1.74)

Table 3.8.7: ΔAIC from different iREMs where animal speed follows a ZAGA.

	Count data models			
	Poisson	NB	ZIP	ZINB
Species				
wallaby	406.30	0	690.18	2.14
water deer	155.96	12.14	48.47	0
mara	0	2.00	2.01	4.00

3.8.6 Estimated density of animals moving in groups

At WWAP the maras and muntjacs were the only species to be found moving in pairs or family groups. We provide estimates of the density by including group size data in iREM. Table 3.8.8 gives the results of the mara species. The results show that including group size data in iREM to estimate density of the mara species induced a moderate, but even worse, change in density estimation across all models. Also, including group size in the model for the muntjac species induced an underestimation of the density when allowing for zero-inflation and overdispersion (Table 3.8.9), compared with the estimates resulting from multiplying density and a fixed value of average group size. However, by only allowing for zero-inflation the density estimated from the model improves and is nearer to the density from the census. For models which do not allow for zero-inflation (Poisson iREM and NB iREM) estimated density is less accurate, which resulted in the density from the census not being captured within an approximate 95% confidence interval. The average group size, g, is estimated exactly as that in Rowcliffe et al. (2008) for both species but it is worth noting that g is relatively small (less than 2), which might suggest that animals are still moving individually.

Table 3.8.8: Parameter estimates from fitting different iREMs to the mara species data. Estimates of average group size are given (standard error in parentheses).

	speed data models		
	gamma	lognormal	Weibull
Poisson			
$\hat{D}$	2 (1.17)	1 (1.32)	2 (1.17)
$\hat{\mu}_x$	3.66 (1.55)	5.03 (4.00)	3.68 (1.59)
$\hat{\nu}$	0.22 (0.14)	1.46 (0.39)	0.88 (0.28)
$\hat{g}$	1.80 (0.30)	1.80 (0.30)	1.80 (0.30)
NB			
$\hat{D}$	2 (1.16)	1 (1.32)	2 (1.17)
$\hat{\mu}_x$	3.65 (1.55)	5.02 (3.99)	3.67 (1.59)
$\hat{\nu}$	0.22 (0.14)	1.46 (0.39)	0.88 (0.28)
$\hat{k}$	0.0002 (0.02)	0.00001 (0.001)	0.00003 (0.002)
$\hat{g}$	1.80 (0.30)	1.80 (0.30)	1.80 (0.30)
ZIP			
$\hat{D}$	2 (1.17)	1 (1.32)	2 (1.17)
$\hat{\mu}_x$	3.65 (1.55)	5.02 (4.00)	3.68 (1.59)
$\hat{\nu}$	0.22 (0.14)	1.46 (0.39)	0.88 (0.28)
$\hat{g}$	1.80 (0.30)	1.80 (0.30)	1.80 (0.30)
$\hat{\rho}$	0.00001 (0.004)	0.00001 (0.0002)	0.00001 (0.02)
ZINB			
$\hat{D}$	2 (1.19)	2 (1.35)	2 (1.18)
$\hat{\mu}_x$	3.58 (1.51)	5.03 (3.99)	3.68 (1.59)
$\hat{\nu}$	0.22 (0.14)	1.46 (0.39)	0.88 (0.28)
$\hat{k}$	0.01 (0.17)	0.005 (0.13)	0.002 (0.07)
$\hat{g}$	1.80 (0.31)	1.80 (0.31)	1.80 (0.31)
$\hat{\rho}$	0.000003 (0.003)	0.01 (0.18)	0.001 (0.06)

Table 3.8.9: Parameter estimates from fitting different iREMs to the muntjac species data. Estimates of the average group size are also given (standard error in parentheses).

	speed data models		
	gamma	lognormal	Weibull
Poisson			
$\hat{D}$	5 (1.46)	5 (1.60)	5 (1.45)
$\hat{\mu}_x$	8.27 (1.76)	8.52 (2.25)	8.32 (1.76)
$\hat{\nu}$	0.27 (0.12)	0.74 (0.17)	1.53 (0.36)
$\hat{g}$	1.50 (0.06)	1.50 (0.06)	1.50 (0.06)
NB			
$\hat{D}$	5 (1.59)	5 (1.74)	5 (1.70)
$\hat{\mu}_x$	8.27 (1.77)	8.68 (2.35)	8.31 (1.67)
$\hat{\nu}$	0.26 (0.12)	0.75 (0.17)	1.62 (0.37)
$\hat{k}$	2.90 (1.16)	2.85 (1.14)	3.00 (1.19)
$\hat{g}$	1.50 (0.06)	1.49 (0.06)	1.57 (0.06)
ZIP			
$\hat{D}$	15 (5.16)	15 (5.49)	15 (5.12)
$\hat{\mu}_x$	8.27 (1.76)	8.52 (2.25)	8.32 (1.76)
$\hat{\nu}$	0.27 (0.12)	0.74 (0.17)	1.53 (0.36)
$\hat{g}$	1.50 (0.06)	1.50 (0.06)	1.50 (0.06)
$\hat{\rho}$	0.68 (0.07)	0.68 (0.07)	0.68 (0.07)
ZINB			
$\hat{D}$	9 (8.13)	9 (7.72)	9 (8.17)
$\hat{\mu}_x$	8.27 (1.76)	8.48 (2.23)	8.32 (1.76)
$\hat{\nu}$	0.27 (0.13)	0.74 (0.16)	1.53 (0.36)
$\hat{k}$	1.06 (2.12)	0.92 (1.76)	0.94 (1.91)
$\hat{g}$	1.50 (0.06)	1.50 (0.06)	1.50 (0.06)
$\hat{\rho}$	0.43 (0.51)	0.46 (0.43)	0.46 (0.47)

3.8.7 Density estimates from iREM compared with REM

To compare estimations from REM with estimations from iREM, we fit a Poisson iREM to the data since the Poisson model is the underlying distribution in REM. We have shown previously in the simulation study (Section 3.7) that there is minimal difference in density estimates between the three speed data models, therefore, as a demonstration we fit a gamma model to the animal speed data. Estimates associated with including the observed zero speeds in REM are compared with estimates from fitting a ZAGA

model to animal speed in iREM. While estimates associated with excluding the observed zero speeds in REM are compared with estimates from fitting a gamma model to animal speed in iREM. Table 3.8.10 compares estimates of the density from iREM with estimates from REM. The results show that there is minimal difference in the density estimates between REM and iREM but clearly REM underestimates the standard error. REM does not account for the variability in speed observations, hence, the underestimation of the standard error of the density. As shown in Table 3.8.1 above the coefficient of variations ($C_v\%$) of the speed estimator, with and without the observed zero speed, are huge suggesting that the variability in the speed is high. Incorporating this high variability in the speed in the modelling process would influence the standard error estimates as shown by the estimates from iREM. There is a huge difference in the density estimates between the models including and excluding the observed zero speeds. We expect the density to be inflated when the observed zero speeds are included in the model as incorporating these would reduce the estimated mean speed, hence, increasing the density.

We compute the confidence intervals from using the Hessian estimates of the standard error and the results are given in Table 3.8.11. Note that there is no result from the ZAGA model for the muntjac species since there were no observed zero speeds. The results from REM show that the density from the census for all species is not captured within an approximate 95% confidence interval. It is worth reiterating that the sampling variability in speed is not considered in REM, therefore, the standard errors of the density would be smaller, and hence the confidence interval would be narrow. Modelling the variability in the speed observations provides larger standard errors of the density and wider confidence intervals, but the density from the census for only two of the four species is captured within an approximate 95% confidence interval. The estimates from Hessian rely on the assumption that sample size approaches infinity, therefore, finite sample properties are sometimes less than optimal, for example maximum likelihood estimates (parameter estimates and standard error estimates) may be biased. And, given that the sample size of the data set for each species is small, we would expect some bias in the estimates from the information matrix. We therefore consider bootstrap estimates and estimates of the standard error from the adjusted variance method, and

the results are given in Table 3.8.12. The bootstrap method and the adjusted variance method incorporate the sampling variability in the speed data and the encounter data, and as shown in Table 3.8.1 above the variability in the speed is high, given the huge coefficient of variation ($C_v\%$). Therefore, we expect larger standard errors and wider confidence intervals. As discussed in Section 2.8.2 in Chapter 2, the bootstrap estimates of the standard error from REM were very skewed, resulting in larger estimates compared with those from the adjusted variance method. However, the similarities between the estimates from the two methods are more clear when computed on the log-scale. This is also the case for the bootstrap estimates from iREM (see Table 3.8.12). Again, for most cases, bootstrap standard error estimates from iREM are larger compared with the standard error estimates from REM, however, the difference in the estimates between the REM and iREM is not substantial as that from using the Hessian matrix. Therefore, we can consider using REM for density estimation but the bootstrap method would be required to estimate the standard error of the density. This is supported by the fact that the density from the census for three of the four species is captured within a 95% confidence interval for both REM and iREM; see Table 3.8.12.

Table 3.8.10: Density from the census compared with estimated density from iREM and REM (standard error from the inverse Hessian matrix in parentheses).

Species	**Census**	**REM**		**iREM**	
		Speed data without zero speed	Speed data with zero speed	gamma	ZAGA
	D	$\hat{D}$	$\hat{D}$	$\hat{D}$	$\hat{D}$
wallaby	468	654 (24.93)	816 (31.10)	653 (243.20)	816 (330.23)
water deer	119	91 (6.26)	155 (10.52)	91 (36.76)	152 (72.73)
mara	68	4 (1.23)	5 (1.76)	4 (2.00)	5 (3.06)
muntjac	13	8 (1.07)	8 (1.07)	8 (1.61)	-

Table 3.8.11: Comparing standard errors from Hessian matrix for REM and iREM. Approximate 95% confidence intervals are also given.

Species	**Census**	**REM**		**iREM**	
	D	Hessian Standard error without zero speed	Hessian Standard error with zero speed	gamma model Hessian Standard error	ZAGA model Hessian Standard error
wallaby	468	24.88	31.10	243.20	330.23
water deer	119	6.31	10.52	36.76	72.73
mara	68	1.23	1.76	2.00	3.06
muntjac	13	1.07	1.07	1.61	-
		Approximate 95% Confidence Interval			
wallaby	468	(606.02, 703.63)	(757.53, 876.54)	(315.80, 1350.13)	(368.97, 1804.36)
water deer	119	(81.64, 106.49)	(136.06, 177.44)	(42.73, 203.24)	(60.77, 397.33)
mara	68	(1.93, 7.12)	(1.61, 8.53)	(2.95, 9.49)	(1.70, 16.44)
muntjac	13	(5.74, 10.00)	(5.74, 10.00)	(4.99, 11.51)	(-)

Table 3.8.12: Bootstrap standard errors are compared with standard errors from the adjusted variance method. We also give 95% confidence intervals.

Species	**Census**	**Bootstrap Standard Error**				**Adjusted Method**	
		REM		**iREM**			
		Standard error without zero speed ($Se\{\log(\hat{D})\}$)	Standard error with zero speed ($Se\{\log(\hat{D})\}$)	gamma model Standard error ($Se\{\log(\hat{D})\}$)	ZAGA model Standard error ($Se\{\log(\hat{D})\}$)	Standard error without zero speed ($Se\{\log(\hat{D})\}$)	Standard error with zero speed ($Se\{\log(\hat{D})\}$)
wallaby		429.36 (0.49)	635.09 (0.52)	443.38 (0.49)	640.17 (0.53)	317.38 (0.48)	414.19 (0.51)
water deer		43.32 (0.35)	118.70 (0.46)	49.84 (0.39)	164.74 (0.51)	29.34 (0.31)	63.63 (0.41)
mara		3.77 (0.56)	11.22 (0.63)	3.90 (0.57)	55.14 (0.72)	1.19 (0.32)	1.80 (0.34)
muntjac		2.43 (0.30)	2.50 (0.31)	2.72 (0.32)	-	1.43 (0.19)	-
		95% **Confidence Interval (CI) of Density**					
		REM (Bootstrap) CI without zero speed	**REM (Bootstrap) CI** with zero speed	**iREM (Bootstrap) CI** from gamma model	**iREM (Bootstrap) CI** from ZAGA model	**Adjusted CI** without zero speed	**Adjusted CI** with zero speed
wallaby	468	(336.24, 1921.36)	(373.26, 2733.97)	(314.68, 1819.40)	(381.17, 2669.95)	(251.89, 1692.90)	(301.43, 2206.75)
water deer	119	(54.13, 210.55)	(81.57, 500.84)	(54.37, 240.71)	(88.71, 476.04)	(50.31, 172.75)	(69.90, 345.39)
mara	68	(1.33, 12.13)	(1.96, 22.64)	(1.40, 12.74)	(1.84, 33.19)	(1.97, 6.69)	(2.71, 10.33)
muntjac	13	(4.14, 13.78)	(-)	(4.29, 14.83)	(-)	(5.19, 11.06)	(-)

3.9 Discussion

In this Chapter we developed an integrated Random Encounter Model (iREM), which simultaneously models the encounter data and animal speed data, to estimate density. In particular, we developed a Poisson iREM, where the Poisson model is the underlying distribution in REM and alternatively other iREMs that allow for zero-inflation and overdispersion; these include an integrated NB iREM, a ZIP iREM, and a ZINB iREM.

We found that a Poisson iREM works well in estimating the density. The simulation results indicated that precise estimates can be obtained for small and large sample sizes and therefore, a Poisson iREM can be used in ecological studies with limited resources. The simulation results also revealed how important it is to account for overdispersion and zero-inflation in the data. We have shown how disregarding overdispersion and zero-inflation can induce bias in the density estimator and an underestimation of the standard error of the density.

We have also demonstrated that modelling the speed data and accounting for its variability is relevant in correctly estimating the standard error of the density. We found that REM correctly estimates the density but the integrated Poisson REM is more appropriate than REM since the REM does not account for the variability in speed. However, when the bootstrap method is used to estimate the standard error of the density comparable estimates are obtained from REM and iREM. We have also shown the importance of accounting for the observed zero speed of movement. However, we would recommend, if possible, to sample more speed observations when there are observed zero speeds, since including the zero speed observations in the estimation process would inflate the density and larger sample sizes would allow for more accurate estimates and better precision.

The use of integrated population modelling for density estimation in REM opens the door to a number of different model developments, including the extension of iREM described to account for animals moving in pairs or family groups, animals with observed zero speeds of movement, and the variation in density by habitat, which we investigate next in Chapter 4. It is also interesting to note the parallelism of the different iREMs

developed here with integrated population models used in other applications. Cole and McCrea (2016) and Besbeas et al. (2002) for example used integrated population modelling in discrete state-space models to estimate parameters that are redundant and or are otherwise inestimable, accounting for the variability in other related parameters of the same species while obtaining more precise estimates of the relevant parameters.

While the underlying model in REM is a Poisson distribution, our analysis of the WWAP data set provides a nice illustration of how accounting for overdipsersion and zero-inflation can change the conclusions regarding the underlying encounters distribution. Hutchinson and Waser (2007) have shown that the deviation from the straight-line pattern of movements by animals has no effect on the expected encounter rate, and therefore, a more variable distribution from the underlying Poisson model can be used for encounters. When we analysed the data using alternative models such as a NB iREM, a ZIP iREM and a ZINB iREM, there was strong support for models that allow for overdispersion, in particular for species with larger data sets such as the wallaby and water deer. For species with fewer encounters such as the mara and muntjac, there was support for a Poisson iREM.

We also found that the estimates of the density from the different iREM are not close to the density from the census. This is because the density from the census is the mean density across habitats, and in this Chapter we have not considered habitat in the models. So some of the variation in the estimates might be explained by spatial covariates such as habitat. In the next Chapter we incorporated covariates such as habitat type in iREM.

In estimating average group size using the joint likelihood developed in Section 3.2 we assumed group sizes are small and are detected within the detection zone. But as explained in Chapter 2, estimating group sizes can be problematic, particularly with the use of camera traps. But if groups are small and counts of the number of individuals in each observed group can be made then an estimate of average group size can be made and estimation of the density becomes straightforward. However, the groups must be

considered to be the object of interest. But given that the probability of detection depends on both the distance from the camera and the group size (Buckland et al., 2001), all animals may not be detected, particularly for large groups. For camera traps the area of detection increases with increasing distances, so more animals would be detected at some distance away from the camera. Another complication is that distances should be measured from the camera to the centre of the group, and since distance between the camera and group centre is unobservable, it therefore becomes impossible to quantify the detection zone, particularly for larger, more dispersed groups. One solution could be to treat each individual in the group as a detection but as pointed out by Buckland et al. (2001) the independence assumption would be severely violated, compromising the analytic variance estimates and model selection procedures. While the difficulty of obtaining analytic variance estimates can be overcome by using robust methods for variance estimation, the issue of model selection remains. To overcome the model selection problem Buckland et al. (2001) suggest, in distance sampling, selecting a model taking groups as the sampling unit, then refitting the model to the data with individuals as the sampling unit. However, as far we are aware, there are no methods available for estimating average group size in camera trapping, particularly for large groups, given that the distance from the camera to the centre of the group is unobservable.

Chapter 4

iREM With Covariates

The underlying model for REM, discussed in Chapter 2, and iREM discussed in Chapter 3, is the "ideal gas" model. The ideal gas model assumes that molecules are distributed randomly, and move with uniform velocity (speed) in straight lines. Movement is also assumed to be independent and equally in all directions (see Maxwell, 1860), and the space which these animals occupy is large (Hutchinson and Waser, 2007). By contrast, however, animals respond to each other and their physical environment. They may spend more time in some habitats than others, deviate frequently from straight line patterns of movement, and are often restricted to partially overlapping home ranges (Hutchinson and Waser, 2007, pages 336-337).

In this Chapter we provide a framework for incorporating covariates in iREM. Since abundance can vary with habitat, we incorporate habitat-specific covariates in iREM to assess the potential of this relationship. In Chapter 2, Section 2.3.4, we provided a framework for incorporating habitat in REM, using a fixed value of animal speed to estimate density. Because REM lacks the potential to account for the variability in animal speed, thereby underestimating the standard error of the density, we address this problem by using an integrated REM to assess the potential relationship between abundance and habitat type. In this Chapter we also consider whether there are camera random effects. Placement strategies of camera traps are crucial in REM. Typical placement strategies involving baiting and luring, or on trails, cannot be used to obtain unbiased estimates in REM as these violate the underlying assumptions of REM. REM requires that placement strategies of camera traps be randomized. It is worth noting

that camera traps placed at random could end up in areas where animals often are, whereas other camera traps could end up where animals do not frequent. Therefore, the placement strategy of camera traps could have some effect on abundance estimation.

The motivation for this Chapter is the data set at Whipsnade Wild Animal Park (WWAP), which is discussed in Chapter 2, Section 2.7. The number of species was expected to vary across WWAP, given that the park has a diverse topography. The park can be divided readily into four areas of contrasting habitats: two areas of open grassland with scattered scrub (the Institute Paddock with gentle slopes, and with size 0.28 km^2, and the Downs with a steep scarp slope, and with size 0.49 km^2), one area of rough grassland and thicket on largely level but highly broken ground (the Old Farm, with size 0.23 km^2) and one area of mixed lawns, roads, buildings and enclosures housing large animals with scattered trees (Central Park with size 1.26 km^2). The Wildlife Park houses several free-ranging species, which are enclosed within an area of 226 ha by an outer perimeter fence but which otherwise move freely throughout the park. Each free-ranging species uses these areas differentially in a predictable way, giving rise to contrasting densities of each across the park (see Rowcliffe et al., 2008). For the purpose of the analyses we focused on the four species given in Chapters 2, and 3: wallaby, water deer, mara and mutjac.

The Chapter begins, in Section 4.1, with the proposed model to incorporate covariates in iREM, for which habitat is a motivating example. We test this model via simulations in Section 4.2. In Section 4.3 we illustrate the application of this model by fitting it to the WWAP data set for four species: wallaby, water deer, mara and muntjac. At WWAP Rowcliffe et al. (2008) adopted a randomized approach for camera placement strategies. In Section 4.4 we propose a model for camera random effect and we investigate camera random effect on estimated density via simulations. An application of the model with random effects to WWAP is given in Section 4.5. It is expected that some of the variation in abundance is likely to be explained by both habitat type and camera random effects, therefore we incorporate both habitat type and random effects in iREM to assess this relationship. We illustrate the application of this model by fitting it to the WWAP data set in Section 4.6.

4.1 Introducing covariates in iREM

In the ideal gas model, the assumption of large abundance of animals in random motion, travelling in straight lines between encounters, and occupying very large spaces (Hutchinson and Waser, 2007, pages 336-337) is likely to be violated. Quite often this quantity varies with diverse factors, some of which may be recorded in the field. The iREM model structure discussed so far can be modified to incorporate covariates to describe how abundance changes with these factors. Animal abundance (density) can vary with the characteristics of the space (or site) which they occupy, and the behaviour of animals. Some of these characteristics are habitat type or climatic conditions. In fact often the primary objective of the study is to assess these potential relationships. Using the log link function, the encounter rate in REM can be expressed as a function of covariates

$$\boldsymbol{\lambda} = \left(\frac{2+\theta}{\pi}\mu_x rt\right)\exp(\mathbf{X}\boldsymbol{\beta}), \tag{4.1.1}$$

where $\boldsymbol{\lambda}$ $(= \lambda_{n\times 1})$, and n is the number of camera trap days, is a vector of expected encounter rates for each camera day, and

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & x_{13} & \dots & x_{1H} \\ 1 & x_{21} & x_{22} & x_{23} & \dots & x_{2H} \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_{n1} & x_{n2} & x_{n3} & \dots & x_{nH} \end{bmatrix}_{n\times H}$$

is a matrix of H covariates. The vector $\boldsymbol{\beta}_{1\times H} = (\beta_1, \beta_2, ..., \beta_H)^T$ contains the regression coefficients. In equation (4.1.1) the camera dimensions are r and θ; the expected speed is μ_x, and $t = 1$ day.

4.1.1 The Model

Suppose the encounter observations, y_{ij}, has probability mass function $h(y_{ij} \mid \boldsymbol{\lambda},\ \tau)$ where $\boldsymbol{\lambda}$ is a vector of encounter rates defined in equation (4.1.1), τ represents any additional parameters in the model, and where $i = 1, 2, ..., c$ is the ith camera trap and $j = 1, 2, ..., n$ is the jth camera trap day. We assume independence of y_{ij} for all i and all j. And, suppose m independent speed observations, $x = \{x_1, ..., x_m\}$, have probability

density function $f(x_l \mid \mu_x,\ \nu)$ where $l = 1, 2, ..., m$, μ_x is the expected animal speed, and ν represents any additional parameters in the model. Therefore, the density and expected animal speed can be estimated by maximizing the joint likelihood function

$$\ell = \sum_{i=1}^{c} \sum_{j=1}^{n} \log\ h(y_{ij} \mid \boldsymbol{\lambda},\ \tau) + \sum_{l=1}^{m} \log\ f(x_l \mid \mu_x,\ \nu). \tag{4.1.2}$$

In this Chapter habitat type is a motivating example, and under the assumption that speed of movement is constant across habitats we provide the integrated likelihood function, which incorporates habitat-specific covariates.

4.1.2 iREM with habitat

In this section we provide an iREM with habitat-specific covariates. The Whipsnade Wildlife Animal Park was divided into four areas of contrasting habitats: Central Park, Downs, Institute Paddock, and Old Farm, which we denote as $p = 1, 2, 3, 4$, respectively. Rowcliffe et al. (2008) estimated the density within each habitat by fitting REM separately to the data set in each habitat. In Chapter 2, Section 2.3.4, we provided a framework to incorporate habitat-specific covariates in REM, modelling the encounter data per habitat simultaneously but using a fixed value for animal speed. Here we incorporate habitat-specific covariates in iREM. The covariates are then indicator variables representing each habitat. We arbitrarily set habitat 1 to be the null habitat and let

$$x_{jp} = \begin{cases} 1 & \text{observation } j \text{ is from habitat } p \\ 0 & \text{observation } j \text{ is not from habitat } p, \end{cases}$$

so that for other habitats

$$\mathbf{X} = \begin{bmatrix} 1 & x_{12} & x_{13} & x_{14} \\ 1 & x_{22} & x_{23} & x_{24} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n2} & x_{n3} & x_{n4} \end{bmatrix}_{\mathbf{n \times 4}}$$

The vector $\boldsymbol{\beta}_{1\times 4} = (\beta_1,\ \beta_2,\ \beta_3,\ \beta_4)^T$ contains the regression coefficients, and the density in each habitat is computed as follows

$$D_p = \begin{cases} \exp(\beta_1) & \text{for } p = 1. \\ \exp(\beta_1 + \beta_p) & \text{for } p = 2, 3, 4. \end{cases}$$

The mean density for a given species, D_T can be computed as

$$D_T = \sum_{p=1}^{4} \frac{A_p \exp(\beta_p)}{A_T} = \frac{\sum_{p=1}^{4} A_p D_p}{A_T}$$

where A_p $(p = 1, 2, 3, 4)$ is the area of habitat p in the surveyed area; and $A_T = A_1 + A_2 + A_3 + A_4$ is the total area. For known variance of D_p the variance of D_T can be computed directly as

$$\begin{aligned} \text{Var}(D_T) &= \text{Var}\left(\frac{\sum_{p=1}^{4} A_p D_p}{A_T}\right) \\ &= \sum_{p=1}^{4} \frac{A_p^2 \text{Var}(D_p)}{A_T^2}. \end{aligned}$$

The distributions used for the count data and animal speed data in Chapter 3 are used here. For the count data we use 1) a Poisson REM, 2) a negative binomial REM (NB REM), 3) a zero-inflated negative binomial REM (ZINB REM), and 4) a zero-inflated Poisson REM (ZIP REM). And, for the animal speed data we use 1) a gamma model, 2) a lognormal model, 3) a Weibull model, and 4) a zero-adjusted gamma model (ZAGA).

4.1.3 iREM with covariates and group density

In this section we provide a framework for incorporating habitat-specific covariates in iREM for animals moving in family groups. As described in Chapter 2, Section 2.3.2, some animals were found to be moving in family groups at WWAP, and to estimate the density of the individual animal Rowcliffe et al. (2008) multiplied the density of the group by an unbiased independent estimate of average group size, g, to give

$$\hat{D} = \hat{D}_G \times \hat{g}, \tag{4.1.3}$$

where $\hat{D}_G$ is the density of the group. In Chapter 3, Section 3.2 we have shown how we can estimate the density of the individual animal if they are found to be moving

in pairs of family groups. We assume group sizes are small and all individuals within the groups are detected within the detection zone. We extended iREM to incorporate data for animals moving in groups. In this Chapter we extend the model in (4.1.2) to include group data and estimate density within and across habitats for animals moving in groups with the same assumptions made in Chapter 3.

Suppose the group data is assumed to follow some discrete distribution whose support is the set of positive integers. The data for group size is s_j, where $j = 1, 2, ..., S$ is the number of groups. Suppose the group size data has probability mass function $k(s_j \mid \phi)$, where ϕ is the parameter that represents the mean of the untruncated discrete distribution, which the group data is assumed to follow. Then the joint likelihood becomes

$$\ell = \sum_{i=1}^{c}\sum_{j=1}^{n} \log\, h(y_{ij} \mid \boldsymbol{\lambda},\ \tau) + \sum_{l=1}^{m} \log\, f(x_l \mid \mu_x,\ \nu) + \sum_{j=1}^{S} \log\, k(s_j \mid \phi), \qquad (4.1.4)$$

where $\boldsymbol{\lambda}$ is a vector of expected encounter rates defined in equation (4.1.1) above. The distribution, $k(\cdot)$, is a zero-truncated Poisson (ZTP) used to model the group size data, which is discussed in Chapter 3.

4.2 Simulation study

This simulation study explores the performance of the models by looking at the seven following cases:

(i) Firstly, we verify that a Poisson iREM with habitat can be used in realistic settings by fitting it to encounter data simulated from a Poisson REM (Section 4.2.1).

(ii) Secondly, we explore how important it is to account for variation in encounter data across habitats. To do this we fit a Poisson iREM with habitat to data simulated from a NB REM. We also fit a NB iREM to these encounters under different parameter settings (Section 4.2.2).

(iii) Thirdly, we explore how important it is to account for zero-inflation in the encounter data across habitats. To do this we fit a Poisson iREM with habitat to

encounters simulated from a ZIP REM, and we compare the results with a ZIP iREM with habitat fitted to the same encounter data (Section 4.2.3).

(iv) Fourthly, we explore how important it is to account for overdispersion and zero-inflation in encounter data. To do this we fit a Poisson iREM with habitat to encounters simulated from a ZINB REM. We also fit a NB iREM with habitat and a ZINB iREM with habitat to the same encounter data (Section 4.2.4).

(v) Fifthly, we verify that an iREM with habitat can be used in realistic settings for animals moving in pairs or family. We fit a Poisson iREM with habitat to encounters simulated from a Poisson REM, and we fit a zero-truncated Poisson (ZTP) to group data simulated from the relevant fitted model (Section 4.2.5).

In all cases we fit a gamma model, a lognormal model or a Weibull model to animal speed data simulated from the relevant fitted models.

(vi) Sixthly, we investigate the relevance of incorporating habitat-specific covariates into iREM. To do this we simulate data with habitat covariates and fit an iREM with habitat and an iREM without habitat. For illustration we fit a Poisson iREM with habitat and a Poisson iREM to encounters simulated from a Poisson REM and we assume animal speed follows a lognormal model (Section 4.2.6).

(vii) Seventhly, we verify that an iREM with habitat, which accounts for observed zeros in speed of movement, can be used in realistic settings. To do this we fit a Poisson iREM with habitat to encounters simulated from a Poisson REM and we fit a zero-adjusted gamma model (ZAGA) to speed data simulated from the relevant fitted model (Section 4.2.7).

For the simulations we generate scenarios in which the true mean encounters in each habitat are based upon reasonable values that might be applicable for data for real species in our motivating WWAP data set. The average parameter estimates (standard error in parentheses), Standard deviation (Sd) and Root Mean Square Error are computed from 200 simulation runs. The parameters used to estimate animal density: $r = 0.012$ (km) and $\theta = 0.175$ (radians), and camera trap time $t = 1$ (day) are held fixed. Here r and θ are detection zone dimensions of the camera. In the simulations the

models were fitted using the optim function in the R software package (version 3.3.2 or earlier) using the default Nelder-Mead algorithm, except where otherwise stated. In the simulation study we used the true parameter values and sample sizes given in Table 4.2.1.

Table 4.2.1: Sample sizes and true parameter values used in the simulation process. The mean encounters in each habitat, and the sample sizes are given in the first section. The regression coefficients are $\boldsymbol{\beta} = \{\beta_1, \beta_2, \beta_3, \beta_4\}$. The densities within habitats, D_p, are derived parameters from β_p. Here β_1 is the null habitat.

		Regression coefficients				
Sample sizes		Habitat 1	Habitat 2	Habitat 3	Habitat 4	Expected speed
n	m	β_1	β_2	β_3	β_4	μ_x
40	40	4.62	0.31	3.13	0.30	0.71
100	100	4.62	0.31	3.13	0.30	0.71
40	40	3.56	1.61	0.18	0.40	0.71
100	100	3.56	1.61	0.18	0.40	0.71
		Density within habitats				
		D_1	D_2	D_3	D_4	
		101.49	138.38	2321.57	137.00	
		35.16	175.91	42.10	52.46	

4.2.1 Investigating the performance of a Poisson iREM with habitat-specific covariates

The simulation results from fitting a Poisson iREM with habitat and from fitting a lognormal model to animal speed data are given in Table 4.2.2. The simulation results show that the regression estimators are reasonably accurate with slight bias under small sample sizes, and as expected the bias is minimal with larger sample sizes. The parameter estimates from the alternative models used to fit the speed data show minimal differences. The results from the gamma model (Table C.1.1) and Weibull model (Table C.1.2) are given in appendix C.1.1.

Table 4.2.2: Average parameter estimates from a Poisson iREM with habitat fitted to encounters simulated from a Poisson REM (standard error in parentheses). The lognormal model is fitted to animal speed data simulated from the relevant fitted model. The sample sizes are n trap days, and m animal speed data. The Standard deviation (Sd) and Root Mean Square Error (RMSE) are also given.

	Parameter estimates				
	Habitat 1 $\hat{\beta}_1$	Habitat 2 $\hat{\beta}_2$	Habitat 3 $\hat{\beta}_3$	Habitat 4 $\hat{\beta}_4$	$\hat{\mu}_x$
$\beta_1 = 4.62$, $\beta_2 = 0.31$, $\beta_3 = 3.13$, $\beta_4 = 0.30$; $\mu_x = 0.71$					
$n = m = 40$					
Estimate	4.60 (0.23)	0.32 (0.27)	3.16 (0.21)	0.34 (0.27)	0.71 (0.06)
Sd	0.21	0.27	0.21	0.26	0.06
RMSE	0.22	0.27	0.21	0.26	0.06
$n = m = 100$					
Estimate	4.60 (0.14)	0.32 (0.17)	3.14 (0.13)	0.29 (0.17)	0.71 (0.04)
Sd	0.14	0.17	0.13	0.17	0.04
RMSE	0.14	0.17	0.13	0.17	0.04
$\beta_1 = 3.56$, $\beta_2 = 1.61$, $\beta_3 = 0.18$, $\beta_4 = 0.40$, and $\mu_x = 0.71$					
$n = m = 40$					
Estimate	3.50 (0.32)	1.65 (0.40)	0.18 (0.46)	0.40 (0.47)	0.71 (0.07)
Sd	0.32	0.34	0.46	0.40	0.07
RMSE	0.33	0.34	0.46	0.40	0.07
$n = m = 100$					
Estimate	3.56 (0.23)	1.61 (0.23)	0.15 (0.30)	0.40 (0.29)	0.71 (0.06)
Sd	0.23	0.23	0.31	0.29	0.04
RMSE	0.23	0.23	0.31	0.29	0.04

4.2.2 *Investigating the importance of accounting for variation in encounter data*

To simulate data from the NB REM we set $\boldsymbol{\beta} = \{4.62, 0.31, 3.13, 0.30\}$; $\kappa = 0.56$; and $\mu_x = 0.71$ (km/day^{-1}); the number of camera trapping days, $n = 40$ or $n = 100$; and the number of animal speed observations, $m = 40$ or $m = 100$. Here we give the results from fitting a lognormal model to animal speed in Table 4.2.3. The estimates from a gamma model (Table C.1.3) and a Weibull model (Table C.1.4), which show similar estimates as a lognormal model are given in appendix C.1.2. The simulation results (Table 4.2.3) reveal that while the parameters are well estimated, ignoring variation in encounter data can introduce an underestimation of the standard error of the density. The standard errors from a Poisson iREM with habitat are consistently smaller, and are larger from a NB iREM with habitat, which may be anticipated as a consequence of accounting for overdispersion. As expected, increasing the sample size improves estimation and precision.

Table 4.2.3: Average parameter estimates from fitting a Poisson iREM with habitat and a NB iREM with habitat to encounters simulated from a NB REM (standard error in parentheses). Animal speed is assumed to follow a lognormal model. The parameters are set to $\beta_1 = 4.62$, $\beta_2 = 0.31$, $\beta_3 = 3.13$, $\beta_4 = 0.30$; $\mu_x = 0.71$ (km/day^{-1}); and $\kappa = 0.56$. The sample sizes are n trap days, and m animal speed data. The Standard deviation (Sd) and Root Mean Square Error (RMSE) are also given.

	Parameter estimates					
	Habitat 1	Habitat 2	Habitat 3	Habitat 4		
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\mu}_x$	κ
$n = m = 40$						
Poisson iREM	4.59 (0.23)	0.31 (0.28)	3.13 (0.22)	0.31 (0.28)	0.71 (0.06)	-
Sd	0.32	0.42	0.38	0.38	0.07	-
RMSE	0.33	0.42	0.38	0.38	0.07	-
NB iREM	4.59 (0.31)	0.32 (0.42)	3.13 (0.37)	0.31 (0.41)	0.71 (0.11)	0.59 (0.11)
Sd	0.33	0.42	0.39	0.39	0.07	0.07
RMSE	0.33	0.42	0.39	0.39	0.07	0.07
$n = m = 100$						
Poisson iREM	4.60 (0.14)	0.32 (0.17)	3.16 (0.13)	0.31 (0.17)	0.71 (0.04)	-
Sd	0.18	0.24	0.21	0.25	0.04	
RMSE	0.18	0.24	0.21	0.25	0.04	
NB iREM	4.59 (0.20)	0.31 (0.26)	3.16 (0.23)	0.31 (0.26)	0.71 (0.07)	0.57 (0.07)
Sd	0.18	0.25	0.21	0.26	0.04	0.07
RMSE	0.18	0.25	0.21	0.26	0.04	0.07

4.2.3 *Investigating the importance of accounting for zero-inflation*

To investigate the importance of accounting for zero-inflation in the encounter data we fit a Poisson iREM with habitat to data simulated from a ZIP REM, and compare the results from fitting a ZIP iREM to the same encounters. We set $\beta_1 = 4.62,\ \beta_2 = 0.31,\ \beta_3 = 3.13,\ \beta_4 = 0.30;\ \mu_x = 0.71$ (km/day^{-1}) and $\rho = 0.35$, respectively. We give the results from a lognormal model in Table 4.2.4, and the results from a gamma model (Table C.1.5) and a Weibull model (Table C.1.6), which gave similar estimates as the lognormal model are given in appendix C.1.3. The simulation results (Table 4.2.4) illustrate how induced-zero-inflation in the data can introduce bias in the regression estimators from a Poisson iREM with habitat, in particular in Habitat 1 which is the null habitat, and hence, bias in the mean encounter in other habitats. Increasing the sample sizes does little in improving estimations in Habitat 1. Therefore, we recommend a ZIP iREM where there is zero-inflation in the encounter data.

Table 4.2.4: Average parameter estimates from fitting a Poisson iREM with habitat and a ZIP iREM with habitat to encounters simulated from a ZIP REM (standard error in parentheses). Animal speed is assumed to follow a lognormal model. The parameters are set to $\beta_1 = 4.62$, $\beta_2 = 0.31$, $\beta_3 = 3.13$, $\beta_4 = 0.30$; $\mu_x = 0.71$ (km/day^{-1}); $\rho = 0.35$. The sample sizes are n trap days, and m animal speed data. The Standard deviation (Sd) and Root Mean Square Error (RMSE) are also given.

	Parameter estimates					
	Habitat 1 $\hat{\beta}_1$	Habitat 2 $\hat{\beta}_2$	Habitat 3 $\hat{\beta}_3$	Habitat 4 $\hat{\beta}_4$	$\hat{\mu}_x$	$\hat{\rho}$
$n = m = 40$						
Poisson iREM	4.14 (0.28)	0.35 (0.35)	3.18 (0.27)	0.33 (0.27)	0.71 (0.06)	-
Sd	0.30	0.40	0.32	0.41	0.07	-
RMSE	0.57	0.40	0.33	0.41	0.07	-
ZIP iREM	4.58 (0.31)	0.33 (0.38)	3.18 (0.31)	0.32 (0.38)	0.70 (0.06)	0.35 (0.10)
Sd	0.33	0.41	0.33	0.41	0.07	0.07
RMSE	0.33	0.41	0.33	0.41	0.07	0.07
$n = m = 100$						
Poisson iREM	4.17 (0.17)	0.33 (0.21)	3.15 (0.17)	0.30 (0.22)	0.71 (0.04)	-
Sd	0.17	0.23	0.18	0.25	0.04	
RMSE	0.48	0.23	0.18	0.25	0.04	
ZIP iREM	4.62 (0.19)	0.31 (0.24)	3.14 (0.19)	0.29 (0.24)	0.71 (0.04)	0.35 (0.06)
Sd	0.18	0.23	0.18	0.23	0.04	0.04
RMSE	0.18	0.23	0.18	0.23	0.04	0.04

4.2.4 *Investigating the importance of accounting for zero-inflation and variation in encounter data*

To assess the impact that zero-inflation and extra variation can have on the estimation of animal density we fit a Poisson iREM with habitat to encounters simulated from a ZINB REM with parameters $\beta_1 = 4.62,\ \beta_2 = 0.31,\ \beta_3 = 3.13,\ \beta_4 = 0.30;\ \ \rho = 0.35,$ $\mu_x = 0.71$ (km/day^{-1}) and $\kappa = 0.56$; and the estimates are compared with estimates from fitting a NB iREM with habitat and a ZINB iREM with habitat to the same data. The simulation results are given in appendix C.1.4.

The results from fitting a gamma model to animal speed data are given in Table C.1.7 and Table C.1.8 for small and large sample sizes, respectively. Table C.1.9 and Table C.1.10 give the results from fitting a lognormal model to speed data for small and large sample sizes, respectively. Table C.1.11 and Table C.1.12 give the results from fitting a Weibull model to the speed data for small and large sample sizes, respectively. The parameter estimates from the alternative speed data models show minimal differences, but the simulation results reveal that not allowing for zero-inflation and variation in encounter data can introduce strong bias and underestimation of the standard errors of the regression estimators from a Poisson iREM with habitat. For the NB iREM with habitat, bias is introduced but the estimated standard errors are close to the estimated standard errors from a ZINB iREM with habitat. As expected, increasing sampling effort improves estimation and its precision.

4.2.5 *Performance of the Poisson iREM with habitat and group size*

Some animals generally move around in pairs or family groups, and using capture records of groups in REM would give rise to the density of the group instead of the individual. To estimate the density of the individual and the group size we simulate scenarios with the integrated likelihood specified above in equation (4.1.4). The parameters are set to $\beta_1 = 4.62,\ \beta_2 = 0.31,\ \beta_3 = 3.13,\ \beta_4 = 0.30$; $g = 2.50$; $\mu_x = 0.71$; and group size $S = 10$ or $S = 50$. We fit a zero-truncated Poisson (ZTP) model to the group size data, and for illustration we fit a Poisson iREM. The results show minimal differences between the three speed data models, so for illustration we give estimates from the gamma model

in Table 4.2.5, while estimates from the lognormal model (Table C.1.13) and estimates from the Weibull model (Table C.1.14) are given in appendix C.1.5.

Table 4.2.5 gives the parameter estimates for animals moving in groups within habitats. The results show some bias in the parameter estimates but with large standard errors the true values are captured within an approximate 95% confidence interval. It was observed at WWAP that the animals moving in family groups had higher expected speeds of movement. For illustration, we investigate the performance of the Poisson iREM with habitat using a higher expected animal speed, $\mu_x = 4.60$, and the same regression parameters and average group size with a smaller expected speed. Animal speed is assumed to follow a gamma model. Table 4.2.6 gives the results for the regression estimators in habitats and group size. The simulation results show that the parameter estimates and precision improve for animals with higher expected speeds, and estimated average group size is closer to the true value.

Table 4.2.5: Average parameter estimates from fitting a Poisson iREM to encounter data simulated from a Poisson REM (standard errors are in parentheses). The parameters are set to $\beta_1 = 4.62$, $\beta_2 = 0.31$, $\beta_3 = 3.13$, $\beta_4 = 0.30$; $\mu_x = 0.71$ (km/day^{-1}); and $g = 2.50$. The sample sizes are n trap days, m animal speed observations, and group size, $S = 10$ or $S = 50$. The Standard deviation (Sd) and Root Mean Square Error (RMSE) are also given.

	Estimated Parameters					
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\mu}_x$	$\hat{g}$
$n = m = 40$; $S = 10$						
Estimate	4.64 (0.18)	0.39 (0.23)	3.20 (0.13)	0.33 (0.23)	0.69 (0.12)	2.69 (0.59)
Sd	0.08	0.20	0.09	0.20	0.05	0.17
RMSE	0.08	0.22	0.11	0.20	0.05	0.25
$n = m = 100$; $S = 10$						
Estimate	4.62 (0.11)	0.37 (0.15)	3.19 (0.09)	0.33 (0.15)	0.69 (0.07)	2.58 (0.40)
Sd	0.05	0.16	0.06	0.17	0.04	0.10
RMSE	0.05	0.17	0.08	0.17	0.04	0.13
$n = m = 40$; $S = 50$						
Estimate	4.68 (0.15)	0.35 (0.23)	3.23 (0.12)	0.30 (0.23)	0.72 (0.12)	3.01 (0.24)
Sd	0.11	0.23	0.12	0.22	0.07	0.22
RMSE	0.13	0.23	0.15	0.22	0.07	0.56
$n = m = 100$; $S = 50$						
Estimate	4.64 (0.10)	0.35 (0.14)	3.22 (0.08)	0.33 (0.14)	0.69 (0.07)	2.74 (0.26)
Sd	0.06	0.16	0.07	0.17	0.04	0.13
RMSE	0.06	0.16	0.11	0.18	0.05	0.27

Table 4.2.6: Average parameter estimates from fitting a Poisson iREM to encounter data simulated from a Poisson REM (standard errors are in parentheses). The parameters are set to $\beta_1 = 4.62$, $\beta_2 = 0.31$, $\beta_3 = 3.13$, $\beta_4 = 0.30$; $\mu_x = 4.60$ (km/day^{-1}); and $g = 2.50$. The sample sizes are n trap days, m animal speed observations, and group size, $S = 10$ or $S = 50$.

	Estimated Parameters					
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\mu}_x$	$\hat{g}$
$n = m = 40$; $S = 10$						
Estimate	4.61 (0.05)	0.36 (0.09)	3.17 (0.05)	0.34 (0.09)	4.56 (0.14)	2.55 (0.14)
Sd	0.03	0.10	0.04	0.09	0.18	0.07
RMSE	0.03	0.12	0.06	0.10	0.19	0.09
$n = m = 100$; $S = 10$						
Estimate	4.61 (0.02)	0.33 (0.05)	3.15 (0.03)	0.32 (0.05)	4.58 (0.12)	2.52 (0.04)
Sd	0.02	0.06	0.03	0.06	0.11	0.04
RMSE	0.02	0.06	0.04	0.06	0.11	0.05
$n = m = 40$; $S = 50$						
Estimate	4.62 (0.07)	0.35 (0.09)	3.17 (0.05)	0.31 (0.09)	4.56 (0.14)	2.60 (0.26)
Sd	0.04	0.11	0.04	0.12	0.19	0.10
RMSE	0.04	0.12	0.06	0.12	0.20	0.13
$n = m = 100$; $S = 50$						
Estimate	4.62 (0.04)	0.31 (0.06)	3.13 (0.04)	0.30 (0.06)	4.59 (0.02)	2.51 (0.06)
Sd	0.01	0.03	0.02	0.04	0.05	0.03
RMSE	0.01	0.03	0.02	0.04	0.05	0.03

4.2.6 Investigating the importance of incorporating habitat-specific covariates

Here we investigate the importance of incorporating habitat-specific covariates in iREM when estimating density. We used the true values in Table 4.2.1 in the simulation process. We set the values for the four areas (in km^2) of the habitats to $A_1 = 1.26$; $A_2 = 0.49$; $A_3 = 0.28$; $A_4 = 0.23$ and total area $A_T = 2.26$. For illustration we fit a Poisson iREM with habitat where animal speed is assumed to follow a lognormal model. Table C.1.15 in appendix C.1.6 gives the simulation results for the parameters $\mu_x = 0.71$; $\beta_1 = 3.56$, $\beta_2 = 1.61$, $\beta_3 = 0.18$, $\beta_4 = 0.40$. Table 4.2.7 gives the simulation results for $\mu_x = 0.71$; $\beta_1 = 4.62$, $\beta_2 = 0.31$, $\beta_3 = 3.13$, $\beta_4 = 0.30$. The simulation results (Table C.1.15; Table 4.2.7) reveal that, under all scenarios, if there are habitat-specific covariates, it is relevant to incorporate these in iREM when estimating density. The results show that a large positive bias can be introduced if habitats are ignored when estimating density. The sign of this bias depends on the specifics of the relative areas and density. For example, if less dense areas are oversampled the bias will be negative.

Table 4.2.7: Average parameter estimates from a Poisson iREM with habitat fitted to encounters simulated from a Poisson REM (standard error in parentheses). Animal speed data is assumed to follow a lognormal model. The expected animal speed, regression coefficients, and hence density within habitats and across habitats are set to $\mu_x = 0.71$; $\beta_1 = 4.62$, $\beta_2 = 0.31$, $\beta_3 = 3.13$, $\beta_4 = 0.30$; $D_1 = 101.49$, $D_2 = 138.38$, $D_3 = 2321.57$, $D_4 = 137.00$; $D_T = 388.16$, respectively. The sample sizes are n trap days, and m animal speed data. The Standard deviation (Sd) and Root Mean Square Error (RMSE) are also given.

	iREM with habitat						iREM	
	Habitat 1 $\hat{\beta}_1$	Habitat 2 $\hat{\beta}_2$	Habitat 3 $\hat{\beta}_3$	Habitat 4 $\hat{\beta}_4$	$\hat{\mu}_x$	$\hat{D_T}$	$\hat{\mu}_x$	$\hat{D}$
Estimates of β_p								
$n = m = 40$								
Estimate	4.62 (0.23)	0.31 (0.27)	3.14 (0.21)	0.30 (0.27)	-	-	-	-
Sd	0.24	0.28	0.23	0.31	-	-	-	-
RMSE	0.24	0.28	0.23	0.31	-	-	-	-
$n = m = 100$								
Estimate	4.62 (0.15)	0.30 (0.19)	3.14 (0.13)	0.30 (0.17)	-	-	-	-
Sd	0.16	0.19	0.14	0.18	-	-	-	-
RMSE	0.16	0.19	0.14	0.18	-	-	-	-
Estimates of D_p and μ_x								
$n = m = 40$								
Estimate	101.55 (22.86)	138.24 (68.93)	2351.38 (1027.32)	136.45 (68.16)	0.70 (0.07)	391.80 (38.54)	0.71 (0.06)	681.54 (66.86)
Sd	23.75	30.07	267.09	28.31	0.07	44.44	0.07	82.82
RMSE	23.92	30.22	270.78	28.40	0.07	44.59	0.07	304.85
$n = m = 100$								
Estimate	101.14 (14.33)	136.87 (42.87)	2327.02 (638.64)	136.10 (42.65)	0.70 (0.07)	388.21 (22.56)	0.71 (0.06)	676.69 (42.90)
Sd	15.60	18.59	161.07	17.44	0.04	27.50	0.07	57.21
RMSE	15.62	18.60	161.44	17.44	0.04	27.50	0.07	294.27

4.2.7 *Investigating the performance of the Poisson iREM for animals with observed zero speed*

During the WWAP survey it was observed that the animals did not move about much, except for the muntjac species which was considerably mobile. And given that the animals were followed around recording their travelled distance to derive an estimate of the speed, in some cases a record of zero speed was observed. These are genuine zeros, which can have an effect on estimated density. As shown in Section 3.7.6 including the observed zeros would have an effect on the density.

For illustration, we investigate the performance of a Poisson iREM for animals with observed zero speed of movement via simulation from a Poisson REM for the encounter data, and a Zero-adjusted gamma model (ZAGA) for animal speed data. We assess the output for parameters and sample sizes in Table 4.2.1 in Section 4.2 above with probability of zero-response category, $w = 0.25$. The simulation results are given in Table 4.2.8. The simulation results (Table 4.2.8) illustrate how well the Poisson iREM and ZAGA model work in estimating the parameters under small and large sample size conditions. Based on the simulation results in Table 4.2.8, we would recommend a ZAGA model if a design allows observations of immobile animals. But it is worth noting that adjusting for zeros in continuous data is not limited to a ZAGA model. Other useful models include a zero-adjusted Weibull model or a zero-adjusted lognormal model, which are not considered here.

We also examine the fit of the ZAGA model and the gamma model when habitat is included in the model assuming the encounter data follows a Poisson REM. We consider parameter values and sample sizes that reflect those of the WWAP data set. We increase the sample size of the speed observations to illustrate the effect on estimated density. We set $\beta_1 = 4.56$, $\beta_2 = 7.00$, $\beta_3 = 6.63$, $\beta_4 = 6.69$; $\mu_x = 0.71$ (km/day^{-1}); and $w = 0.21$. The sample sizes are $n = 42$ camera trap days, $m = 10$ or $m = 40$ animal speed observations. We generate speed observations from a ZAGA model, and we fit a ZAGA model to the data. We also fit a gamma model to the data, excluding the zeros. The results for these parameters are given Table 4.2.9. The results show that a slight

positive bias is introduced by the ZGA model. For the gamma model, a negative bias is introduced for the parameter estimate of the reference habitat (Habitat 1), but the parameters from the other habitats are estimated with reasonable accuracy. Hence, we expect that the mean density across habitats would be lower compared with the mean density from the ZAGA model. Increasing the sample size improves the estimates and precision from the ZAGA, and precision from the gamma model.

Table 4.2.10 gives the results for the following parameter values and sample sizes: $\beta_1 = 4.29$, $\beta_2 = 3.58$, $\beta_3 = 6.36$, $\beta_4 = 4.28$; $\mu_x = 1.17$ (km/day^{-1}); and $w = 0.40$; and $n = 42$ camera trap days, $m = 10$ or $m = 40$ animal speed observations. These parameter values are associated with smaller density values. The results show that it is relevant to account for the observed zero speeds. A strong negative bias is introduced in the parameter estimate of the reference habitat (Habitat 1), hence, the density within the other habitats and the mean density would be underestimated. Increasing the sample size of the speed observations improves estimation and precision.

Table 4.2.8: Parameter estimates from fitting a Poisson iREM to encounters simulated from a Poisson REM, and a ZAGA model to data simulated from ZAGA model (standard error in parentheses). The sample sizes are n trap days, and m animal speed observations. The Standard deviation (Sd) and Root Mean Square Error (RMSE) are also given.

	Estimated Parameters					
	Habitat 1	Habitat 2	Habitat 3	Habitat 4		
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\mu}_x$	$\hat{w}$
$\beta_1 = 4.62$, $\beta_2 = 0.31$, $\beta_3 = 3.13$, $\beta_4 = 0.30$; $\mu_x = 0.71$; $w = 0.25$						
$n = m = 40$						
Estimate	4.62 (0.10)	0.28 (0.27)	3.13 (0.21)	0.29 (0.27)	0.71 (0.10)	0.26 (0.10)
Sd	0.10	0.29	0.21	0.27	0.07	0.07
RMSE	0.10	0.29	0.21	0.27	0.07	0.07
$n = m = 100$						
Estimate	4.61 (0.07)	0.33 (0.17)	3.15 (0.13)	0.31 (0.17)	0.71 (0.07)	0.26 (0.06)
Sd	0.07	0.17	0.13	0.17	0.05	0.05
RMSE	0.07	0.17	0.13	0.17	0.05	0.05
$\beta_1 = 3.56$, $\beta_2 = 1.61$, $\beta_3 = 0.18$, $\beta_4 = 0.40$, $\mu_x = 0.71$, and $w = 0.25$						
$n = m = 40$						
Estimate	3.55 (0.18)	1.60 (0.39)	0.13 (0.49)	0.39 (0.46)	0.71 (0.18)	0.24 (0.09)
Sd	0.17	0.37	0.45	0.47	0.11	0.07
RMSE	0.17	0.37	0.45	0.47	0.11	0.07
Estimate	3.55 (0.11)	1.63 (0.25)	0.17 (0.30)	0.42 (0.29)	0.70 (0.11)	0.25 (0.06)
Sd	0.11	0.25	0.30	0.30	0.09	0.04
RMSE	0.12	0.25	0.30	0.30	0.09	0.04

Table 4.2.9: Parameter estimates from fitting a Poisson iREM to encounters simulated from a Poisson REM, and a ZAGA model to data simulated from a ZAGA model. We also fit a gamma model to the speed data, excluding the zero observations (standard error in parentheses). The sample sizes are n trap days, and m animal speed observations. The Standard deviation (Sd) and Root Mean Square Error (RMSE) are also given.

	Estimated Parameters					
	Habitat 1	Habitat 2	Habitat 3	Habitat 4		
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\mu}_x$	$\hat{w}$
$\beta_1 = 4.56$, $\beta_2 = 7.00$, $\beta_3 = 6.63$, $\beta_4 = 6.69$; $\mu_x = 0.71$; $w = 0.21$						
$n = 42$; $m = 10$						
ZAGA model						
Estimate	4.57 (0.48)	7.01 (1.46)	6.64 (1.38)	6.70 (1.39)	0.76 (0.08)	0.22 (0.17)
Sd	0.71	0.22	0.22	0.22	0.39	0.13
RMSE	0.71	0.22	0.22	0.22	0.39	0.13
gamma model						
Estimate	4.32 (1.19)	7.01 (1.46)	6.64 (1.38)	6.70 (1.39)	0.91 (0.16)	-
Sd	0.32	0.22	0.22	0.22	0.20	-
RMSE	0.40	0.22	0.22	0.22	0.28	-
$n = 42$; $m = 40$						
ZAGA model						
Estimate	4.56 (0.48)	7.01 (1.46)	6.64 (1.38)	6.70 (1.39)	0.76 (0.08)	0.22 (0.17)
Sd	0.55	0.21	0.21	0.21	0.36	0.12
RMSE	0.55	0.21	0.21	0.21	0.37	0.12
gamma model						
Estimate	4.32 (1.19)	7.02 (1.46)	6.65 (1.39)	6.71 (1.40)	0.90 (0.16)	-
Sd	0.28	0.21	0.21	0.21	0.17	-
RMSE	0.37	0.21	0.21	0.21	0.25	-

Table 4.2.10: Parameter estimates from fitting a Poisson iREM to encounters simulated from a Poisson REM, and a ZAGA model to data simulated from a ZAGA model. We also fit a gamma model to the speed data, excluding the zero observations (standard error in parentheses). The true parameter values and sample sizes reflect those of the WWAP data. The sample sizes are n trap days, and m animal speed observations. The Standard deviation (Sd) and Root Mean Square Error (RMSE) are also given.

	Estimated Parameters					
	Habitat 1	Habitat 2	Habitat 3	Habitat 4		
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\mu}_x$	$\hat{w}$
$\beta_1 = 4.29$, $\beta_2 = 3.58$, $\beta_3 = 6.36$, $\beta_4 = 4.28$; $\mu_x = 1.17$; $w = 0.40$						
$n = 42$; $m = 10$						
ZAGA model						
Estimate	4.31 (0.40)	3.58 (0.67)	6.36 (1.17)	4.28 (0.79)	1.19 (0.08)	0.41 (0.28)
Sd	0.36	0.18	0.17	0.18	0.31	0.16
RMSE	0.36	0.18	0.17	0.18	0.31	0.16
gamma model						
Estimate	3.85 (1.48)	3.58 (0.67)	6.36 (1.17)	4.28 (0.79)	1.95 (0.65)	-
Sd	0.44	0.18	0.17	0.18	0.76	-
RMSE	0.62	0.18	0.17	0.18	1.09	-
$n = 42$; $m = 40$						
ZAGA model						
Estimate	4.29 (0.40)	3.54 (0.67)	6.37 (1.18)	4.29 (0.80)	1.17 (0.11)	0.40 (0.13)
Sd	0.18	0.18	0.18	0.18	0.22	0.08
RMSE	0.18	0.18	0.18	0.18	0.22	0.08
gamma model						
Estimate	3.80 (0.98)	3.59 (0.67)	6.36 (1.17)	4.29 (0.80)	1.93 (0.35)	-
Sd	0.24	0.18	0.18	0.18	0.35	-
RMSE	0.55	0.18	0.18	0.18	0.84	-

4.3 Application of iREM with habitat to the WWAP data set

Here we illustrate the application of iREM with habitat model proposed in this Chapter with the analysis of the 4 habituated mammals data set from Whipsnade Wild Animal Park. The data were analysed in Chapter 2, initially assuming that the Park where the data came from was not divided into habitats, and when habitat is incorporated in Rowcliffe et al. (2008) REM. The data were also analysed in Chapter 3 assuming that the Park where the data came from was not divided into habitats. In this section we present the results of an analysis with habitats in iREM.

We give the results for each of the four species obtained from iREM with habitat. These results are compared with the results from iREM given in Chapter 3, and the density from the census in Rowcliffe et al. (2008). The models were fitted using the optim function in the R software package (version 3.3.2 or lower) using the default Nelder-Mead algorithm.

The results for the wallaby species are given in Section 4.3.1. It then goes on, in Section 4.3.2 to give the results for the water deer species. The results for the munjac and mara species are given in Section 4.3.3, and Section 4.3.4, respectively. To determine whether the density from the census is captured within an approximate 95% confidence, we use the natural logarithm method discussed in Chapter 2 Section 2.5.1.

4.3.1 Estimated density of the wallaby species

The results show minimal differences between the three speed data models, so for illustration, we give the results from fitting a gamma model to the speed data in this section, and the results from the lognormal model (Table C.2.1) and the Weibull model (Table C.2.2) are given in appendix C.2.1.

Table 4.3.1 gives estimates for the wallaby species where animal speed follows a gamma model. Looking in isolation at the results from iREM, which does not consider the variation in density within habitats, would suggest that including habitat in the modelling process is very relevant for the wallaby data set. Estimated density from the ZIP iREM

is inflated but with large standard errors the density from the census is captured within an approximate 95% confidence interval. It is worth noting that the probability of the zero-response category from the encounter data for the wallaby species, $\rho \approx 0.1458$ and the encounter rate, $\lambda \approx 3.429$. The variance of the encounter data computed from a ZIP REM, $\text{Var}(y_{ij}) \approx 1.464$. Since $\text{Var}(y_{ij}) < \mathbb{E}(y_{ij})$, the ZIP REM suggests underdispersion and it is expected that a Poisson REM would show a better model fit to the encounter data. From fitting a gamma model to animal speed and ZIP REM to encounter data, the probability of the zero-response category is estimated as $\hat{\rho} = 0.126$ (0.284) does have an effect on estimated density as shown in Table 4.3.1. The lognormal and Weibull models gave similar estimates of the parameters. The results also show that density estimated in Central Park is large for all models, which might suggest that the animals generally frequented this area.

Table 4.3.2 compares the ΔAIC values from iREM with habitat and iREM. The iREM with habitat had the lowest AIC values and the most favoured models are the NB iREM where animal speed is assumed to follow a lognormal model, but all three alternative models for the speed data have close AIC values. It is worth noting that the sample size for the animal speed data is small, which may explain why the AIC values are close. Note that the ΔAIC values from the ZINB iREM are close to the ΔAIC values from the NB iREM, particularly, where animal speed is assumed to follow a lognormal model. This suggests that the ZINB iREM where animal speed follows a lognormal model would also be a suitable model in practice.

Table 4.3.1: Density estimates from iREM with habitat compared with density estimates from an iREM and the density from the census of the wallaby species (standard error in parentheses). Animal speed is assumed to follow a gamma model

	Habitat				**Total density**	
	Central Park	Downs	Institute Paddock	Old Farm	**iREM with habitat**	**iREM**
Census density ((D), km^2)	96	1101	760	803	468	468
Poisson	267 (106.41)	667 (371.29)	942 (527.09)	647 (364.67)	476 (178.26)	653 (243.20)
NB	267 (118.61)	667 (485.58)	942 (709.02)	647 (483.99)	476 (183.07)	653 (249.52)
ZIP	348 (336.95)	684 (796.66)	977 (1141.07)	659 (772.26)	531 (504.70)	763 (283.85)
ZINB	225 (90.13)	554 (376.49)	935 (664.76)	591 (414.90)	421 (141.44)	653 (249.51)

Table 4.3.2: Model selection for wallaby species. The ΔAIC values from fitting different iREM with habitat models and iREM models for the wallaby species.

	iREM with habitat				**iREM**			
	Poisson	NB	ZIP	ZINB	Poisson	NB	ZIP	ZINB
Speed data models								
gamma	346.22	1.16	286.86	6.40	416.70	10.40	320.20	12.40
lognormal	345.06	0	285.82	2.08	415.54	9.24	319.04	11.24
Weibull	346.06	3.66	273.18	3.74	416.54	10.24	320.06	12.78

4.3.2 Estimated density of the water deer species

Table 4.3.3 gives the results from fitting a Weibull model to animal speed data, while the results from a lognormal model (Table C.2.3) and a gamma model (Table C.2.4) are given in appendix C.2.2.

Estimates of the density from a Poisson iREM, a NB iREM and a ZINB iREM (Table C.2.4) show minimal differences, but there is a large difference between estimated density within three of the four habitats and the density from the census. Overall estimated density from the three count data models is close to the density from the census. Estimated density from the ZIP iREM within habitats is different from the other count data models, and the overall density is overestimated. If there is a problem with overdispersion a ZIP iREM would be a better fit than a Poisson iREM but like the wallaby species the data reflects underdispersion where $\text{Var}(y_{ij}) \approx 0.908 < \mathbb{E}(y_{ij}) = \lambda \approx 1.932$ and a Poisson iREM would be a better fit. However, despite the poor fit of the ZIP iREM the standard errors are large and capture the density from the census within an approximate 95% confidence interval. For all models, the density from the census for only two of the four habitats (Central Park and Downs) is captured within an approximate 95% confidence interval. Except for the ZINB iREM with habitat, which obtained similar estimate of the density as a ZINB iREM without habitat, the models with habitat obtained better overall estimates of the density. The results show that there is support for the Weibull model for animal speed (Table 4.3.4).

Table 4.3.3: Estimates from iREM with habitat compared with estimates from iREM and the density from the census of the water deer species (standard error in parentheses). Animal speed is assumed to follow a Weibull model.

	Habitat				**Total density**	
	Central Park	Downs	Institute Paddock	Old Farm	**iREM with habitat**	**iREM**
Census density ((D), km^2)	72	73	36	577	119	119
Poisson	150 (56.76)	52 (29.92)	6 (6.03)	188 (101.90)	115 (42.26)	92 (33.16)
NB	150 (63.65)	52 (37.78)	6 (6.65)	188 (135.79)	115 (45.33)	92 (34.89)
ZIP	204 (77.91)	75 (45.69)	9 (8.51)	250 (137.75)	157 (57.98)	176 (63.56)
ZINB	162 (66.94)	56 (39.28)	7 (7.05)	205 (140.18)	124 (48.21)	117 (52.50)

Table 4.3.4: Model selection for water deer species. The ΔAIC values are given for iREM with habitat and iREM models.

	iREM with habitat				**iREM**			
	Poisson	NB	ZIP	ZINB	Poisson	NB	ZIP	ZINB
Speed data models								
gamma	58.66	0.14	17.92	2.14	192.90	49.08	85.32	50.54
lognormal	61.44	3.02	21.00	4.62	195.38	51.56	87.80	53.02
Weibull	58.50	0	17.78	0.70	192.76	48.94	85.18	50.40

4.3.3 Estimated density of the muntjac species

In this section we give estimates of the density for the muntjac species. Here we give the results from a Weibull model in Table 4.3.5, used to model the animal speed data. Estimates from the three speed data models show minimal differences, therefore, we give the results from a lognormal model (Table C.2.5) and gamma model (Table C.2.6) in appendix C.2.3.

The results (Table 4.3.5) show that non-zero density is estimated in the habitat (Institute Paddock) where the observed census count was zero, and that incorporating habitat in iREM has improved estimated density substantially, particularly for a Poisson iREM and a NB iREM. Estimations from a ZIP iREM with habitat and ZINB iREM with habitat remained practically the same as the iREM without habitat but precision improves when habitat is added to the model. The ΔAIC values are given in Table 4.3.6, and they suggest the best model is a NB iREM with habitat where animal speed is assumed to follow a Weibull model. But the difference in ΔAIC values between the other iREM with habitat models is minimal because the sample sizes in the speed data are small.

Table 4.3.5: The estimated density from an iREM with habitat model is compared with estimated density from an iREM and the density from the census of the muntjac species (standard error in parentheses). Animal speed is assumed to follow a Weibull model.

	Habitat				**Total density**	
	Central Park	Downs	Institute Paddock	Old Farm	**iREM with habitat**	**iREM**
Census density ((D), km^2)	18	6	0	22	13	13
Poisson	25 (6.99)	6 (3.60)	1 (0.95)	4 (3.09)	15 (4.20)	8 (1.92)
NB	24 (8.40)	6 (4.43)	1 (1.01)	4 (3.63)	15 (4.95)	8 (2.18)
ZIP	35 (11.01)	10 (6.59)	1 (1.52)	7 (5.27)	23 (6.80)	23 (6.86)
ZINB	25 (10.57)	6 (32.09)	1 (4.25)	5 (22.67)	16 (71.41)	16 (10.72)

Table 4.3.6: Model selection for muntjac species. The ΔAIC values from iREM with habitat models and iREM models for the muntjac species are given.

	iREM with habitat				**iREM**			
	Poisson	NB	ZIP	ZINB	Poisson	NB	ZIP	ZINB
Speed data models								
gamma	3.90	13.50	0.92	2.92	41.72	20.56	21.48	20.48
lognormal	4.26	0.08	1.28	2.56	42.10	20.92	21.84	22.48
Weibull	4.26	0	1.28	1.98	42.08	20.92	21.82	22.46

4.3.4 Estimated density of the mara species

At WWAP, Central Park was the largest habitat and was heavily frequented by people. In this habitat the camera traps were strategically placed to avoid human interference and film wastage due to overcrowding of cameras with human photographs. This placement strategy coupled with the fact that around 90% of the mara were observed in the Central Park area during the census counts at WWAP explains the dramatic underestimation of the mara species (see Rowcliffe et al., 2008).

Here we give estimates of the density of the mara species where animal speed is assumed to follow a gamma model, shown in Table 4.3.7. The results from the lognormal model (Table C.2.7) and the Weibull model (Table C.2.8) are given in appendix C.2.4. The results show that estimation of the density is highest in Central Park for the speed data models (Table 4.3.7, Table C.2.7, and Table C.2.8), and estimated density in Old Farm is non-zero, which has observed census count of zero. The gamma model (Table 4.3.7) and Weibull model (Table C.2.8) obtained similar estimates of the density, but the lognormal model (Table C.2.7) is less accurate than these models. Adding habitat as a covariate only marginally improved the estimated density of the mara species. The results show support for a Poisson iREM where animal speed is assumed to follow a gamma model, but the difference between a Weibull model is small (Table 4.3.8).

Table 4.3.7: Density estimates from iREM with habitat compared with density estimates from iREM and the density from the census of the mara species (standard error in parentheses). Animal speed is assumed to follow a gamma model.

	Habitat				**Total density**	
	Central Park	Downs	Institute Paddock	Old Farm	**iREM with habitat**	**iREM**
Census density ((D), km^2)	108	30	7	0	68	68
Poisson	7 (5.31)	3 (4.95)	2 (3.60)	3 (5.53)	3 (3.35)	4 (2.00)
NB	7 (5.31)	3 (4.96)	2 (3.70)	3 (5.53)	5 (3.35)	4 (2.00)
ZIP	7 (5.31)	3 (4.95)	2 (3.69)	3 (5.53)	5 (3.37)	4 (2.00)
ZINB	8 (5.44)	3 (5.02)	2 (3.76)	3 (5.60)	5 (3.46)	4 (2.14)

Table 4.3.8: Model selection for muntjac species. The ΔAIC values for iREM with habitat models and iREM models for the mara species are given.

	iREM with habitat				**iREM**			
	Poisson	NB	ZIP	ZINB	Poisson	NB	ZIP	ZINB
Speed data models								
gamma	4.27	6.27	8.27	10.30	0	2.00	2.01	4.00
lognormal	5.12	7.12	9.12	11.22	0.95	2.95	2.95	4.97
Weibull	4.37	6.37	8.45	10.43	0.09	2.09	2.09	4.09

4.3.5 Estimated density of species with observed zero speed of movement

In Chapter 2, Section 2.7 we discussed the data at WWAP where it was observed that animals generally did not move about much. And given that the survey design involved following the animals around for 30 minutes recording the distance travelled and duration of passage, some animals did not move during this period, hence, zero speeds were recorded. In Chapter 3, Section 3.3, we gave a ZAGA model for data where there are animals that are observed to not move. In this section we fit this model to the speed data to estimate density within habitats and overall density. At WWAP three of the four species, wallaby, water deer and mara, were observed to not move in one or more speed samples. We fit a Poisson iREM, a NB iREM, a ZIP iREM, or a ZINB iREM to the data. Estimates of the density from iREM with habitat are compared with estimates of the density from iREM, and the density from the census. Tables with the results are given in appendix C.2.5. As shown in the simulation study in Section 4.2.7 above it is relevant to include the observed zeros to avoid bias in the estimates. The simulation results (Section 4.2.7) showed that not accounting for the observed zeros in the speed of movement can introduce a negative bias in the estimate in the reference habitat, and hence, and underestimation of the mean density across habitats.

We give the results of the wallaby species in Table C.2.9, in appendix C.2.5. Note that the census in Institute Paddock and Downs is recorded this way but are actually observed the other way around. The results show that the observed zeros of animal speed does have an effect on estimated density. Estimated densities within all the habitats,

except Central Park, are similar to the census (considering the observed data in Downs and Institute Paddock and the recording of these data). Also, the estimated density of the wallaby species at WWAP improved substantially when habitat is included in the model.

For the water deer species (Table C.2.10, appendix C.2.5) a Poisson iREM with habitat does considerably better than the other three iREM with habitat models, with the estimated density of the water deer species at WWAP (124 per km^2) being close to the density from the census. This is a considerable improvement compared with a density of 152 per km^2 from a Poisson iREM. For the NB iREM with habitat, ZIP iREM with habitat and ZINB iREM with habitat, estimations of the density in Central Park are quite large, and given that Central Park is the largest area the density of the water deer species is expected to be large overall.

Table C.2.11, in appendix C.2.5, compares estimated density from iREM with habitat with estimated density from iREM without habitat for the mara species. When the observed zero speeds of the mara species are considered estimated density in Central Park improves, a result which is consistent with the behaviour of the mara species. The iREM with habitat model does better in estimating the density of the mara species than iREM without habitat. The ΔAIC values are given in Table 4.3.9. The NB iREM best explains the observed data for species with a larger data set (wallaby and water deer), while the Poisson iREM best explains the observed data of the mara species. Note that the ΔAIC values of NB iREM are relatively small compared with best model for the mara species. Therefore, we might be able to use a NB iREM if the data set increases.

Table 4.3.9: ΔAIC values from iREM with habitat and from iREM adjusting for zeros in animal speed. The ΔAIC values are given for the wallaby species, water deer species and mara species.

	iREM with habitat				**iREM**			
	Poisson	NB	ZIP	ZINB	Poisson	NB	ZIP	ZINB
Species								
wallaby	406.30	0	309.82	2.14	335.82	149.96	284.42	4.10
water deer	181.58	0	17.88	1.98	192.74	48.92	85.28	36.78
mara	4.28	6.30	6.30	8.55	0	2.00	2.01	4.00

4.4 iREM with random effect

The model structure in Section 4.1.1 may readily be adjusted to allow the encounter rate ($\boldsymbol{\lambda}$) to dependent upon some random components, such as camera random effects. As with the covariate model, the random effects are introduced via a log-linear link on the encounter rate.

Suppose the encounter records is defined by y_{ij} where $j = 1, 2, ..., n$ is the jth observation on the ith $(= 1, 2, ..., c)$ camera trap. With random effect b_i, the probability mass function of the encounter data is given as $h(y_{ij}\ ;\ \lambda_i)$, where

$$\lambda_i = \left(\frac{2+\theta}{\pi}\mu_x rt\right)\exp\{\log(D) + b_i\} \tag{4.4.1}$$

is expected encounter rates of each camera day; b_i is a vector of camera random effects; the camera dimensions are r and θ; $t = 1$ day; μ_x is the expected animal speed; D is the density with or without the camera random effect since the density from equation (4.4.1) is $\exp\{\log(D)\} = D$. We expect the location of the camera trap to have an effect on the expected encounter rate. As discussed in Section 2.6, REM does not allow camera traps to target focal species - if, for example, traps are baited or set on trails, this would violate the underpinning assumption of REM - but rather REM requires randomized placement of camera traps to avoid bias in the estimate of the density. As we have seen in Chapters 2, and 3 for the mara species, the location of the camera trap had an effect on the expected encounter rate as shown by 1) the limited capture records

and 2) the dramatic underestimation of the density.

Here we assume that the random camera effect, b_i, follows a normal distribution with mean 0 and variance σ_b^2, that is

$$b_i \sim \mathrm{N}(0,\ {\sigma_b}^2),$$

with probability density function

$$g(b_i \mid {\sigma^2}_b) = \frac{1}{\sqrt{2\pi}\sigma_b}\mathrm{e}^{-{b_i}^2/(2{\sigma_b}^2)}.$$

Then to obtain the marginal likelihood of y_{ij}, the random effect b_i must be integrated out. That is, the observed encounter distribution is obtained after integration as

$$h(y_{ij}) = \int_{-\infty}^{\infty} h(y_{ij};\ \lambda_i)\ g(b_i \mid {\sigma^2}_b)\ db_i. \tag{4.4.2}$$

No closed-form solution for this integral is available, therefore to obtain the maximum likelihood estimate numerical approximation of the integral is required. There are several approximation methods but in this thesis we used the Gauss-Hermite quadrature, which is a form of Gaussian quadrature for approximating the value of integrals of an unknown function over a specified domain such as

$$\int_{-\infty}^{+\infty} e^{-x^2} f(x). \tag{4.4.3}$$

The Gauss-Hermite approximation for the function in equation (4.4.3) is

$$\sum_{i}^{n} w_i f(x_i), \tag{4.4.4}$$

where w_i are called weights and x_i are called abscissas. In the next section we show how the function, h_{ij} above can be written in the form of equation (4.4.3), and the approximation in equation (4.4.4) can be used.

4.4.1 Gauss-Hermite quadrature

The starting point for the Gauss-Hermite quadrature is the integral

$$h(y_{ij}) = \int_{-\infty}^{\infty} h(y_{ij} \ ; \ \lambda_i) \ g(b_i \mid {\sigma^2}_b) db_i. \tag{4.4.5}$$

where λ_i is the expected encounter rates of each camera day as defined in equation (4.4.1). The function $h(\cdot)$ cannot be solved by analytical methods. By assuming an appropriate change of variable equation (4.4.5) can be brought into form as in equation (4.4.4). Let us consider the probability density function $h(y_{ij} \ ; \ \lambda_i)$ where $b_i \sim \mathrm{N}(0, \ {\sigma^2}_b)$ such that

$$g(b_i) = \frac{1}{\sqrt{2\pi}\sigma_b} \exp\left\{ -\frac{1}{2}\left(\frac{b_i}{\sigma_b}\right)^2 \right\}.$$

The expectation of $h(y_{ij})$ corresponds to

$$\mathbb{E}\{h(y_{ij})\} = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_b} \exp\left\{ -\frac{1}{2}\left(\frac{b_i}{\sigma_b}\right)^2 \right\} h(y_{ij} \ ; \ \lambda_i) db_i.$$

Using a change of variable b_i to v where $v = b_i/\sqrt{2}\sigma_b$ and coupled with integration of substitution we have

$$\mathbb{E}\{h(y_{ij})\} = \int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi}} \mathrm{e}^{-v^2} h(y_{ij} \ ; \ \lambda_i) dv,$$

which is of the form in equation (4.4.3) above. Therefore,

$$\mathbb{E}\{h(y_{ij})\} \approx \frac{1}{\sqrt{\pi}} \sum_{q=1}^{Q} w_q h_{ij}(\sqrt{2}\sigma_b v_q),$$

where w_q are weights and v_q are evaluation points (Winkelmann, 2008, pages. 286-287).

4.4.2 iREM with camera random effect

In Rowcliffe et al. (2008) camera placement strategies are not allowed to target species, and camera traps must be randomized, otherwise placement strategies would violate the underpinning assumption of REM. And the nonrandom location of camera traps would have an effect on estimated density. In this section we give an iREM with camera

random effect.

Suppose y_{ij}, the jth $(= 1, 2, ..., n)$ encounter frequency on the ith $(= 1, 2, ...c)$ camera has probability mass function $h(y_{ij}\ ;\ \lambda_i)$ where λ_i is the expected encounter rates given in equation (4.4.1) above. We assume that the encounters, y_{ij}, are independent for all i and all j. And suppose m independent speed observations $x = \{x_1, ..., x_m\}$ have joint probability density function $f(x_l \mid \mu_x,\ \nu)$, where $l = 1, 2, ..., m$, μ_x is the expected speed, and ν represents any additional parameters in the model. Using the Gauss-Hermite quadrature derived above for the approximation of the integral then the following is the construction of the log-likelihood function for iREM with random effect

$$\ell \approx \sum_{i=1}^{c} \log \sum_{q=1}^{Q} \prod_{j=1}^{n_i} \frac{w_q h_{ij}(\sqrt{2}\sigma_b d_q)}{\sqrt{\pi}} + \sum_{l=1}^{m} \log\ f(x_l \mid \mu_x,\ \nu). \tag{4.4.6}$$

where

$$h_{ij}(\sqrt{2}\sigma_b d_q) = h(y_{ij} \mid b_i = \sqrt{2}\sigma_b d_q;\ \lambda_i)$$

is the conditional density function of the encounters, evaluated at a camera random effect, b_i. The symbol d_q represents the quadrature evaluation points; w_q are the quadrature evaluation weights; and Q is the number of quadrature points. We use $Q = 20$ quadrature points following recommendations by Choquet and Cole (2012) and Cole et al. (2003). As iREM is an extension of REM, the assumptions in REM are also the assumptions in iREM. In iREM with camera random effect, the assumptions are:

(i) cameras are located in clear view of the area to capture animals

(ii) camera traps are randomly placed relative to the space used by animals

(iii) animals move independently of camera traps.

The distributions used in Chapter 3, Section 3.4 to model the encounters, y_{ij}, and animal speed observations, x used in Chapter 3, Section 3.3 are used here. We assume the encounters follow a Poisson REM, a NB REM, a ZINB REM or ZIP REM, and the speed data is assumed to follow a gamma model, a lognormal model or a Weibull model.

4.4.3 Simulation Study

In this section we test the performance of iREM with camera random effect via simulations. We use a Poisson iREM for illustration where animal speed is assumed to follow a lognormal model. The parameter required to generate encounters to fit a Poisson iREM with camera random effects is the expected encounter rate, λ_i, which is dependent on the density, D; camera random effects b_i; the camera dimensions, r and θ; expected animal speed, μ_x, and the camera trap time, t. In the simulation study we assume a fixed value of the number of camera trap days, n, for ease of computation.

Step 1: First we start with c camera traps, which have been set up for n camera trap days. Note that n can vary for each trap, which we illustrate in the R codes in appendix C.3. The random component of the camera, b_i, is generated from a normal distribution with mean zero, and variance ${\sigma_b}^2$ such that $b_i \sim \mathrm{N}(0, {\sigma_\mathbf{b}}^2)$. We then calculate density, $\mathbf{D} = \exp\{\log(D_s) + b_i\}$, where D_s is the density value without camera random effect, and b_i is the effect of camera i. Next we calculate the expected encounter rate, $\lambda_i = ((2+\theta)/\pi) r t \bar{v} \mathbf{D}$, where r (km) and θ (radians) are the dimensions of the camera trap; $\bar{v}$ is the mean animal speed and t is the camera trap time, which are held fixed.

Step 2: Next we generate encounters, $\mathbf{Y}$, on each camera trap from a Poisson distribution with encounter rate, λ_i, computed in *Step* 1.

Step 3: For m animals moving at speeds $x = \{x_1, ..., x_m\}$, we generate speeds x, which are random numbers from either a gamma distribution $\mathrm{X} \sim Ga(\alpha; \nu)$, a lognormal distribution $\mathrm{X} \sim lnN(\mu; \sigma)$ or a Weibull distribution $\mathrm{X} \sim Wei(\tau; \beta)$. We then fit a Poisson REM with camera random effect to the encounters data, $\mathbf{Y}$, and a Poisson REM without random effect to the encounters, $\mathbf{Y}$. As a demonstration we use a lognormal model from which the speed, X, is generated, and we fit a lognormal model to X.

For the simulations we generate scenarios in which the density and the camera random effect are plausible ecologically for real species. We set $D = 200$; $\mu_x = 0.60$ (ms^{-1}); $\sigma_b = 0.10$ or $\sigma_b = 1.10$, $n = 40$ or $n = 100$; and $m = 40$ or $m = 100$. For each simulation

case, we compute the average parameter estimates (standard error in parentheses), Standard deviation (Sd) and Root Mean Square Error (RMSE) for 100 simulation runs. The parameters used to estimate density: $r = 0.012$ (km); $\theta = 0.175$ (radians), and camera trap time period $t = 1$ (day) are held fixed. Under all scenarios, estimates from iREM with random effect are compared with estimates from iREM without random effect. Both methods are fitted to encounter data with camera random effect. For illustration we give the results from fitting a Poisson iREM with random effects, and a Poisson iREM without random effects where animal speed is assumed to follow a lognormal model. We also compute the coefficient of variation ($C_v\%$), expressed as a percentage, using the equation (2.7.1) given in Chapter 2 Section 2.7.1. The percentage bias (% bias) of the parameters is also given, which is computed as

$$\%\text{bias} = \frac{\text{average parameter estimate} - \text{true parameter value}}{\text{true parameter value}} \times 100.$$

Table 4.4.1 compares the estimated parameters from a Poisson iREM with random effect with a Poisson iREM without random effect. The simulation results (Table 4.4.1) show that the parameter estimates from the two methods have minimal differences when the camera random effect is small, and the estimated random effect is not different from zero at the 5% significance level. Also, for both methods the bias in the density is negligible (less than 3% of the true population density, on average) for small sample sizes, and 1% or less for larger sample sizes. The $C_v\%$ of the density is also small, suggesting low variability in estimated densities. However, the $C_v\%$ of the estimated camera random effect ($\hat{\sigma}_b$) is huge suggesting that the variability in the estimates is high. Increasing the camera random effect results in a large difference in the estimated density between the two methods, and the bias is substantial for the model without camera random effect (more than 60% of the true population parameter) for small sample sizes. Increasing sampling effort does not improve the density estimate. The estimated camera random effect ($\hat{\sigma}_b$) is different from zero (that is, an approximate 95% does not contain zero), suggesting that accounting for the random location of camera traps is relevant in estimating the density. Also, the $C_v\%$ of the density estimator from the model without camera random effect is huge (more than 30% for both small and large sample sizes) suggesting high variability in the estimates.

Table 4.4.1: Simulation results from a Poisson iREM with random effect and Poisson iREM without random effect. The camera random effect is set to, $\sigma_b = 0.10$ or $\sigma_b = 1.10$. The standard errors, Standard deviation (Sd) and Root Mean Square Error (RMSE) are given. The coefficient of variation ($C_v\%$) and percentage bias, %bias, of the estimators are given (negative bias (%bias) is given in parentheses).

	Model with random effect			**Model without random effect**	
	$\hat{D}$	$\hat{\mu}_x$	$\hat{\sigma}_b$	$\hat{D}$	$\hat{\mu}_x$
$\mu_x = 0.60;\ D = 200;\ \sigma_b = 0.10;\ n = m = 40$					
estimate	204 (15.96)	0.59 (0.04)	0.08 (0.07)	205 (14.97)	0.60 (0.04)
Sd	16.96	0.04	0.06	16.86	0.04
RMSE	17.52	0.04	0.004	17.57	0.04
%bias	+2%	−1.6%	−20%	+2.5%	0%
$C_v\%$	8.31%	0.07%	75.00%	8.27%	6.60%
$\mu_x = 0.60;\ D = 200;\ \sigma_b = 0.10;\ n = m = 100$					
estimate	201 (11.93)	0.60 (0.02)	0.08 (0.06)	202 (10.50)	0.60 (0.02)
Sd	14.03	0.02	0.05	10.50	0.02
RMSE	14.12	0.02	0.003	14.17	0.02
%bias	+0.5%	0%	−20%	+1%	0%
$C_v\%$	6.98%	3.30%	62.50%	5.20%	3.30%
$\mu_x = 0.60;\ D = 200;\ \sigma_b = 1.10;\ n = m = 40$					
estimate	206 (32.51)	0.61 (0.04)	1.05 (0.11)	334.74 (31.04)	0.65 (0.05)
Sd	28.85	0.05	0.21	149.69	0.10
RMSE	30.87	0.05	0.22	201.40	0.11
%bias	+3%	+1.6%	+4.54%	+67.5%	+8.3%
$C_v\%$	14.00%	8.20%	20%	44.72%	15.38%
$\mu_x = 0.60;\ D = 200;\ \sigma_b = 1.10;\ n = m = 100$					
estimate	203 (18.16)	0.61 (0.03)	1.02 (0.05)	354 (20.03)	0.69 (0.03)
Sd	25.97	0.05	0.06	153.75	0.42
RMSE	26.09	0.05	0.07	217.81	0.43
%bias	+1.5%	+1.6%	+7.27%	+77%	+15%
$C_v\%$	12.79%	8.20%	5.88%	43.43%	60.87%

4.5 Application of iREM with random effects to WWAP data set

We illustrate the application of iREM with random effect proposed in this Chapter with an analysis of the data sets for the four species (wallaby, water deer, mara and muntjac) at WWAP, which consists of the location of the camera traps in habitat p and the number of encounters recorded on each camera trap. We fit a Poisson iREM, a NB iREM, a ZINB iREM, or a ZIP iREM. In each case a gamma model, a lognormal model or a Weibull model is fitted to animal speed data. Guass-Hermite quadrature is used to approximate the marginal likelihood of the encounters. We use the natural logarithm in equation (2.5.1) in Chapter 2, Section 2.5.2 to compute approximate 95% confidence intervals of the density. We also compare models using AIC criteria as described in Chapter 1. We defined the AIC criteria as

$$\text{AIC} = -2\log L + 2p$$

where p is the 'degrees of freedom' correction, or the number of parameters in the model. But Vaida and Blanchard (2005) state that the definition of the AIC is not straightforward when random effects are contained within the model under consideration. Questions such as what likelihood should be used (marginal likelihood or conditional likelihood), and should the random effects be counted as parameters are often raised. Vaida and Blanchard (2005) argued that the answer to these questions depends on the focus of the research. If the focus is on making inference about population parameters, then the likelihood is the marginal likelihood, and p is the number of fixed parameters, counting the mean parameters and the variance components. Here our interest is estimating density and determining the overall effect of the random location of the camera traps on the encounter rate, and hence, the density. Therefore, the AIC formula in current use, which Vaida and Blanchard (2005) defined as the marginal AIC, mAIC, is appropriate to use in our context: $\text{mAIC} = -2\log L + 2p$.

4.5.1 Estimated parameters of the wallaby species

In this section we give the results of the wallaby species. Table 4.5.1 compares estimates from iREM with random effect with estimates from iREM without random effect. We

assume the speed data without observed zeros follows a gamma model and the speed data with observed zeros follows a zero-adjusted gamma (ZAGA) model.

The results (Table 4.5.1) show that ignoring estimation uncertainty in the random effects covariance matrix induces a much larger bias in the density estimator. Moreover, the effect of the random placement of camera traps on the density estimator is significant (that is, an approximate 95% confidence interval of σ_b does not contain zero), particularly when the observed variation in the encounter data is unaccounted for (i.e., using a Poisson iREM with random effect and from a ZIP iREM with random effect). In fact the estimated $\hat{\sigma}_b$ is larger for these models compared with the estimated $\hat{\sigma}_b$ from a NB iREM and a ZINB IREM. Note that including the camera random effect can account for unobserved variation in the data (Gschlößl and Czado, 2008), hence, smaller estimates of the camera random effect from a NB iREM with random effect and a ZINB iREM with random effect. These results correspond with the simulation results in Section 4.4.3 above, giving clear indication of the importance of accounting for the randomness in the placement of camera traps when estimating the density.

We also compare the results from modelling the observed zero speeds with the results from excluding these zeros in the estimation the process (Table 4.5.1). We expect the observed zero speed of movement to have an effect on estimated density. As discussed in Chapter 3 and as shown in Section 4.2.7 not accounting for observed zero speeds can induce a negative bias in the density estimator. As can be seen in Table 4.5.1 estimates from iREM with random effect (without zero speed) is smaller compared with estimates from iREM with random effect (with zero speed). Note that the sample size of the speed observations is small, and therefore, including the observed zero speeds would inflate the density (see, for example, Section 3.7.6). Also, it worth reiterating that the density from the census at WWAP is the mean density across habitats so we expect that the estimated density and the density from the census will be different since the variation in habitat is not accounted for.

The results (Table 4.5.1) also suggest that it is relevant to account for the sampling variability in the data (NB iREM) as indicated by the ΔAIC values but the density

estimate from the Poisson iREM with random effect (without zero speed) is closer to the density from the census. Note that the probability of selecting models that have too many parameters using the AIC method increases when the sample size is not many times larger than the square of the number of parameters in the model, that is, overfitting (see Bedrick and Tsai, 1994), and the WWAP data set is a rather small data set.

The results from the lognormal model (Table C.4.1) and from the Weibull model (Table C.4.2) are given in appendix C.4. The difference in estimated density and the density from the census from the lognormal model (C.4.1, in appendix C.4) is considerably smaller compared with the difference from the gamma model (Table 4.5.1). For example, for a Poisson iREM with random effect the difference in estimated density from the gamma model is 2 times more than the difference in estimated density from the lognormal model and the density from the census. Including the camera random effect can account for unobserved variation in the data (Gschlößl and Czado, 2008), hence, smaller estimates of the camera random effect from a NB iREM with random effect and a ZINB iREM with random effect. The results from a Weibull model (Table C.4.2, appendix C.4) show that this model does not perform well in estimating density when camera random effect is incorporated in the model. For all count data models, the difference in estimated density and the density from the census is substantially large. For a Poisson iREM in particular, the estimated $\hat{\sigma}_b$ is larger than the estimated camera random effect from the gamma and lognormal models, and it is statistically different from zero.

The ΔAIC values from iREM, iREM with habitat and iREM with random effect used to estimate density of the wallaby species are given in Table 4.5.2. The results show that there is support for accounting for variation in the encounter data. The best model does not include camera random effect but includes habitat. It is also worth noting that camera random effect will account for some of the variation in habitat. The results show no support for the underlying Poisson iREM and the ZIP iREM. The ΔAIC values for the alternative speed data models are close but the best model selected is the lognormal model for the speed data and the NB iREM with random effects for the encounter data. When we account for the observed zero speeds the model that bets describes the wallaby data set is iREM.

Table 4.5.1: Estimated parameters for the wallaby species from iREM with random effect models and iREM without random effect models where animal speed data is assumed to follow a gamma distribution, and where animal speed is assumed to follow a ZAGA model (standard error in parentheses).

	iREM with random effect (without zero speed)				**iREM (without zero speed)**			
	Poisson	NB	ZIP	ZINB	Poisson	NB	ZIP	ZINB
D = 468								
$\hat{D}$	498 (191.55)	614 (241.60)	732 (274.99)	620 (258.30)	653 (243.20)	653 (249.55)	763 (283.85)	653 (249.51)
$\hat{\mu}$	0.88 (0.33)	0.88 (0.33)	0.88 (0.33)	0.87 (0.35)	0.88 (0.65)	0.88 (0.66)	0.88 (0.65)	0.88 (0.66)
$\hat{\nu}$	1.03 (0.59)	1.03 (0.59)	1.03 (0.59)	0.89 (0.23)	1.03 (0.59)	1.03 (0.59)	1.03 (0.59)	1.03 (0.59)
$\hat{\sigma}_b$	0.54 (0.09)	0.30 (0.18)	0.53 (0.08)	0.30 (0.18)	-	-	-	-
$\hat{k}$	-	1.03 (0.17)	-	0.97 (0.16)	-	0.95 (0.14)	-	1.03 (0.15)
$\hat{\rho}$	-	-	0.12 (0.03)	2e+05 (0.001)	-	-	0.14 (0.03)	3e-07 (8e+03)
$\hat{\omega}$	-	-	-	-	-	-	-	-
$(-\ell)$	550.51	389.87	517.91	389.79	594.59	390.42	545.32	390.42
AIC	1113.02	791.74	1049.82	793.58	1197.11	790.84	1100.64	792.84
ΔAIC	321.28	0	258.08	1.84	406.29	0	309.80	2.00
	iREM with random effect (with zero speed)				**iREM (with zero speed)**			
	Poisson	**NB**	**ZIP**	**ZINB**	**Poisson**	**NB**	**ZIP**	**ZINB**
$\hat{D}$	621 (258.82)	769 (325.65)	647 (256.49)	615 (242.39)	816 (330.23)	816 (337.58)	952 (385.18)	882 (249.55)
$\hat{\mu}$	0.71 (0.28)	0.71 (0.28)	0.78 (0.29)	0.88 (0.33)	0.66 (0.76)	0.71 (0.80)	0.66 (0.76)	0.88 (0.66)
$\hat{\nu}$	0.88 (0.19)	0.88 (0.19)	0.88 (0.19)	1.11 (0.24)	0.88 (0.33)	0.88 (0.33)	0.88 (0.33)	1.03 (0.59)
$\hat{\sigma}_b$	0.60 (0.11)	0.30 (0.18)	0.46 (0.09)	0.26 (0.16)	-	-	-	-
$\hat{k}$	-	1.03 (0.17)	-	1.01 (0.17)	-	0.95 (0.14)	-	1.06 (0.15)
$\hat{\rho}$	-	-	0.12 (0.03)	0.0002 (0.001)	-	-	0.17 (0.04)	0.001 (0.01)
$\hat{\omega}$	0.20 (0.16)	0.20 (0.16)	0.20 (0.16)	0.20 (0.16)	-	-	0.25 (0.20)	0.24 (0.18)
$(-\ell)$	555.51	394.87	522.51	394.87	599.57	395.42	550.33	395.41
AIC	1123.02	803.74	1059.02	805.74	1207.14	800.84	1110.66	802.82
ΔAIC	319.28	0	255.28	2	406.30	0	309.82	1.98

Table 4.5.2: Model selection for wallaby species. ΔAIC values of the wallaby species split by methods: iREM with habitat, iREM with random effect and iREM.

	iREM with habitat				**iREM with random effects**				**iREM**			
	Poisson	NB	ZIP	ZINB	Poisson	NB	ZIP	ZINB	Poisson	NB	ZIP	ZINB
Speed data models												
gamma	346.22	1.16	288.52	6.40	330.58	11.30	267.33	11.30	416.70	10.40	320.20	12.40
lognormal	345.06	0	285.82	2.08	329.42	10.12	265.38	10.12	415.54	9.24	319.04	11.24
Weibull	346.06	3.66	286.88	3.74	332.44	9.14	266.40	3.58	416.54	10.24	320.06	12.78
ZAGA	335.82	149.96	284.42	4.10	320.22	0.90	256.18	0.90	406.30	0	309.82	2.14

4.5.2 *Estimated parameters of the water deer species*

The results of the water deer species are given in this section. Table 4.5.3 compares the estimated parameters from iREM with random effect models with the estimated parameters from iREM without random effect models. We also give the estimates from iREM (with zero speed) and iREM (without zero speed) for comparison purposes. The results show that it is relevant to account for the uncertainty in the random placement of camera traps as this can have a significant effect on the density estimator (i.e., an approximate 95% confidence interval of σ_b does not contain zero). Also, more support is given to models that allow for zero inflation when the uncertainty in the random effects variance-covariance matrix is considered (see the ΔAIC values for ZIP iREM with or without the observed zero speeds). One possible reason for this could be as result of the behaviour of animals in terms of their movement and the area they frequent, and given that camera placements are randomized (not set on trails). For example, some camera traps may be placed where animals do not frequent, and hence, zero capture records would be observed.

Including the observed zeros inflated the density estimator for all models, except the ZIP iREM. Note that the number of speed observations for the water deer species is $m = 10$ but 4 of these observations are zeros, hence, the estimated mean speed would be smaller compared with the mean speed estimated when these zeros are excluded. Therefore, we expect the density estimator to be inflated. However, the estimated standard errors are large and the density from the census is captured within an approximate 95% confidence interval; this is the case for all models, except ZIP iREM with random effect. The Poisson iREM with random effect (without zero speed) gave a dramatic underestimation of the density and the standard error. Also, for the Poisson iREM with random effect and the ZIP iREM with random where the observed zero speeds are not included, the density from the census is not captured within an approximate 95% confidence interval.

Again, it is worth reiterating that the density from the census is the mean density across habitats, and since the variation in habitat type is not accounted for then we expect the estimated density and the density from the census to be different.

The lognormal model (Table C.4.3, in appendix C.4) performs better than the gamma model as three of the four count data models gave estimates than are similar to the density from the census. The difference between estimated density from a ZIP iREM and the density from the census is substantially large, but the density from the census is just captured within an approximate 95% confidence interval. The ΔAIC value shows that a ZIP iREM is the best model, however, estimated density from a Poisson iREM, which has an AIC of at least 18 units more, is more accurate.

For the Weibull model there is a positive difference between the estimated densities from a Poisson iREM, a NB iREM, a ZINB iREM, and the density from the census, while the difference between estimated density from a ZIP iREM and the density from the census is negative (Table C.4.4, appendix C.4). The best model according to the ΔAIC values is a ZIP iREM. However, the difference between estimated density from a ZIP iREM and the density from the census is almost 2 times more than the difference between estimated density from a NB iREM and the density from the census, but a NB iREM has an AIC of at least 10 units more. The estimated camera random effect, $\hat{\sigma}_b$, from all models are different from zero, since zero is not contained within an approximate 95% confidence interval.

Table 4.5.4 gives the ΔAIC values across all models for the water deer species. The results show that a NB iREM with habitat is the best model where animal speed is assumed to follow a Weibull model. However, the difference between the minimum AIC value and the AIC values from a ZIP iREM with random effect is minimal. This suggests that accounting for the variation in the random location of camera traps is relevant in estimating the density of the water deer species. Also, when we account for the observed zero speeds the results indicate that there is more support for iREM with random effect.

Table 4.5.3: Estimated parameters for the water deer species from iREM with random effect models and iREM without random effect models where animal speed data is assumed to follow a gamma distribution, and where animal is assumed to follow a ZAGA model (standard error in parentheses).

	iREM with random effect (without zero speed)				**iREM (without zero speed)**			
	Poisson	NB	ZIP	ZINB	Poisson	NB	ZIP	ZINB
D = 119								
$\hat{D}$	37 (17.27)	193 (90.38)	72 (31.13)	184 (77.21)	91 (36.76)	91 (38.29)	174 (70.35)	115 (55.95)
$\hat{\mu}$	1.95 (0.78)	1.89 (0.85)	1.90 (0.73)	1.95 (0.78)	1.95 (2.54)	1.95 (2.58)	1.95 (2.55)	1.95 (2.57)
$\hat{\nu}$	0.54 (0.35)	0.42 (0.30)	0.57 (0.36)	0.54 (0.35)	0.54 (0.35)	0.54 (0.35)	0.54 (0.35)	0.54 (0.35)
$\hat{\sigma}_b$	1.25 (0.15)	2.03 (0.28)	1.54 (0.31)	1.80 (0.25)	-	-	-	-
$\hat{k}$	-	2.77 (1.16)	-	0.20 (0.13)	-	0.51 (0.10)	-	1.21 (0.75)
$\hat{\rho}$	-	-	0.14 (0.07)	0.09 (0.08)	-	-	0.48 (0.09)	0.27 (0.32)
$\hat{\omega}$	-	-	-	-	-	-	-	-
$(-\ell)$	225.01	223.88	220.57	222.67	318.52	245.61	263.73	245.34
AIC	460.02	459.76	453.14	459.34	643.04	499.22	535.46	500.68
ΔAIC	6.88	6.62	0	6.20	143.82	0	36.24	1.46
	iREM with random effect (with zero speed)				**iREM (with zero speed)**			
	Poisson	NB	ZIP	ZINB	Poisson	NB	ZIP	ZINB
$\hat{D}$	310 (148.99)	202 (134.05)	49 (23.91)	193 (23.91)	152 (72.73)	152 (74.89)	310 (145.41)	192 (105.55)
$\hat{\mu}$	1.71 (0.57)	1.71 (1.11)	1.71 (0.69)	1.80 (0.89)	0.65 (1.28)	1.17 (1.33)	0.62 (1.19)	1.17 (2.40)
$\hat{\nu}$	1.95 (0.50)	3.09 (0.88)	1.95 (0.50)	5.35 (1.57)	1.95 (0.78)	1.95 (0.78)	1.81 (0.70)	1.96 (0.78)
$\hat{\sigma}_b$	1.77 (0.18)	1.87 (0.28)	3.36 (0.56)	1.80 (0.26)	-	-	-	-
$\hat{k}$	-	2.40 (0.94)	-	0.20 (0.09)	-	0.51 (0.10)	-	1.21 (0.75)
$\hat{\rho}$	-	-	0.13 (0.07)	0.14 (0.07)	-	-	0.66 (0.43)	0.40 (0.26)
$\hat{\omega}$	0.40 (0.26)	0.45 (0.28)	0.40 (0.26)	0.66 (0.44)	0.67 (0.43)	0.40 (0.26)	0.87 (0.16)	0.21 (0.25)
$(-\ell)$	237.53	230.98	227.35	230.23	325.25	252.34	270.52	245.27
AIC	487.06	475.06	468.70	476.46	658.50	514.68	551.04	502.54
ΔAIC	18.36	6.36	0	7.76	155.96	12.14	48.50	0

Table 4.5.4: Model selection for water deer species. ΔAIC values of the water deer species split by methods: iREM with habitat, iREM with random effect and iREM.

	iREM with habitat				**iREM with random effect**				**iREM**			
	Poisson	NB	ZIP	ZINB	Poisson	NB	ZIP	ZINB	Poisson	NB	ZIP	ZINB
Speed data models												
gamma	58.66	0.14	17.92	0.86	7.88	7.62	1.00	7.20	192.90	49.08	85.32	50.54
lognormal	61.44	3.02	21.00	19.42	21.94	10.78	3.48	9.68	195.38	51.56	87.80	53.02
Weibull	58.50	0	17.78	4.14	19.32	11.04	0.86	7.06	192.76	48.94	85.18	50.40
ZAGA	211.98	8.17	9.68	0.22	11.10	0	7.26	4.50	184.54	40.72	97.08	28.68

4.5.3 *Estimated parameters of the muntjac species*

Table 4.5.5 presents the results from fitting a gamma model to the speed data. The results from fitting a lognormal model (C.4.5) and Weibull model (Table C.4.6) to animal speed data are given in appendix C.4. Note that the muntjac species was the only species observed to be moving throughout the period when they were followed around so as to derive an estimator of the speed of movement. Hence, there were no observed zero records.

Estimated density ($\hat{D} = 15$) from a Poisson iREM is better than the other count data models, with a minimal positive difference between estimated density and the density from the census (Table 4.5.5). There is a larger negative difference between estimated density and the density from the census from the other models. According to the ΔAIC values a NB iREM is the best model for the muntjac data but it gave the largest difference between estimated density and the density from the census.. The estimated camera random effect, $\hat{\sigma}_b$, is different from zero for all models, since an approximate 95% confidence interval does not contain zero.

The lognormal model (Table C.4.5, appendix C.4) and Weibull model (Table C.4.6, appendix C.4)gave similar estimations as the gamma model, with a NB iREM being the best model that explains the muntjac data set. A Poisson iREM does well in estimating the density but its AIC is almost 13 units higher than the AIC from a NB iREM.

Table 4.5.6 gives the ΔAIC values from the three iREM methods for the muntjac species. The ΔAIC values show that an iREM with habitat model best explains the muntjac data where animal speed is assumed to follow a Weibull model and the encounters are assumed to follow a NB REM model. But the difference between ΔAIC values from an iREM with habitat model and an iREM with random effect where animal speed is assumed to follow a Weibull model and encounters are assumed to follow a NB REM is minimal.

Table 4.5.5: Estimated parameters for the muntjac species from iREM with random effect models and iREM without random eefect models where animal speed data is assumed to follow a gamma distribution (standard error in parentheses).

	iREM with random effect				**iREM**			
	Poisson	NB	ZIP	ZINB	Poisson	NB	ZIP	ZINB
D = 13								
$\hat{D}$	15 (4.27)	4 (1.77)	6 (2.64)	5 (2.76)	8 (1.94)	8 (2.19)	23 (6.92)	16 (10.65)
$\hat{\mu}$	8.27 (1.76)	8.27 (1.76)	8.27 (1.76)	8.27 (1.76)	8.27 (14.70)	8.27 (14.80)	8.27 (14.77)	8.28 (14.98)
$\hat{\nu}$	0.27 (0.12)	0.27 (0.12)	0.27 (0.12)	0.27 (0.12)	0.27 (0.12)	0.27 (0.12)	0.27 (0.12)	0.26 (0.12)
$\hat{\sigma}_b$	2.17 (0.39)	1.90 (0.49)	1.86 (0.82)	1.89 (0.49)	-	-	-	-
$\hat{k}$	-	1.85 (1.59)	-	0.38 (0.73)	-	0.37 (0.15)	-	0.63 (1.17)
$\hat{\rho}$	-	-	0.24 (0.21)	0.09 (0.40)	-	-	0.67 (0.20)	0.52 (0.62)
$(-\ell)$	133.47	126.86	127.25	126.84	149.66	138.08	138.54	137.86
AIC	276.94	265.72	266.50	267.68	308.32	284.16	284.08	285.08
ΔAIC	11.22	0	0.78	1.96	24.24	0.08	0	0.92

Table 4.5.6: Model selection for muntjac species. ΔAIC values of the muntjac species split by methods: iREM with habitat, iREM with random effect and iREM.

	iREM with habitat				**iREM with random effect**				**iREM**			
	Poisson	NB	ZINB	ZIP	Poisson	NB	ZINB	ZIP	Poisson	NB	ZINB	ZIP
Speed data models												
gamma	3.90	13.50	0.92	2.92	11.34	0.12	2.08	9.72	41.72	20.56	20.48	21.48
lognormal	4.26	0.08	1.28	2.65	13.54	0.48	1.96	6.92	42.10	20.92	22.48	21.84
Weibull	4.26	0	1.28	1.98	11.68	0.02	1.95	10.06	42.08	20.92	22.46	21.82

4.5.4 *Estimated parameters of the mara species*

The results from a gamma model are given in Table 4.5.7, while the results from a lognormal model (Table C.4.7) and a Weibull model (Table C.4.8) are given in appendix C.4.

The results show that the difference in parameter estimates between the three speed data models is minimal (Table 4.5.7, Table C.4.7, appendix C.4) and Table C.4.8, appendix C.4). Also, the results suggest that it is relevant to account for the variation in the random placement of camera traps, the estimated $\hat{\sigma}_b$ is significant (i.e., zero is not contained within an approximate 95% confidence interval). Large estimates of κ from a NB iREM suggest that a Poisson iREM would be more appropriate. Note that as $1/\kappa$ $\rightarrow 0$ an NB iREM reduces to a Poisson iREM.

The ΔAIC values between the three methods iREM with habitat, iREM with random effect and iREM (Table 4.5.8) indicate that a Poisson iREM where animal speed is assumed to follow a gamma model is the best model that explains the mara species data set but the difference in ΔAIC values between a Poisson iREM and a Poisson iREM with habitat is minimal. The iREM is more supported when we account for the observed zero speeds.

Table 4.5.7: Estimated parameters for the mara species from iREM with random effect models and iREM without random eefect models where animal speed data is assumed to follow a gamma distribution, and where animal is assumed to follow a ZAGA model (standard error in parentheses).

	iREM with random effect (without zero speed)				**iREM (without zero speed)**			
	Poisson	NB	ZIP	ZINB	Poisson	NB	ZIP	ZINB
D = 68								
$\hat{D}$	2 (1.57)	2 (1.58)	2 (1.58)	2 (1.58)	4 (2.00)	4 (2.00)	4 (2.00)	4 (2.14)
$\hat{\mu}$	3.67 (1.56)	3.66 (1.55)	3.65 (1.55)	3.66 (1.55)	3.65 (11.26)	3.66 (11.29)	3.66 (11.30)	3.61 (10.97)
$\hat{\nu}$	0.22 (0.14)	0.22 (0.14)	0.22 (0.14)	0.22 (0.14)	0.22 (0.14)	0.22 (0.14)	0.22 (0.14)	0.22 (0.14)
$\hat{\sigma}_b$	1.35 (0.62)	1.34 (0.62)	1.34 (0.62)	1.34 (0.62)	-	-	-	-
$\hat{\kappa}$	-	3.5e+06 (4.1e+08)	-	1e-04 (8.5e-19)	-	1.1e+07 (2.4e+08)	-	0.001 (0.02)
$\hat{\rho}$	-	-	1e-05 (0.002)	1e-04 (0.002)	-	-	4e-05(0.01)	0.01 (0.18)
$\hat{\omega}$	-	-	-	-	-	-	-	-
$(-\ell)$	48.27	48.27	48.27	48.27	49.89	49.89	49.89	49.89
AIC	106.54	108.54	108.54	110.54	105.78	107.78	107.78	109.78
ΔAIC	0	2	2	4	0	2	2	4
	iREM with random effect (with zero speed)				**iREM (with zero speed)**			
	Poisson	NB	ZIP	ZINB	Poisson	NB	ZIP	ZINB
$\hat{D}$	3 (2.32)	3 (2.32)	2 (2.48)	3 (2.39)	5 (3.06)	5 (3.06)	5 (3.62)	5 (3.13)
$\hat{\mu}$	2.56 (1.21)	2.56 (1.21)	3.56 (0.84)	3.60 (1.80)	2.56 (2.89)	2.56 (2.89)	2.55 (2.88)	2.54 (131.62)
$\hat{\nu}$	3.66 (0.84)	3.66 (0.84)	3.65 (0.84)	5.72 (1.34)	3.66 (1.55)	3.66 (1.55)	3.63 (1.54)	3.63 (1.54)
$\hat{\sigma}_b$	1.34 (0.62)	1.34 (0.62)	1.34 (0.62)	1.34 (0.71)	-	-	-	-
$\hat{\kappa}$	-	7.8e+04 (2.25e+07)	-	4e-04 (0.001)	-	2.0e+05 (3.9e+08)	-	1.6e-8 (7.0e-23)
$\hat{\rho}$	-	-	3e-06 (5e-06)	0.06 (0.23)	-	-	0.03 (0.32)	0.003 (1.74)
$\hat{\omega}$	0.30 (0.21)	0.30 (0.21)	0.30 (0.21)	0.37 (0.24)	0.30 (0.21)	0.30 (0.21)	0.30 (0.21)	0.30 (0.21)
$(-\ell)$	54.37	54.37	54.37	54.37	56.00	56.00	56.00	56.00
AIC	120.74	122.74	122.74	124.74	120.00	122.00	122.00	124.00
ΔAIC	0	2	2	4	0	2	2	4

Table 4.5.8: Model selection for mara species. ΔAIC values of the mara species split by methods: iREM with habitat, iREM with random effect and iREM.

	iREM with habitat				**iREM with random effect**				**iREM**			
	Poisson	NB	ZINB	ZIP	Poisson	NB	ZINB	ZIP	Poisson	NB	ZINB	ZIP
Speed data models												
gamma	4.27	6.27	8.27	10.30	0.76	2.76	2.76	4.76	0	2.00	2.01	4.00
lognormal	5.12	7.12	9.12	11.22	2.06	3.70	3.70	5.70	0.95	2.95	2.95	4.97
Weibull	4.37	6.37	8.45	10.43	0.84	2.84	2.84	4.84	0.09	2.09	2.09	4.09
ZAGA	4.28	6.30	6.30	8.55	0.74	2.74	2.74	4.74	0	2.00	2.01	4.00

4.6 iREM with random effects and habitat

In this section we assess the potential relationship between abundance and habitat type and camera random effect. In Section 4.1.2 we introduced habitat-specific covariates in iREM, investigating the potential relationship between abundance and habitat type, and as discussed in that section, at WWAP there were four areas of contrasting habitats, p: Central Park, Downs, Institute Paddock, and Old Farm, where $p = 1, 2, 3, 4$, respectively. And, in each of these habitats there were four camera traps. In REM it is required that camera traps be placed randomly, and therefore, we expect that abundance estimation could depend on the location of the camera traps and habitat type.

Here we expand on the model structure in Section 4.4 to allow the encounter rate to depend on habitat type and camera random effect, using a log link function

$$\boldsymbol{\lambda} = \left(\frac{2+\theta}{\pi}\mu_x rt\right)\exp\{\mathbf{X}\boldsymbol{\beta} + b_i\}, \tag{4.6.1}$$

where $\boldsymbol{\lambda}$ is a vector of encounter rates for each camera day; b_i is a vector of random effects; the camera dimensions are r and θ; $t = 1$ (day) is the camera trap time period; μ is the expected animal speed. As discussed earlier in Section 4.1.2 the covariates, $\mathbf{X}$, are indicator variables representing each habitat and camera random effect. We arbitrarily set habitat 1 to be the null habitat and let

$$x_{jp} = \begin{cases} 1 & \text{observation } j \text{ is from habitat } p \\ 0 & \text{observation } j \text{ is not from habitat } p, \end{cases}$$

so that for other habitats

$$\mathbf{X} = \begin{bmatrix} 1 & x_{12} & x_{13} & x_{14} \\ 1 & x_{22} & x_{23} & x_{24} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n2} & x_{n3} & x_{n4} \end{bmatrix}_{\mathbf{n\times 4}}$$

The vector $\boldsymbol{\beta}_{1\times 4} = (\beta_1, \beta_2, \beta_3, \beta_4)^T$ contains the regression coefficients, and the density

becomes $\exp\{\mathbf{X}\boldsymbol{\beta} + b_i\}$.

Suppose encounters, y_{ij}, for jth $(= 1, 2, ..., n)$ day on the ith $(= 1, 2, ..., c)$ camera trap, has probability mass function $h(y_{ij};\ \boldsymbol{\lambda})$, where $\boldsymbol{\lambda}$ is a vector of encounter rates defined in equation (4.6.1). And, suppose there are m animal speed observations, such that $x = \{x_1, ..., x_m\}$, has probability density function $f(x_l \mid \mu,\ \nu)$, , where $l = 1, 2, ..., m$, and μ is the expected speed and ν represents any additional parameter in the model. Then using Gauss-Hermite quadrature discussed in Section 4.4.2 to approximate the marginal likelihood of the encounter data, the joint log-likelihood function is

$$\ell = \sum_{i=1}^{c} \log \sum_{q=1}^{Q} \prod_{j=1}^{n_i} \frac{w_q h_{ij}(\sqrt{2}\sigma_b d_q)}{\sqrt{\pi}} + \sum_{l=1}^{m} \log\ f(x_l \mid \mu,\ \nu), \tag{4.6.2}$$

where

$$h_{ij}(\sqrt{2}\sigma_b d_q) = h(y_{ij} \mid b_i = \sqrt{2}\sigma_b d_q;\ \boldsymbol{\lambda})$$

is the conditional density function for the encounters, evaluated at a camera random effect. The symbol d_q represents the quadrature evaluation points and w_q are the quadrature evaluation weights where we set $Q = 20$ as recommended by Choquet and Cole (2012) and Cole et al. (2003), which we discussed in Section 4.4.2. The assumptions of the models discussed in Section 4.1.2 and Section 4.4.2 are held here.

4.6.1 Application of iREM with random effect and habitat to WWAP data set

In this section we illustrate the application of iREM with random effect and habitat to WWAP data set. For illustration we fit a Poisson iREM with habitat and random effect to the data assuming the speed data follows a gamma model, a lognormal model or a Weibull model. We give results for the wallaby species and water deer species.

It is sometimes not possible to make meaningful inference about population abundance using classical methods due to the inability to estimate, or estimate well, all the parameters of a model. This may be because of lack of data or in some instances, parameters may be confounded and only ever appear as a product. In such cases a model is termed non-identifiable or parameter-redundant (see Cole et al., 2010). According to

Cole and McCrea (2016) and Cole and Morgan (2010) models are said to be parameter-redundant when they contain too many parameters to be estimated, however much data are collected, so that using classical inference it would not be possible to estimate all the original parameters. In practice, a model that is parameter-redundant will cause problems with the estimation of parameters, because the likelihood surface will not possess a unique maximum and the standard errors will not exist (Cole et al., 2014).

There are several approaches to determine whether a model is parameter-redundant. But one approach is the use of a symbolic algebra computer package, which involves forming a suitable derivative matrix and then calculating its symbolic rank (Cole and Morgan, 2010; Cole et al., 2010). With this approach, a model will be parameter-redundant if and only if this derivative matrix is singular, which occurs if and only if the row rank of the derivative matrix is less than the number of parameters. If a model is parameter-redundant then the question of which (if any) of the parameters are estimable is solved by considering components of the eigenvectors of the derivative matrix corresponding to the zero eigenvalues (Catchpole et al., 2001).

In our model, if there was only one camera per habitat this model would be parameter redundant since there would not be enough distinct information to separate out the fixed effects (habitat) from the the camera random effect. However, there are actually 4 cameras in each habitat, giving a total of 16 camera traps but the model is near redundant (near-singular). A model is near-singular if the smallest eigenvalue of the expected information matrix is small rather than zero (see Catchpole et al., 2001).

The data sets for the muntjac and mara species are rather small and the likelihoods from the data for all four species are fairly flat with various local maxima, which provides a poor fit of the model. For the wallaby species and water deer species, however, there is enough data available for model fitting, and it is still possible for the maximum likelihood to be maximized on a nearly flat ridge or plane (Catchpole et al., 2001). Therefore, for the purpose of analyse we fit an iREM with random effect and habitat model to the wallaby and water deer species. We compute the eigenvalues of the Hessian matrix,

which shows that the smallest eigenvalue is small (non-zero), hence, our model is near-singular (near parameter-redundant).

Table 4.6.1 gives the parameter estimates of the wallaby species from fitting a Poisson iREM with random effect and habitat, while Table 4.6.2 gives the parameter estimates of the water deer species. We also give the standardized eigenvalues (all eigenvalues divided by largest eigenvalue), which we denote by ξ_s for the estimated regression coefficients of the habitats, estimated random effect and estimated expected speed. Clearly, fitting a Poisson iREM with random effect and habitat model to the data sets of both species produces very poor results of the density within and across habitats. For both species, the results show that some standard errors are inestimable. Also, we expected the estimated random effect, $\hat{\sigma}_b$ from iREM with habitat and random effect to be smaller than those estimated from the less complicated iREM with random effect model, as in Tables 4.5.1 and 4.5.3 above, since some of the variation in density within habitats would be mopped up by the camera random effects (Gschlößl and Czado, 2008). But this is not case for the wallaby species (Table 4.6.1). Also, the eigenvalues are "small" to the smallest eigenvalue is $\xi_s = 0.001$ for the wallaby species and $\xi_s = 0.002$ for the water deer species, suggesting that the model is a near parameter-redundant model.

The ΔAIC values for the wallaby species are given in Table 4.6.3. We compare a Poisson iREM with random effect and habitat with other models discussed so far in this Chapter and Chapter 3. The results show that a NB iREM with habitat where animal speed is assumed to follow a lognormal model is the best model with the lowest AIC value. It is worth noting that a NB iREM with random effect and habitat would be impossible to fit. We also provide the ΔAIC values across all the models fitted in Chapter 3, and in the previous sections of this Chapter for the water deer species. These values are given in Table 4.6.4. The results show, in this case, that a Poisson iREM with random effect and habitat where animal speed is assumed to follow a gamma model is the best model, but we note that the difference between the minimum AIC value and the AIC value from a NB iREM with habitat where animal speed is assumed to follow a gamma model or a Weibull model is minimal. This suggests that a NB iREM with habitat would work well in practice, and it also a simpler model to implement.

Table 4.6.1: Estimated regression coefficients, $\hat{\beta}_p$ for each habitat, density, $\hat{D}_p$, and standardized eigenvalues of all parameters, ξ_s of the wallaby species. The standard errors are in parentheses.

	Habitat					Estimated mean speed	Estimated random effect
	Central Park	Downs	Institute Paddock	Old Farm	Total density	$\hat{\mu}$	$\hat{\sigma}_b$
Census (D, animals km^2)	96	1101	760	803	468		
gamma							
$\hat{\beta}_p$	5.20 (2.18)	0.82 (0.16)	1.01 (0.20)	0.64 (0.13)	-	0.88 (0.33)	0.74 (0.08)
$\hat{D}_p$	180.64 (75.76)	408.47 (249.21)	496.55 (304.11)	342.29 (211.99)	285.63 (110.71)	-	-
lognormal							
$\hat{\beta}_p$	5.13 (2.97)	0.79 (0.15)	1.52 (0.31)	1.13 (0.22)	-	0.94 (0.51)	0.77 (0.08)
$\hat{D}_p$	168.31 (97.40)	371.70 (285.73)	771.66 (602.54)	522.78 (405.84)	323 (178.20)	-	-
Weibull							
$\hat{\beta}_p$	5.69 (2.18)	-0.14 (-)	1.04 (0.21)	0.19 (0.04)	-	0.87 (0.35)	0.72 (0.09)
$\hat{D}_p$	297.24 (130.41)	259.47 (-)	842.86 (536.65)	359.06 (241.98)	362.94 (149.23)	-	-
	Standardized eigenvalues, ξ_s						
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\mu}$	$\hat{\sigma}_b$	
gamma	1	0.112	0.096	0.031	0.004	0.001	
lognormal	1	0.168	0.102	0.041	0.018	0.001	
Weibull	1	0.180	0.09	0.020	0.008	0.002	

Table 4.6.2: Estimated regression coefficients, $\hat{\beta}_p$ for each habitat, density, $\hat{D}_p$, and standardized eigenvalues of all parameters, ξ_s of the water deer species. The standard errors are in parentheses.

	Habitat					Estimated mean speed	estimated random effect
	Central Park	Downs	Institute Paddock	Old Farm	Total density	$\hat{\mu}$	$\hat{\sigma}_b$
Census (D, animals km^2)	72	73	36	577	119		
gamma							
$\hat{\beta}_p$	4.44 (2.45)	-0.89 (-)	-2.87 (-)	0.70 (0.21)	-	1.95 (0.77)	0.71 (0.16)
$\hat{D}_p$	84.69 (46.74)	34.89 (-)	4.81 (-)	170.72 (145.37)	72.75 (36.13)	-	-
lognormal							
$\hat{\beta}_p$	3.93 (2.45)	-0.87 (-)	-2.88 (-)	0.71 (0.21)	-	3.23 (2.77)	0.71 (0.16)
$\hat{D}_p$	50.66 (47.58)	21.15 (-)	2.84 (-)	103.16 (127.51)	43.68 (39.67)	-	-
Weibull							
$\hat{\beta}_p$	4.53 (2.35)	-0.66 (-)	-2.37 (-)	0.98 (0.25)	-	1.54 (0.65)	0.69 (0.15)
$\hat{D}_p$	93.05 (48.28)	47.91 (-)	8.71 (-)	248.10 (192.77)	88.59 (42.47)	-	-
	Standardized eigenvalues, ξ_s						
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\mu}$	$\hat{\sigma}_b$	
gamma	1	0.121	0.063	0.026	0.005	0.004	
lognormal	1	0.122	0.073	0.058	0.014	0.002	
Weibull	1	0.116	0.06	0.025	0.02	0.005	

Table 4.6.3: Model selection for wallaby species. The ΔAIC values for the four methods: iREM with random effect and habitat, iREM with habitat, iREM with random effect and iREM are given here. We fit a Poisson iREM with random effect and habitat and compare the AIC values with other models fitted in the earlier sections of this Chapter and Chapter 3.

	iREM with habitat and random effect			**iREM with random effect**			**iREM with habitat**			**iREM**		
	gamma	lognormal	Weibull	gamma	lognormal	Weibull	gamma	lognormal	Weibull	gamma	lognormal	Weibull
Count data models												
Poisson												
(-ℓ)	466.42	465.06	468.21	550.51	549.93	551.44	556.33	555.75	556.25	594.59	593.99	594.49
AIC	948.42	946.12	952.42	1113.02	1109.86	1112.88	1124.66	1123.00	1124.50	1197.18	1195.98	1196.68
ΔAIC	169.98	167.68	173.98	330.58	329.42	332.44	346.22	345.06	346.22	416.70	415.54	416.54
NB												
(-ℓ)	-	-	-	389.87	389.28	389.79	382.80	**382.22**	384.05	390.42	389.84	390.34
AIC	-	-	-	791.74	790.56	791.74	779.60	778.44	782.10	790.84	789.68	790.68
ΔAIC	-	-	-	11.30	10.12	9.14	1.16	**0**	3.66	10.40	9.24	10.24
ZIP												
(-ℓ)	-	-	-	517.91	516.91	517.42	526.48	525.13	525.66	545.32	544.74	545.25
AIC	-	-	-	1049.82	1048.82	1058.84	1066.96	1064.26	1065.32	1100.64	1099.48	1100.50
ΔAIC	-	-	-	267.33	265.38	266.40	288.52	282.82	286.88	320.20	319.04	320.06
ZINB												
(-ℓ)	-	-	-	389.79	389.28	389.71	382.42	382.22	383.09	390.42	389.84	390.34
AIC	-	-	-	793.59	792.56	793.42	780.84	780.44	782.16	792.84	791.68	792.68
ΔAIC	-	-	-	11.30	10.12	3.58	6.40	2.08	3.74	12.40	11.24	12.78

Table 4.6.4: Model selection for water deer species. The ΔAIC values for the four methods: iREM with random effect and habitat, iREM with habitat, iREM with random effect and iREM are given here. We fit a Poisson iREM with random effect and habitat and compare the AIC values with other models fitted in the earlier sections of this Chapter and Chapter 3.

	iREM with habitat and random effect			**iREM with random effect**			**iREM with habitat**			**iREM**		
	gamma	lognormal	Weibull	gamma	lognormal	Weibull	gamma	lognormal	Weibull	gamma	lognormal	Weibull
Count data models												
Poisson												
(-ℓ)	**216.78**	218.02	218.22	225.01	232.04	230.73	248.40	249.79	248.32	318.52	319.76	318.45
AIC	449.56	452.04	452.44	460.02	474.08	471.46	1124.66	1123.50	1124.50	645.04	647.52	644.90
ΔAIC	**0**	2.48	2.88	10.46	24.52	21.90	59.24	62.02	59.08	195.48	197.96	195.34
NB												
(-ℓ)	-	-	-	223.88	225.46	225.59	218.14	219.58	248.32	245.61	246.85	245.54
AIC	-	-	-	459.76	462.92	463.18	450.28	453.16	508.64	501.22	503.70	501.08
ΔAIC	-	-	-	10.20	13.36	13.62	0.72	3.60	0.58	51.66	54.14	51.52
ZIP												
(-ℓ)	-	-	-	220.57	221.81	220.50	227.03	228.57	226.96	263.73	264.97	263.66
AIC	-	-	-	453.14	455.62	453.00	468.06	471.14	467.92	537.46	539.94	537.32
ΔAIC	-	-	-	3.58	6.06	3.44	18.50	21.58	18.36	87.90	90.38	87.76
ZINB												
(-ℓ)	-	-	-	222.67	223.91	222.60	218.14	219.38	217.42	245.34	246.58	245.27
AIC	-	-	-	459.34	461.82	459.20	452.28	454.76	450.84	502.68	505.16	502.54
ΔAIC	-	-	-	9.78	12.26	9.64	2.72	5.20	1.28	52.60	55.60	52.98

4.7 Discussion

In this Chapter we have extended the integrated Random Encounter model developed in Chapter 3 to incorporate covariates which have some potential relationship with animal abundance. In Chapter 2 we showed how REM can be extended to incorporate habitat-specific covariates by modelling the encounter data and using a fixed estimate of animal speed. Our model (iREM) builds on REM accounting for the variation in travel speed. We have demonstrated that incorporating habitat-specific covariates into iREM can be relevant when estimating abundance. We have also shown how disregarding zero-inflation and variation in encounter data across habitats can induce bias in the density estimator and an underestimation of the standard error of the density. We have proposed models with habitat-specific covariates for animals moving in groups and animals with observed zero travel speed. One issue with the proposed model for animals moving in groups is that less precise estimates are obtained for animals with lower expected speeds. However, our simulations indicated that precise estimates can be obtained for faster moving species with large sample sizes. But it is worth noting that, as shown in Chapter 3 with the analysis of the data set for species moving in pairs of family groups, the difference between estimated density using the iREM with group size data and estimated density from REM where the density is obtained by multiplying density of groups by an independent estimate of average group size is minimal.

The application of iREM with habitat to the WWAP data showed how relevant it is to incorporate habitat as non-zero estimates of the density were obtained in habitats which had zero values for the density from the census. For three of the four species, there was support for models which allow for variation in the encounter data but the parameter estimates from the alternative models show minimal differences. A Poisson iREM with habitat was shown to be more appropriate for the species with limited data set, for example the mara species. As in Chapter 3, this Chapter also highlights the importance of accounting for the observed zero speed of movement data. Not accounting for these zeros resulted in larger estimates of the expected speed, and hence, induced negative bias in the density estimator.

When investigating the use of iREM with covariates to analyse data collected from camera traps, the question of whether camera random effect would influence abundance estimation was raised. We therefore extended iREM to incorporate camera random effect. To deal with the problem of the intractable integral in the model component for the encounter data, we used Gauss-Hermite quadrature to approximate the marginal likelihood of the encounters and maximize the likelihood. One limitation of the proposed iREM with random effect model using Gauss-Hermite quadrature to maximize the likelihood is that it is slightly demanding in terms of computing time, owing to the large number of quadrature points needed to obtain a close approximation to the likelihood. For example, fitting an iREM with Poisson with random effect model with 20 quadrature points to the WWAP data set took in our implementation around 1 minute (on a 2.20 GHz computer) compared with a few seconds required by an iREM with Poisson model without random effect. In the case of the simulation procedure with sample sizes of 100 camera trapping days, $m = 100$ animal speed data and 100 simulation runs took around 50 minutes compared with a several seconds (under 1 minute) required by an iREM with Poisson without random effect. Our simulation results indicated that precise estimates of the density can be obtained and an iREM with random effect performs better than an iREM without random effect.

Despite the limitation of the model our analysis of the WWAP data provided an illustration of how accounting for the random location of camera traps can affect abundance estimation. In camera trapping analysis a potential relationship exists between abundance and habitat type and location of the camera trap within the habitat. We examined this relationship by fitting an iREM with random effect and habitat model to real data. This model is a further extension of the iREM. Given the greater complexity of the iREM with random effect and habitat model, which unlike the iREM with habitat models data in a particular habitat on a given camera trap, it is understandable that model fitting becomes increasingly difficult as the complexity increases, particularly because of the number of Gauss-Hermite quadrature points required to approximate the integral, and for species with large number of data points. Also, the complexity of model implies that, unless enough data are available the likelihood can be rather flat with various local maxima, which can be of similar magnitude while providing different

explanations for the data. This means that the outcome of the analysis can be sensitive to the point from which the optimization routine is initialized and that, therefore, it is a good approach to explore the likelihood function by trying out different starting values. In our WWAP data analysis we encountered this problem: some of the standard errors were inestimable, and the some of the parameters were poorly estimated, with relatively small eigenvalues, which suggested that the model is near parameter redundant.

Finally, when applying the models proposed in this Chapter, it is important to remember the appropriateness of interpreting the estimates obtained as abundance estimates is contingent on how well the model assumptions are met. First of all Rowcliffe et al. (2008) REM assumes that placement strategies of camera traps are random, and the movement of animals are independent of camera traps. For instance, we have seen that the violation of the randomized placements of camera traps can cause an underestimation of the density, in particular for the mara species. An appropriate method of randomization of camera traps is therefore critical for reliable estimates of density. Secondly, the underlying distribution for encounters in Rowcliffe et al. (2008) REM is a Poisson model. We have seen in iREM and its extension that not accounting for variation in the encounter data can cause bias in the density estimator and an underestimation of its standard error. If some measurable factors are thought to affect encounter rates, and hence, density appreciably, these can and should be incorporated into the model as covariates. Thirdly, the models assume that all camera traps are located in clear view of the area to capture animals. Whether this assumption is satisfied depends on the characteristics of the study site, and the size of the species. Fourthly, animals are assumed to move about randomly, and for long distances in straight-line. While for all species this would be an unusual behaviour, Hutchinson and Waser (2007) have shown that violation of this assumption does not change the expected encounter rate, and a more variable distribution can be used to model encounters, as we have shown by the possible cases of a NB REM, a ZINB REM or a ZIP REM.

Chapter 5

iREM With Detection Zone Dimensions

Prior to this Chapter the detection zone dimensions of the camera were assumed to be fixed. In practice, the camera dimensions vary (Rowcliffe et al., 2011), and as we have shown in the previous Chapters it is crucial to account for any sampling variability of the estimators of parameters associated with the density in REM to ensure accurate treatment of precision and correlation in the estimators. The motivation for this Chapter is the Barro Colorado Island (BCI) data set, which includes detection distance data and angular data collected from the camera traps. In this Chapter we develop a single integrated likelihood to estimate animal density, which builds on iREM from Chapter 3, and which includes models for detection distance and detection angle. Our approach, which adopts distance sampling methodology (Buckland et al., 2001), provides a comprehensive framework for abundance estimation of unmarked animals.

This Chapter begins with a discussion of distance sampling theory and its application in Section 5.1. A description of the detection zone for camera traps is given in Section 5.2. The BCI data set, which is the motivating factor for this Chapter is given in Section 5.3. This data set is different from the Whipsnade Wild Animal Park data given in earlier Chapters since it was collected from a tropical moist forest at Barro Colardo Island (BCI), Panama and census count is unknown. The BCI data set is used in the remaining core Chapters of this thesis. In Section 5.4 we propose a model for the detection distance data and provide estimates of the effective detection distance for the species at BCI, Panama. To account for the variation in the detection distance we develop an integrated REM with detection distance (iREM-dd), which is given in Section 5.5.

The model is an extension of iREM developed in Chapter 3, which provides a structure for estimating abundance accounting for the variation in detection distance data. We test iREM with detection distance model (iREM-dd) via simulations in Section 5.6. An application of iREM-dd to the BCI data is given in Section 5.7.

It is also likely that the variation in the angle of approach in which the animals enter the camera trap detection zone would have an effect on estimated density. In Section 5.8 we consider a model relevant for circular or angular data, which we test via simulations. For illustration, this model is applied to angular data at BCI, Panama. The Chapter goes on, in Section 5.9, to give a single integrated likelihood to estimate abundance of unmarked animals. This model is an extension of the integrated REM with detection distance, which incorporates the angle to detection data. The integrated Random Encounter with detection zone dimensions (iREM-D) merges four independent steps required in REM to estimate density into a comprehensive framework for abundance estimation using camera trap data. In the simulation study, in Section 5.10, the performance of iREM-D and the importance of accounting for variation in the detection zone dimensions are investigated. Finally, the Chapter concludes with an application of iREM-D to the data set of a community of terrestrial mammals at BCI, Panama in Section 5.11.

5.1 Distance Sampling and its application

It is rarely practical to count all the objects of interest in a large area. In practice, we sample, counting the objects in a few small areas called plots and estimating the density of these objects. It is the idea of these plots on which distance sampling is based. Distance sampling is a widely used methodology for estimating animal density or abundance by using measured or estimated distances. There are various analysis methods used in distance sampling to model these measured or estimated distances. One such method is point transects (Buckland et al., 2001; Borchers et al., 2002), which is mostly used in avian surveys, and which, for example, has been used to estimate density of small mammals and whales (Smith et al., 1975; Royle et al., 2004; Marques et al., 2010). Point transect sampling, which is discussed in detail in Section 5.1.1, involves an observer standing at a given position and counting the number of objects

seen (Buckland et al., 2001). Point transects are similar to camera traps, which are fixed at a given position, recording images of passing animals.

5.1.1 Point transect sampling

In point transect sampling an observer visits a number of points, which are determined by a random design (Marques et al., 2010). By recording from the point, the observer can concentrate on detecting the animals (or plants or species) of interest and recording the distances to the detected animals. Figure 5.1.1 shows an example of point transect sampling with five randomly selected points (open circles), detected animals and also undetected animals. The circles (•) represent the positions of the animals. In this example 11 animals are detected, which are represented by the (•) connected to the lines (—) (see Buckland et al., 2001).

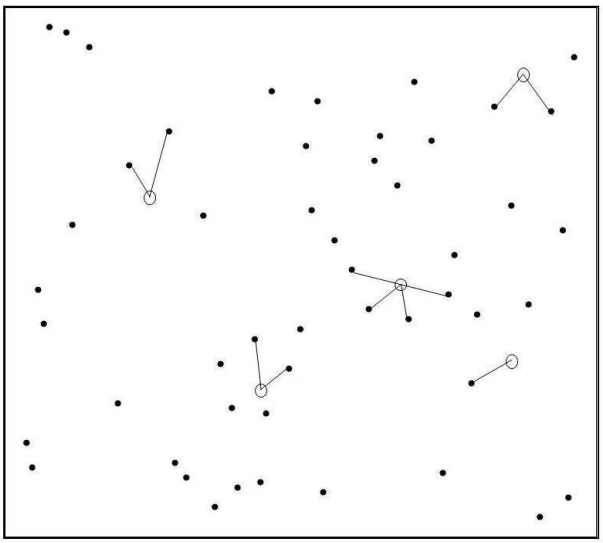


Figure 5.1.1: Point transect sampling approach with five randomly spaced points ($k = 5$), denoted by the open circles where eleven objects were detected (Buckland et al., 2001).

Point transects are a generalisation of traditional circular plot surveys. Consider k circular plots randomly positioned within the survey area, and that objects further than some distance w from a point are not recorded. Then the surveyed area is $a = k\pi w^2$, within which n objects are detected. By definition, the density D is the number per

unit area, which can be estimated by

$$\hat{D} = \frac{n}{k\pi w^2}. \tag{5.1.1}$$

The assumption of these k points is that they are far enough apart such that the areas do not overlap. In point transect sampling, however, only the area close to the random point can be fully censused; a proportion of objects away from the random point but within the survey area remains undetected, therefore, the density D can no longer be estimated by equation (5.1.1). If we let P_a be the probability that a randomly chosen object within the surveyed area is detected, and suppose an estimate $\hat{P}_a$ is available. Then, the density D in point transect sampling can be estimated by

$$\hat{D} = \frac{n}{k\pi w^2 \hat{P}_a}. \tag{5.1.2}$$

Buckland et al. (2001) showed that P_a can be reformulated as a statistical estimation problem by considering an annulus of width dz at distance z from a point (Figure 5.1.2).

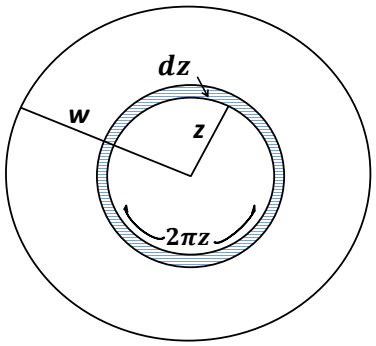


Figure 5.1.2: An annulus of width dz and distance z from a point. The annulus is divided up into an infinite number of annuli of of infinitesimal width dz, so we can compute the area of the annulus by taking the difference between the area of the outer circle with radius $z + dz$ and the inner circle with radius z, i.e., $A = \pi(z + dz)^2 - \pi z^2 = \pi z^2 + 2\pi z dz + \pi(dz)^2 - \pi z^2$. Since $(dz)^2$ is the square of an infinitesimal quantity, it can be taken as zero, hence an approximate area of the annulus is $2\pi z dz$ (see Buckland et al., 2001).

The proportion of the circle of radius w that falls within this annulus is

$$\frac{2\pi z dz}{\pi w^2},$$

which is also the expected proportion of objects in the circle that lie within the annulus. Hence, the proportion of objects in the circle that are both in the annulus and detected is

$$g(z) \times \frac{2\pi z dz}{\pi w^2},$$

where $g(z)$ represents the probability of detecting an object, given it is at distance z from the point. The unconditional probability of detecting an object that is in one of the k circular plots is

$$\begin{aligned} P_a &= \int_0^w \frac{2\pi z g(z) dz}{\pi w^2} \\ &= \frac{2}{w^2} \int_0^w z g(z) dz, \end{aligned} \tag{5.1.3}$$

Substituting equation (5.1.3) into equation (5.1.2) and cancelling the w^2 terms, the estimator of the density is

$$\hat{D} = \frac{n}{2k\pi \int_0^w z\hat{g}(z) dz}. \tag{5.1.4}$$

Defining

$$v = 2\pi \int_0^w z g(z) dz, \tag{5.1.5}$$

which is the critical quantity to be estimated from the distance data, z_j $(j = 1, 2, ..., s)$ for a point transect survey, then

$$\hat{D} = \frac{n}{k\hat{v}}. \tag{5.1.6}$$

Note that in a sampled plot it may not be possible to detect all animals within this plot. In fact the further an animal is from the observer the less likely the observer will record it. Earlier literature such as Gates et al. (1968) assumed that the detection of animals is "spiked" close to the transect, and quickly decreases away from the transect. This suggests that the distribution of objects detected is exponential, and inflexible. However, Eberhardt (1968) suggests that the distribution of the detected animals would change from survey to survey, and conceptualized a fairly general model where probabilities of detection decreases with increasing distances from the transect. Therefore, it is usually assumed that the probability of detection is 1 for objects on the transect, that is,

$g(0) = 1$, and the detection function $g(z)$ is decreasing with increasing distance, and $0 \leq g(z) \leq 1$ always (Buckland et al., 2001). Figure 5.1.3 shows the general detection function, which decreases with increasing distances.

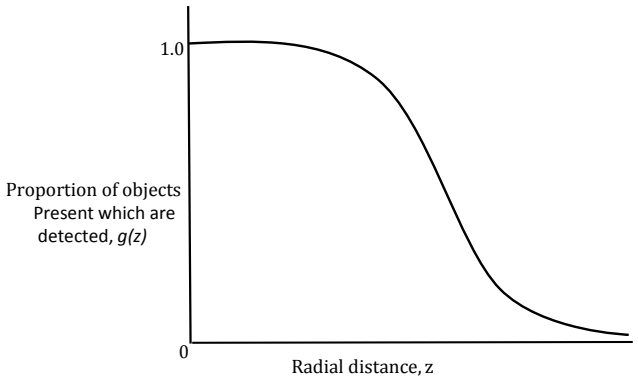


Figure 5.1.3: $g(z)$ equals the probability of detecting an object, given that it is at radial distance z from the point. It is usually monotonic decreasing and assumed to start at 1. (Buckland et al., 2001).

In point transect sampling, the relationship between the detection function, $g(z)$ and the probability density function of the distances of the detected objects, $f(z)$, can be derived from Figure (5.1.2) above. The area of the ring of incremental width dz at distance z from the observer is proportional to z. Thus, $f(z)$ is proportional $z \cdot g(z)$. With the constraint that $f(z)$ integrates to unity, the probability density function of the distances is defined as

$$f(z) = \frac{zg(z)}{\int_0^w zg(z)dz}, \tag{5.1.7}$$

(see Buckland et al., 2001).

Sometimes, it is not possible to sample the entire area of interest because of an obstruction of some kind. When this happens, a modification of the formula in equation (5.1.2) is required to estimate the density. If, for instance, there is an obstruction that obscures the visibility in part of the circle, the full circle will not be surveyed. This is demonstrated in Figure 5.1.4, which shows a circular plot with an obstruction. Due to the obstruction a sector with angle ψ is not visible to the observer. Therefore, any detections in front of the obstruction that fall within this sector are disregarded (Buckland et al., 2001).

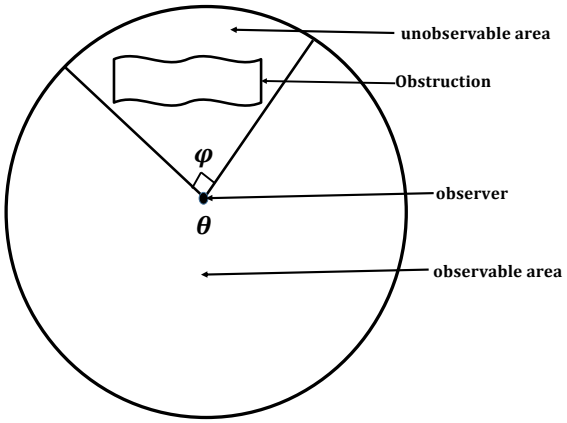


Figure 5.1.4: In point transect sampling, obstructions can reduce the surveyed area around a point. Buckland et al. (2001) suggest that a simple solution to this problem is to define a sector of angle ψ that span the obstructed region, and disregard any detections in front of the obstruction that fall within this sector. By excluding the detections within the sector shown here, and calculating the area about the point as $(\pi - \psi/2)w^2$, density can be estimated.

The contribution to the surveyed area from this circle is given below.

$$\begin{aligned} \text{The area of a circle} &= \pi w^2, \\ \text{therefore, the area of a sector} &= \frac{\theta}{2\pi}\pi w^2. \\ \text{Given that } \theta &= 2\pi - \psi, \\ \text{the area of the sector} &= \frac{2\pi - \psi}{2} w^2 = (\pi - \psi/2)w^2. \end{aligned}$$

The density is then estimated by

$$\hat{D} = \frac{n}{k(\pi - \psi/2)w^2 \hat{P}_a}, \tag{5.1.8}$$

which is equivalent to 5.1.1, and where $\psi = 2\pi - \theta$; n is the number of objects detected; and w is the known radius. The density is then

$$\hat{D} = \frac{n}{k(\pi - \pi + \psi/2)w^2 \hat{P}_a} = \frac{n}{k\psi w^2 \hat{P}_a/2}, \tag{5.1.9}$$

where $\hat{P}_a$ is equation (5.1.3) defined above (see Buckland et al., 2001).

5.1.1.1 *Effective Radius in Point Transect Sampling*

As only a proportion of animals will be detected from a sampled point, it is important for the observer to determine an effective detection distance, such that the number of animals missed within this distance exactly equals that recorded beyond this distance. Figure 5.1.5 shows the effective detection distance, where the number of animals outside this distance equals the number of animals missed within this distance.

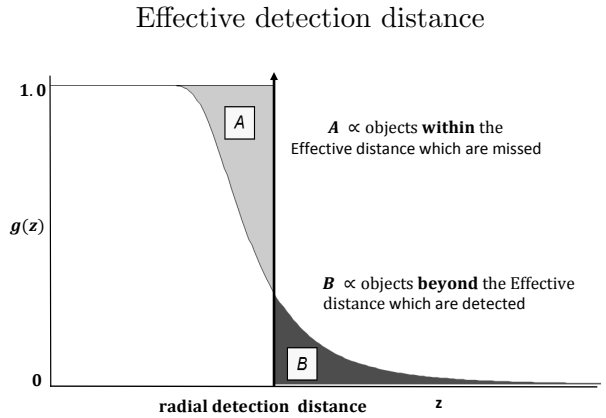


Figure 5.1.5: Effective detection distance from the point (Rowcliffe et al., 2011).

In point transect sampling an estimate of the effective detection distance, say $\hat{\gamma}$ can be computed from radial distances z_j $(j = 1, 2, ..., s)$, which can then be used in the estimation of animal density. As shown in equation (5.1.2), the density can be estimated as

$$\hat{D} = \frac{n}{k\pi\hat{w}^2\hat{P}_a}. \tag{5.1.10}$$

So replacing $\hat{w}^2$ with $\hat{\gamma}^2$ we have

$$\hat{D} = \frac{n}{k\pi\hat{\gamma}^2\hat{P}_a}, \tag{5.1.11}$$

which is analogous to equation (5.1.6) and where

$$\pi\hat{\gamma}^2\hat{P}_a = \hat{v}$$

as defined in equation (5.1.5), that is,

$$\hat{D} = \frac{n}{k\hat{v}}. \tag{5.1.12}$$

It is this effective detection distance (γ) that is required in REM to estimate animal density. Buckland et al. (2001) define the effective detection distance (radius) as

$$\gamma = \sqrt{\frac{v}{\pi}}, \tag{5.1.13}$$

which can be estimated by

$$\hat{\gamma} = \sqrt{\frac{2}{\hat{h}(0)}} \tag{5.1.14}$$

where $\hat{h}(0) = \lim_{z \to 0} \hat{f}(z)/z = 2\pi/\hat{v}$. Here $\hat{h}(0)$ is the slope of estimated density $\hat{f}(z)$ of the observed detection distances evaluated at $z = 0$. In the next section we give the assumptions required in distance sampling theory. The validity of the assumptions allows the investigator assurance that valid inference can be made concerning the density of the population sampled.

5.1.2 Assumptions in distance sampling theory

Buckland et al. (2001) suggest that the survey must be competently designed and conducted in order to make valid statistical inferences. The first three assumptions are critical to achieving reliable estimates of the density from the point transect sampling.

1. objects that are very close to the transect will always be detected
2. objects are detected at their initial location
3. all distances are measured without error
4. sightings of different objects are independent events
5. transects are placed randomly or systematic random
6. there is no movement of objects in response to the observer and none are counted twice
7. the survey is a "snapshot"

5.2 *Detection zone in camera trapping*

In camera trapping analysis,"detection" by an observer is replaced by animals being caught in camera traps (Buckland et al., 2001). These camera traps are fixed at a completely random location in the field, and detect moving animals through infra-red sensors. Unlike point transect sampling, camera traps are set up for a longer period of time, and can take "snapshots" and video recordings of the same animal multiple times. From these snapshots and video recordings, the observer can determine the radial detection distance of the animal from the trap when it first triggers the motion sensor of the camera, as well as the angle of detection. Figure 5.2.1 gives an example of the detection zone and the locations of animals when they triggered the camera sensors, and other animals captured in the camera field view.

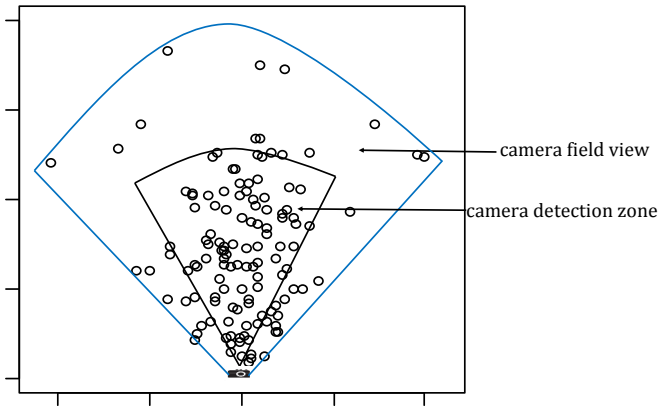


Figure 5.2.1: Camera detection zone and camera field view with animals being captured. Note that the detection zone is not necessarily equal to the camera field view.

The detection zone is the area in which a camera trap is able to detect moving animals through its sensor. This detection zone is not necessarily equal to the camera field view, that is, the area included in the actual photograph. The detection zone varies depending on the type of camera trap; and for some, only a small proportion of the field view actually corresponds to the cameras' detection zone. Camera traps can sense through either a conical shaped detection zone or through a combination of horizontal bands and vertical axis zones, depending on the technical specifications of the Passive-Infra-red (PIR) sensor (see Rovero et al., 2013). Note that whether a camera trap has

a narrow detection width, i.e., smaller than the camera field view, or a large detection zone, the further away the animal is from the trap the less likely it would be detected. That is, the probability of detection decreases with increasing distance from the trap. Hence, an effective detection distance and angle is required to maximize trapping rates.

5.2.1 Effective detection zone in REM

Given that camera detection zones are roughly triangular, with the camera at one corner with a maximum detection width of $\pi/2$ (see Chapter 2), animals can approach the camera closely before being detected. This suggests that a distance-based approach is possible to provide reasonable estimates in practice. Rowcliffe et al. (2011) used a distance-based approach to estimate the detection dimensions for application in REM. A critical step was to determine whether the detection distance and detection angle are independent or whether they are correlated, as the outer edges of the detection zone becomes weaker at greater distances. Testing their method on the species with the most records at BCI (agouti species), they found that a weak correlation exists between the detection distance and angle, which was driven by a small number of extreme records. Given that such extreme records are generally considered lacking in information and are frequently truncated in distance analyses, Rowcliffe et al. (2011) concluded that the correlation is not important in practice. As such, they assumed that the detection distance and angle are independent of one another, therefore, the declining detectability patterns in camera trapping can be modelled using standard distance sampling theory. Rowcliffe et al. (2011) used a line transect model for detection angle and a point transect model for detection angle assuming two possible models halfnormal, and hazard rate for describing detection probability as a function of distance and angle, respectively. These models were fitted to the distance or angle data and the likelihoods were maximized. Using standard distance sampling theory, the effective detection angle or distance was estimated by finding the threshold value at which the expected number missed within is equal to the expected number detected beyond.

In REM density, D, is a function of the detection distance, detection angle, an estimate of average speed of movement, and the camera trapping time period. Therefore, for known detection distances z_j $(j = 1, 2, ..., s)$ of animals, we can estimate an effective

detection distance, $\hat{\gamma}$, which is then used in the estimation of animal density,

$$D = \frac{\lambda}{t} \frac{\pi}{(2+\theta)\hat{\gamma}\bar{v}}. \tag{5.2.1}$$

The detection angle θ and camera trap time period, t, are fixed. Note that in REM an independent estimate of the mean speed, $\bar{v}$, is required in the estimation process.

5.3 The data

This section describes the data set used in the analysis. The data were recorded in the years 2008 to 2010 at a tropical moist forest, Barro Colorado Island (BCI), Panama (see Rowcliffe et al., 2011). Data were collected from 19 species of terrestrial mammals, however, we concentrate on nine of these species as used in Rowcliffe et al. (2011). The data collected include: 1) detection distance, 2) detection angle, 3) encounters (trap records), and 4) speed of movement of animals.

5.3.1 Placement strategies of camera traps

There were 20 camera traps at randomly selected locations in ten different 1-ha forest plots. All cameras were mounted around 20 cm off the ground and angled to be parallel to the slope of the ground. The traps were aimed in the direction of less vegetation or slope to maximize the view. There was no directed placement of camera traps, baiting or luring to maximise trapping rates, instead traps were placed relative to the distribution of animals in the surveyed area (Rowcliffe et al., 2011).

As discussed in Chapter 2, Rowcliffe et al. (2008) REM requires the data collected to be in the form of independent contacts between animals (individual or group) and camera traps. Therefore, the cameras were set to be inactive at 2 minutes interval. This interval period allows an animal (group of animals) to leave the camera trap detection zone after a contact, and later the same or a different animal re-enters to give a second, independent contact. Rowcliffe et al. (2008) suggest that this interval is sufficient to provide independent contact between animals and camera traps. An animal can approach the camera trap from any direction and enter the trap zone. When motion

sensors are triggered the cameras made 10 low-resolution pictures, and without delay the camera was triggered again producing a short video clip of animals moving in front of the camera. These provide the detailed information required to estimate parameters such as detection distance and angle; and distance travelled by the animal and the duration, that is, the time spent to cover the distance travelled.

5.3.2 *Extracting the data from camera trap records*

The detection distances, z, and angles to detection, θ, come from the photographs taken by the camera traps. The detection distances and angles were recorded using the position of the animal when it first triggered the camera trap. The position of the animal was obtained by examining a subset of the photographs from a portable card reader used in the field before removing the camera. The position of the animal was noted relative to nearby landmarks such as trees and rocks from the camera trap. The detection distance and angle were then measured from the camera trap to that position using a measuring tape and compass. There are 1555 records of animal positions on first detection for the 19 species (Rowcliffe et al., 2011).

Unlike the WWAP set used in Chapters 2, 3, and 4, the animal speed data is not independently observed but comes from the same photo records taken by the camera traps. Before removing the camera traps in the field the speed of animals moving in front of the camera traps were measured. Using a measuring tape, the length of each animal's path through the environment was measured and recorded, and this distance was then divided by the time between the first and last image recordings of the animal. The standard mean of these speed measures was then used as the average speed of the species in REM fitted in Rowcliffe et al. (2011).

5.3.3 *Summary of real data at BCI, Panama*

The recording of the data set at BCI, Panama is different from the data recorded at Whipsnade Wild Animal Park. At WWAP, the number of encounters, y_{ij}, where $i = 1, 2, ..., c$ is the camera traps and $j = 1, 2, ..., n$ is the number of camera trap days, were collected for a fixed camera trap time period of $t = 1$ (day). However, for the

BCI data, the encounters, a_i $(i = 1, 2, ..., c)$, were collected at the camera level, i.e., the data were recorded over a period of t_i days on the ith camera trap. The total number of camera trap days is $\sum_{i=1}^{c} t_i = 7569.78$. Also, for each species, a sample of the detection distance r, and angle to detection θ were collected from the camera traps and an estimate of the detection distance and angle were computed. In Table 5.3.1 we give the first three recordings of the species with the largest data set at BCI (agouti).

Table 5.3.1: First three data points for agouti species

Camera ID (i)	Camera duration (t_i days)	Number of encounters (a_i) per camera
11	8.975694	3
12	9.025694	19
13	8.92083	16

In our analyses, we concentrate on 9 species investigated in Rowcliffe et al. (2011) for which at least 40 records of the position on first detection were available. Table 5.3.2 below shows the number of encounters and the sample mean speed ($\bar{v}$), and detection distance ($\bar{z}$) for nine species at BCI. An approximate standard error is given in parentheses, which is computed as given in Section 1.4. Note, to get an estimate of the angle to detection, θ for each species, we compute the mean angle to the detections, $\bar{\theta}$ to estimate density for each species. The formula is given below. Let

$$S_\theta = \sum_{i}^{s} \sin(\theta_i) \quad \text{and} \quad C_\theta = \sum_{i}^{s} \cos(\theta_i). \text{ Then } \bar{\theta} = \arctan(S_\theta, C_\theta).$$

For brevity, we will use the common names of the species in parentheses in the thesis.

Table 5.3.2: Summary of key traits of nine Panamanian forest mammal species with at least 40 records of position on first detection. The sample mean of speed ($\bar{v},\ \mathrm{ms}^{-1}$), detection distance ($\bar{z},\ m$) and mean direction (angle) ($\bar{\theta},$ radians) are given with their standard error in parentheses. For brevity, the abbreviated common names given in brackets are used hereafter.

Species	**Sample size**			**sample mean**		
	encounters	Speed	Distance	$\bar{v}$	$\bar{z}$	$\bar{\theta}$
Ocelot *Leopardus pardalis* (ocelot)	330	93	73	0.41 (0.04)	1.86 (0.22)	0.24 (0.02)
White-nosed coati *Nasua narica* (coati)	840	125	60	0.33 (0.03)	2.20 (0.28)	0.20 (0.01)
Central American spiny rat *Proechimys semispinosus* (rat)	916	132	114	0.23 (0.02)	1.30 (0.12)	0.24 (0.02)
Collared peccary *Tayassu tajacu* (peccary)	5066	265	128	0.31 (0.02)	3.09 (0.27)	0.27 (0.02)
Red brocket deer *Mazama Temama* (brocket)	854	181	86	0.28 (0.02)	2.92 (0.32)	0.24 (0.02)
Paca *Agouti paca* (paca)	1049	195	139	0.27 (0.02)	2.24 (0.19)	0.23 (0.02)
Central American agouti *Daysyprocta punctata* (agouti)	11220	953	753	0.26 (0.01)	2.15 (0.08)	0.23 (0.02)
Red-tailed squirrel *Sciurus granatensis* (squirrel)	604	66	56	0.26 (0.03)	1.82 (0.24)	0.24 (0.02)
Mouse unknown species (mouse)	104	43	43	0.17 (0.03)	1.29 (0.20)	0.20 (0.01)

5.4 Estimating effective detection distance from the BCI data

Most species are substantially more detectable in open areas (habitats) such as farmland than in closed habitats such as woodland. Therefore, it would be important to have some measure of detectability when designing surveys. As discussed in Section 5.1.1.1, in point transect sampling an effective radius γ, which is the radius of the circle around each point such that as many objects are detected beyond γ as remain undetected within γ, can be used (see Buckland, 1992). The effective radius can be computed from the radial distances, which are assumed to follow some probability detection function. Buckland (1992) state that the modelling process of the radial distances has two components: 1) a "key function", which is selected at the starting point, and is possibly based on visual inspection of the histogram of distances after truncation of visual outliers; and 2) a "series expansion", which is a flexible form used to adjust the key function. In some cases, however, the key function by itself will be adequate, with no terms in the series expansion. Buckland et al. (2001) recommended some key functions that are useful in modelling distance data. One such function is the halfnormal model since it has the three desired properties of:

- model robustness: this means that the model is a general, flexible function that can take a variety of plausible shapes for the detection function;
- shape criterion: the detection function should have a "shoulder" near the point (i.e., detection remains almost certain at small distances from the point);
- estimator efficiency: it is desirable that the model selected provides estimates that are relatively precise (that is, have small variance).

A halfnormal probability detection function of distances z_{ij} for $i = 1, 2, ..., c$ camera trap, and $j = 1, 2, ..., s$ observed distances with standard deviation σ of the corresponding zero-mean normal distribution is defined as

$$p(z_{ij} \mid \sigma) = \frac{\sqrt{2}}{\sigma\sqrt{\pi}} \exp\left(-\frac{{z_{ij}}^2}{2\sigma^2}\right) \qquad \text{for } \sigma > 0;\ z_{ij} > 0. \tag{5.4.1}$$

The mean detection distance $\mathbb{E}(z) = \sigma\sqrt{2}/\sqrt{\pi}$ and variance $\mathrm{Var}(z) = \sigma^2\,(1 - 2/\pi)$. The parameter σ is referred to as the scale parameter. Using equation (5.1.7) to define the

probability density function of the observed radial distances and equation (5.1.14) given in Section 5.2.1, we can estimate the effective detection distance (radius) for each species at BCI, Panama. We adopt the method of maximum likelihood estimation to estimate the effective detection distance.

So suppose we have s_i observed distances such that z_{ij} for $i = 1, 2, ..., c$ camera traps and $j = 1, 2, ..., s_i$, then likelihood function can be written as

$$L(\sigma \mid z_{ij}) = \prod_{i=1}^{c} \prod_{j=1}^{s_i} g(z_{ij} \mid \sigma), \tag{5.4.2}$$

where $g(z_{ij} \mid \sigma)$ is the probability density function of the observed distances, which is defined as

$$g(z_{ij} \mid \sigma) = \frac{z_{ij} \cdot p(z_{ij} \mid \sigma)}{\int_0^w z_{ij} p(z_{ij} \mid \sigma) dz_{ij}}, \tag{5.4.3}$$

where w is the radius of the point (distance of detection of the camera trap). Note that equation (5.4.3) also holds for infinite w. This equation (5.4.3) is analogous to equation (5.1.7). The effective detection distance can be estimated as

$$\hat{\gamma} = \sqrt{\frac{2}{\hat{h}(0)}}$$

where $\hat{h}(0)$ is the slope of $g(\cdot)$ evaluated at $z_{ij} = 0$ (see Buckland et al., 2001).

Before attempting to estimate the effective detection distance, we examine the data by looking at histograms (Figure 5.4.1). At small distances, detection increases because the detection zone area of the camera trap increases with distance from the trap. As shown in Figure 5.2.1 and as discussed in Rowcliffe et al. (2011) relatively few animals are detected near the camera trap where the area of view of the camera trap is small. And this may occur because small animals may pass under the field of view of the camera, and or larger animals may be unwilling or unable to pass close to the tree where the camera trap is mounted. To examine how probability of detection falls off with distance, Buckland et al. (2001) state that we must "correct" for the increase in area away from the point (trap) by assigning the jth detection distance on the ith camera

trap z_{ij} a weight $1/z_{ij}$. So scaling frequencies by $1/z_{ij}$ causes the first histogram bar (Figure 5.4.2, left panel), corresponding to the few short distances, to increase in size the most, making these distances appear more influential than they are. Figure 5.4.2 (left panel) gives the estimated detection function (halfnormal) on a histogram of scaled frequencies. According to Buckland et al. (2001) the detection function may sometimes appear to fit badly at small distances, see for example graph (a) in Figure 5.4.2 (left panel), and this arises because of the deceptive nature of point transect data. Note that relatively few distances are recorded close to the camera trap, where the detection zone area is small, so the fit of the model is not heavily influenced by distances close to zero.

Estimates of the effective detection distance, $\hat{\gamma}$ for each species at BCI are given in Table 5.4.1. We also give the estimated mean detection distance, $\bar{\hat{z}}$. The results show that effective detection distance increases strongly with the size of the animal (Table 5.4.1) with a difference of 50% between the highest and smallest effective detection distances. The corresponding (fitted) probability density function (PDF) of observed distances for the nine species are shown in Figure 5.4.2 (right panel). In the next Section (5.5) we use the concept of effective detection distance to estimate density in iREM.

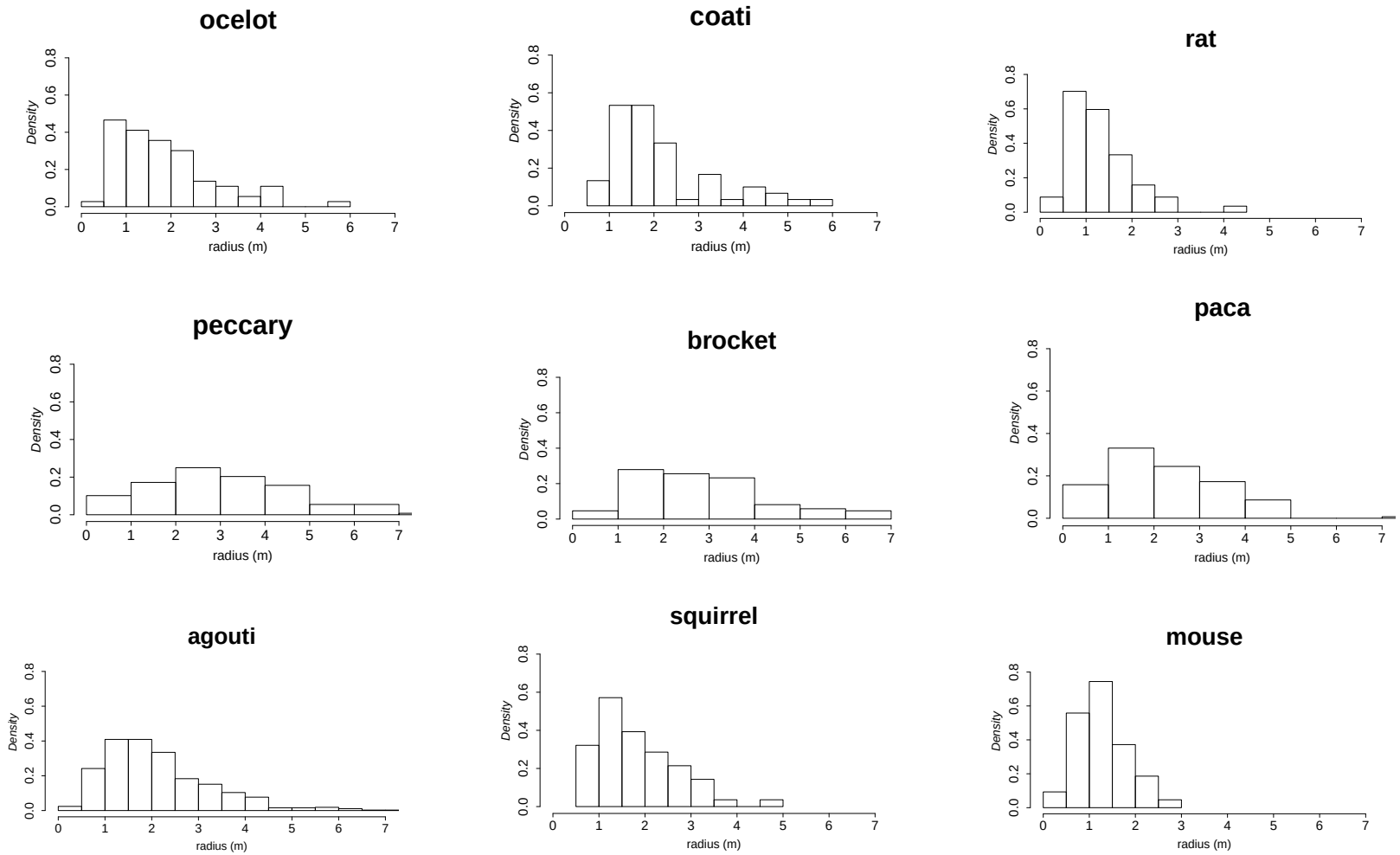
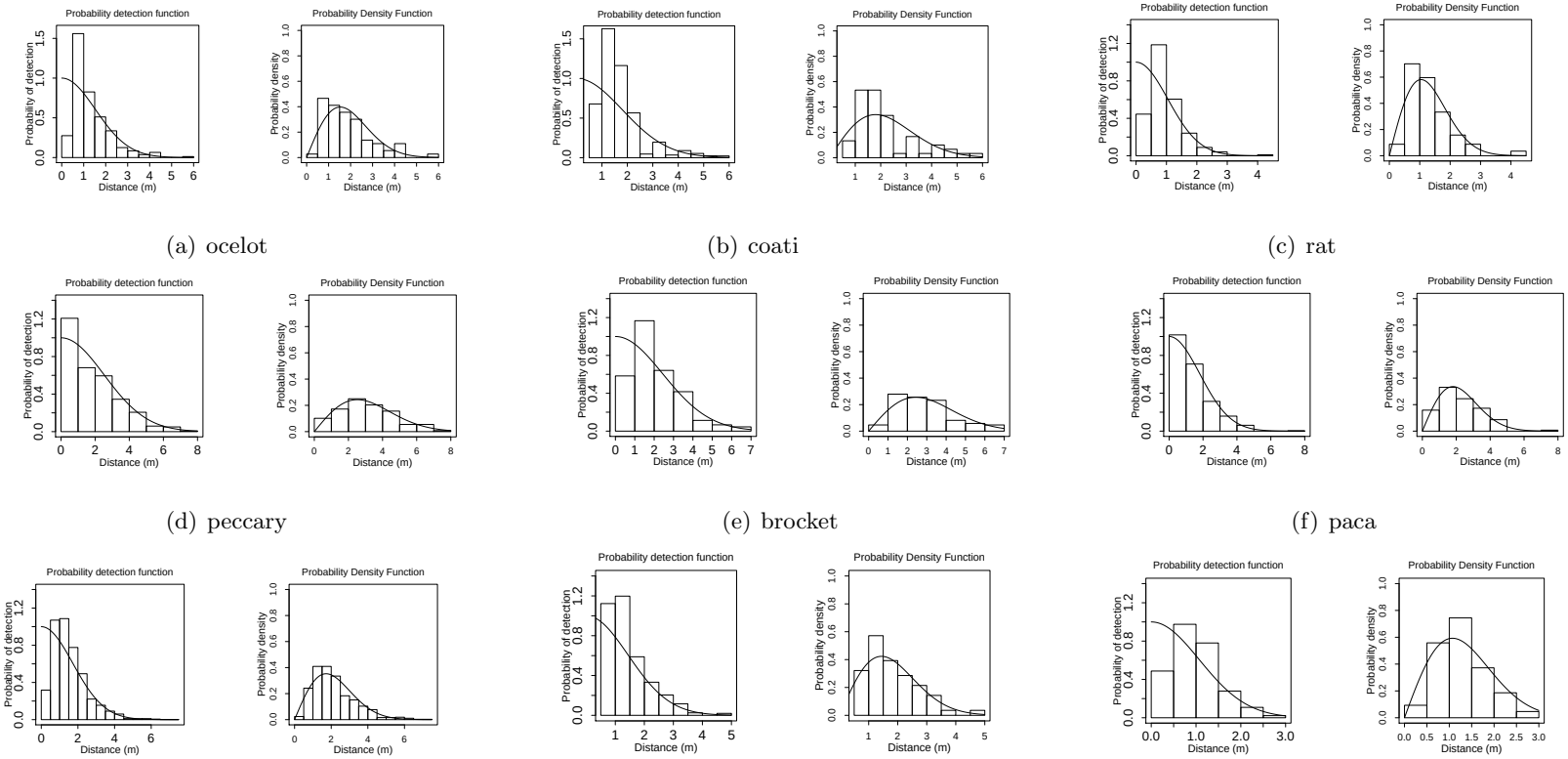


Figure 5.4.1: Histograms of detection distances of the nine species.

Table 5.4.1: Estimates of the effective detection distance, γ (in m) and mean detection distance, $\bar{\hat{z}}$ (in m) (standard error in parentheses).

Species	Parameter estimates	
	Mean detection distance ($\bar{\hat{z}}$)	Effective detection distance ($\hat{\gamma}$)
ocelot	1.22 (0.07)	1.56 (0.05)
coati	1.41 (0.09)	1.68 (0.05)
rat	0.83 (0.04)	1.29 (0.03)
peccary	1.96 (0.09)	1.98 (0.04)
brocket	1.84 (0.10)	1.92 (0.05)
paca	1.44 (0.06)	1.70 (0.04)
agouti	1.37 (0.02)	1.66 (0.02)
squirrel	1.13 (0.08)	1.51 (0.05)
mouse	0.79 (0.06)	1.26 (0.05)



(a) ocelot (b) coati (c) rat

(d) peccary (e) brocket (f) paca

(g) agouti (h) squirrel (i) mouse

Figure 5.4.2: Histograms of BCI data. The fitted halfnormal detection function is shown in the left panel, in which each histogram frequency has been scaled by dividing by the midpoint of the corresponding group interval. The fitted probability density function (pdf) is given in the right panel.

5.5 iREM with detection distance (iREM-dd)

In REM the independent parameters required to estimate density are: detection distance, r, detection angle, θ, expected animal speed, μ, camera trap time (in days), t, and average group size g if animals are found in pairs or family groups. The detection zone parameters are assumed to be fixed. In reality however, these vary and appropriate estimators are required. In this section, we develop an integrated likelihood, simultaneously modelling the encounter data, animal speed data and detection distance data to estimate animal density. It is worth reiterating that the speed data and the encounter data at BCI are collected from the same source, hence, there are not independent. So here we assume independence between the data sets so that their contributions to the likelihood could be multiplied.

In Chapter 2 Section 2.3 we give the nonparametric model used by Rowcliffe et al. (2008) to estimate density as

$$D = \frac{\sum_{i=1}^{c} a_i}{\sum_{i=1}^{c} t_i} \times \frac{\pi}{(2+\theta) r \bar{v}}, \tag{5.5.1}$$

where a_i is the encounters on the ith camera trap for $i = 1, 2, ..., c$ camera trap, t_i is the camera trapping time period of the ith camera trap, the fixed parameters θ and r are the detection angle and detection distance, respectively, and $\bar{v}$ is the mean speed. We showed that D in equation (5.5.1) could be derived by maximum likelihood estimation assuming the encounters follow a Poisson model. Rowcliffe et al. (2008) computed the standard error by bootstrapping the camera locations with replacement and taking the variance of a large number of resampled density estimates.

In this Chapter the information required for density estimation in REM is collected at the camera level, where the camera trap time, t_i $(i = 1, 2, ..., c)$, for the ith camera is no longer fixed. Therefore, assuming the rate of encounter, which is dependent on a variable time period t_i, is from a Poisson model, we show that the above formula for the density is a maximum likelihood estimator.

Suppose a_i is a Poisson random variable with encounter rate λt_i, that is, $a_i \sim \text{Poisson}(\lambda t_i)$. Then the probability mass function is

$$h_{pois}(a_i \mid \lambda,\ t_i) = \frac{e^{-\lambda t_i}}{a_i!}(\lambda t_i)^{a_i}, \tag{5.5.2}$$

where

$$\lambda = \left(\frac{2+\theta}{\pi}\right)\gamma\mu D, \tag{5.5.3}$$

and μ is expected speed; θ is the detection angle, which is held fixed, and the effective detection distance is γ. We assume the number of encounters, a_i are independent for all i. That is, all animals are captured independently of each other at each camera trap. Note that in an ideal gas model, from which is REM derived, a key assumption is that movement of individuals is random and independent of each other, and equally likely in all directions; see Section 2.1. Thus, from equation (5.5.2) the likelihood function is

$$L(a_i; \lambda,\ t_i) = \prod_{i=1}^{c} \frac{e^{-\lambda t_i}(\lambda t_i)^{a_i}}{a_i!}.$$

Taking the logarithm, we have

$$\ell(a_i; \lambda,\ t_i) = \text{constant} - \sum_{i=1}^{c} \lambda t_i + \sum_{i=1}^{c} a_i \log(\lambda t_i).$$

Differentiating with respect to λ and setting equal to zero, we have

$$\hat{\lambda} = \frac{\sum_{i=1}^{c} a_i}{\sum_{i=1}^{c} t_i}.$$

An estimate of the density, D, from equation (5.5.3) is therefore $\hat{D} = \hat{\lambda} \times 1/C$, where $C = \{(2+\theta)/\pi\}\,\gamma\mu$ is a constant. An estimate of the standard error of the density can be computed using the Fisher Information Matrix. Since density, $\hat{D} = \hat{\lambda} \times 1/C$, the standard error of the density estimator, $\hat{D}$ is the standard error of the encounter rate multiplied by some constant, that is, $Se(\hat{\lambda}) \times 1/C$. The Fisher Information is

$$I(\hat{\lambda}) = -\mathbb{E}\left[\frac{\partial^2}{\partial \hat{\lambda}^2}\ \log f(a_i; \hat{\lambda})\right] = \mathbb{E}\left[\frac{\sum_{i=1}^{c} a_i}{\hat{\lambda}^2}\right] = \frac{\sum_{i=1}^{c} t_i}{\hat{\lambda}}.$$

The variance is computed by taking the inverse of the Fisher Information, that is,

$$I^{-1}(\hat{\lambda}) = \frac{\hat{\lambda}}{\sum_{i=1}^{c} t_i} = \frac{\sum_{i=1}^{c} a_i}{\left(\sum_{i=1}^{c} t_i\right)^2}.$$

We can therefore find an estimate of the standard error of the density estimator, which is an alternative method to the bootstrapping method used by Rowcliffe et al. (2008) in REM. Since $\hat{D} = \hat{\lambda} \times 1/C$, then the variance of $\hat{D}$ is $I^{-1}(\hat{\lambda}) \times (1/C)^2$. The standard error of the density is therefore, $\sqrt{I^{-1} \times (1/C)^2}$.

5.5.1 The Model

In this section we give the integrated REM for estimating density accounting for the sampling variability in speed and detection distance. We propose an iREM with detection distance model (iREM-dd) to estimate the density of unmarked species. Suppose the encounter data, a_i, has probability mass function $h(a_i \mid \lambda,\ t_i,\ \tau)$ where $i = 1, 2, ..., c$ is the ith camera trap, t_i is the camera trap time period for the ith camera trap, τ represents any additional parameters in the model, and

$$\lambda = \left(\frac{2+\theta}{\pi}\right) \gamma \mu D. \tag{5.5.4}$$

And suppose for m speed observations such that $x = \{x_1, x_2, ..., x_m\}$, has probability density function $f(x_l \mid \mu,\ \nu)$ where $l = 1, 2, ..., m$; μ is the expected speed and ν represents any additional parameters in the model. Also, suppose the detection distance data, z_{ij}, has probability density function $g(z_{ij} \mid \sigma)$ where $j = 1, 2, ...s_i$ is the jth detection distance on the ith ($i = 1, 2, ..., c$) camera trap, σ is the scale parameter from a zero- mean normal distribution. As discussed in Section 3.2 in Chapter 3, certain assumptions of the combined likelihood are made. These include: 1) encounters between animals and camera are independent, 2) animals move randomly and they encounter camera traps independently of each other, 3) speeds are independent and identically distributed, 4) if animals are found to be moving in pairs or family group sizes, these are assumed to be small; given that the probability of detection depends on the size of the group and the distance, we assume groups are small and are detected within the sector-shaped detection zone of the camera trap. Therefore, the log-likelihood function

can be constructed as

$$\ell = \sum_{i=1}^{c} \log\, h(a_i \mid \lambda,\ t_i\ \tau) + \sum_{l=1}^{m} \log\, f(x_l \mid \mu,\ \nu) + \sum_{i=1}^{c}\sum_{j=1}^{s_i} \log\, g(z_{ij} \mid \sigma). \tag{5.5.5}$$

We consider a gamma model, a lognormal model, or a Weibull model for the speed data as proposed and used in Chapter 3, and a Poisson REM described above for the encounter data. We could also use a negative binomial model as shown in Section 2.1.1 in Chapter 2. Suppose $a_i \mid \lambda \sim \text{Poi}(\lambda t_i)$ has probability mass function

$$h_{pois}(a_i \mid \lambda,\ t_i) = \frac{e^{-\lambda t_i}}{a_i!}(\lambda t_i)^{a_i}, \tag{5.5.6}$$

where $\lambda \sim \text{Ga}(\kappa, \beta)$ with probability density function

$$g(\lambda) = \frac{\beta^{\kappa}}{\Gamma(\kappa)}\lambda^{\alpha-1}\text{exp}(-\lambda\beta). \tag{5.5.7}$$

To obtain the density function of the encounters a_i we need to integrate out λ.

$$h(a_i) = \int_{\lambda=0}^{\infty} h(a_i \mid \lambda,\ t_i)g(\lambda)d\lambda = \int_0^{\infty} \frac{e^{-\lambda t_i}}{a_i!}(\lambda t_i)^{a_i}\frac{\beta^{\kappa}}{\Gamma(\kappa)}\lambda^{\kappa-1}\text{exp}(-\lambda\beta)d\lambda$$

$$= \frac{{t_i}^{a_i}\beta^{\kappa}\Gamma(\kappa+a_i)}{a_i\Gamma(\kappa)(t_i+\beta)^{\kappa+a_i}}\int_0^{\infty}\frac{(t_i+\beta)^{\kappa+a_i}\lambda^{a_i+\kappa-1}\text{exp}\{-\lambda(t_i+\beta)\}}{\Gamma(\kappa+a_i)}$$

$$= \frac{{t_i}^{a_i}\beta^{\kappa}\Gamma(\kappa+a_i)}{a_i\Gamma(\kappa)(t_i+\beta)^{\kappa+a_i}} = \frac{\Gamma(a_i+\kappa)}{\Gamma(a_i+1)\Gamma(\kappa)}\left(\frac{t_i}{t_i+\beta}\right)^{a_i}\left(\frac{\beta}{t_i+\beta}\right)^{\kappa},$$

where the probability of encounter, $p = \beta/(t_i+\beta)$ and $1-p = t_i/(t_i+\beta)$. Therefore, $a_i \sim NB\,(\kappa,\ \beta/(t_i+\beta))$ with mean $\kappa t_i/\beta$. Given that the encounter rate is λt_i, where $\lambda = \{(2+\theta)/\pi\}\,\gamma\mu D$, and the mean encounters from the Poisson-gamma mixture is $\kappa t_i/\beta$, then in this case $\lambda = \kappa/\beta$ and $\beta = \kappa/\lambda$. Therefore, the probability density function of the encounters, a_i has NB-2 form

$$h(a_i; \lambda,\ t_i,\ \kappa) = \frac{\Gamma(\kappa+a_i)}{\Gamma(a_i+1)\Gamma(\kappa)}\left(\frac{\kappa}{\lambda t_i+\kappa}\right)^{\kappa}\left(\frac{\lambda t_i}{\lambda t_i+\kappa}\right)^{a_i}.$$

where κ is the dispersion parameter. The negative binomial distribution has NB-2 form as given in Hilbe (2011).

5.6 Performance of the iREM-dd model

In this section we test the performance of the iREM-dd model. We simulate N_A individuals uniformly and select individuals that are detected based on their distance from the centre point of a circle. We assume a halfnormal probability detection function for the distances. For illustration, we fit a Poisson REM to the encounter data, and a lognormal model is fitted to the speed data simulated from the relevant fitted model. We assess the output for the true values and sample sizes given in Table 5.6.1 based upon 100 simulations. We set $\theta = 0.28$ (radians). The chosen parameter values are realistic for real species in our motivating data set. The average parameter estimates (standard error in parentheses), Standard deviation (Sd), and Root Mean Square Error (RMSE) are computed. We test the Poisson iREM-dd under various parameter settings and population size, N_A. For each simulation run, the number of individuals detected varies. Increasing the number of individuals in the population (N_A), is likely to increase the number of individuals detected. The simulation results show that the Poisson iREM-dd model is working as it is expected to under all scenarios. As expected, increasing the number of individuals in the population available for capture, the number of speed observations, and the camera trapping days improves estimates of the density and its precision (see Table 5.6.1 below). Note that higher variability in the distance data means that the value of the effective detection distance would be larger.

Table 5.6.1: Simulation results from a Poisson iRM-dd. Detection distance is assumed to follow a halfnormal distribution and animal speed is assumed to follow a lognormal model. Density, D is estimated in km^2, expected speed, μ is in ms^{-1}, and expected effective radius, γ is in m. The number of animals in the population available for detection is denoted by N_A; n is the number of camera trapping days, and m is the number of speed observations. The standard errors (in parentheses), Standard deviation (Sd) and Root Mean Square Error (RMSE) are also given.

	$\hat{D}$	$\hat{\mu}$	$\hat{\gamma}$	$\hat{D}$	$\hat{\mu}$	$\hat{\gamma}$
	Increasing D			**Larger N_A**		
	$N_A = 100$; $n = 500$; $m = 40$			$N_A = 1000$; $n = 500$; $m = 40$		
	$D = 0.96$; $\mu = 0.156$; $\gamma = 1.22$			$D = 0.96$; $\mu = 0.156$; $\gamma = 1.22$		
Estimates	0.99 (0.16)	0.158 (0.01)	1.22 (0.16)	0.98 (0.09)	0.155 (0.01)	1.20 (0.06)
Sd	0.16	0.01	0.16	0.08	0.01	0.05
RMSE	0.16	0.01	0.16	0.08	0.01	0.05
	$D = 15.96$; $\mu = 0.156$; $\gamma = 1.22$			$D = 15.96$; $\mu = 0.156$; $\gamma = 1.22$		
Estimates	16.90 (2.82)	0.154 (0.01)	1.19 (0.19)	16.32 (1.50)	0.155 (0.03)	1.20 (0.06)
Sd	3.00	0.01	0.16	0.29	0.003	0.02
RMSE	3.14	0.01	0.16	1.36	0.01	0.05
	$D = 30.96$; $\mu = 0.156$; $\gamma = 1.22$			$D = 30.96$; $\mu = 0.156$; $\gamma = 1.22$		
Estimates	31.94 (5.36)	0.155 (0.01)	1.80 (0.20)	31.24 (1.81)	0.156 (0.01)	1.20 (0.06)
Sd	4.74	0.01	0.15	3.19	0.01	0.05
RMSE	4.84	0.01	0.15	3.24	0.01	0.05
	Increasing $\text{Var}(z_{ij})$			**Larger survey effort**		
	$N_A = 100$; $n = 500$; $m = 40$			$N_A = 1000$; $n = 2000$; $m = 200$		
	$D = 0.96$; $\mu = 0.156$; $\text{Var}(z_{ij}) = 0.52$; $\gamma = 1.70$			$D = 0.96$; $\mu = 0.156$; $\text{Var}(z_{ij}) = 0.52$; $\gamma = 1.70$		
Estimates	0.99 (0.19)	0.155 (0.01)	1.70 (0.27)	0.97 (0.06)	0.155 (0.002)	1.69 (0.10)
Sd	0.19	0.01	0.29	0.06	0.002	0.10
RMSE	0.19	0.01	0.30	0.06	0.002	0.10
	$N_A = 100$; $n = 500$; $m = 40$			$N_A = 1000$; $n = 2000$; $m = 200$		
	$D = 0.96$; $\mu = 0.156$; $\text{Var}(z_{ij}) = 3.06$; $\gamma = 4.10$			$D = 0.96$; $\mu = 0.156$; $\text{Var}(z_{ij}) = 3.06$; $\gamma = 4.10$		
Estimates	0.98 (0.16)	0.157 (0.01)	4.07 (0.34)	0.97 (0.04)	0.155 (0.002)	4.07 (0.10)
Sd	0.18	0.01	0.55	0.04	0.002	0.17
RMSE	0.18	0.01	0.55	0.04	0.002	0.17
	$N_A = 100$; $n = 500$; $m = 40$			$N_A = 1000$; $n = 2000$; $m = 200$		
	$D = 0.96$; $\mu = 0.156$; $\text{Var}(z_{ij}) = 11.40$; $\gamma = 7.92$			$D = 0.96$; $\mu = 0.156$; $\text{Var}(z_{ij}) = 11.40$; $\gamma = 7.92$		
Estimates	1.01 (0.15)	0.154 (0.01)	7.68 (0.41)	0.98 (0.04)	0.156 (0.002)	7.72 (0.13)
Sd	0.14	0.01	0.90	0.03	0.003	0.25
RMSE	0.15	0.01	0.93	0.04	0.002	0.32

5.7 Application of iREM-dd to real data at BCI

An application of iREM-dd to the BCI data is given in this section. With the assumption that the encounter data, speed data and detection distance data are independent we combine the data sets and maximize the likelihood to estimate the density. We fit an iREM-dd to the data of nine terrestrial Panamanian species at BCI rainforest. We assume a halfnormal probability detection function for distance data; the speed data is assumed to follow a gamma model, a lognormal model or a Weibull model that was used in Chapters 3, and 4, and the encounter data is assumed to follow a Poisson REM described in Section 5.5 above.

Table 5.7.1 gives the parameter estimates (standard errors in parentheses) and the ΔAIC values for each of the nine species. The results show minimal differences between the estimates from the three speed data models. However, there is more support for the lognormal as shown by the ΔAIC values. For large species such as ocelots, paca, brocket and peccary, the estimated effective detection is larger compared with smaller species such as mouse and rat for example. Note that larger species tend to be more detectable at distances further away from the trap, than smaller species. In the next Section we consider modelling the angle to detection, to an estimate effective detection to estimate the density.

Table 5.7.1: Estimates of density (D, km^2), expected animal speed (μ, ms^{-1}) and effective detection distance (γ, m) where the observed distances are assumed to follow a halfnormal distribution (standard error in parentheses). The ΔAIC values are also given.

Species	**Speed data models**											
	gamma				lognormal				Weibull			
	$\hat{D}$	$\hat{\mu}$	$\hat{\gamma}$	ΔAIC	$\hat{D}$	$\hat{\mu}$	$\hat{\gamma}$	ΔAIC	$\hat{D}$	$\hat{\mu}$	$\hat{\gamma}$	ΔAIC
ocelot	0.96 (0.09)	0.41 (0.02)	2.15 (0.16)	0	0.96 (0.10)	0.41 (0.03)	2.15 (0.16)	6.97	0.96 (0.09)	0.41 (0.02)	2.15 (0.16)	0.08
coati	2.36 (0.24)	0.29 (0.02)	2.50 (0.22)	4.21	2.60 (0.30)	0.33 (0.03)	2.49 (0.21)	0	2.49 (0.28)	0.33 (0.03)	2.49 (0.21)	9.30
rat	6.20 (0.54)	0.23 (0.02)	1.30 (0.05)	12.65	6.73 (0.67)	0.23 (0.02)	1.47 (0.07)	0	6.73 (0.67)	0.23 (0.02)	1.47 (0.07)	25.63
peccary	19.57 (1.67)	0.20 (0.02)	2.97 (0.16)	0	19.57 (2.01)	0.20 (0.02)	2.97 (0.16)	4.89	19.57 (1.51)	0.20 (0.01)	2.97 (0.16)	5.87
brocket	2.80 (0.12)	0.23 (0.01)	3.26 (0.27)	15.38	2.80 (0.29)	0.23 (0.02)	3.26 (0.27)	0	2.80 (0.12)	0.23 (0.02)	3.26 (0.27)	25.63
paca	3.83 (0.26)	0.27 (0.01)	2.55 (0.15)	24.12	3.83 (0.28)	0.27 (0.01)	2.55 (0.15)	0	3.83 (0.30)	0.27 (0.02)	2.55 (0.15)	60.41
agouti	43.05 (1.39)	0.27 (0.01)	2.07 (0.04)	19.16	43.38 (1.72)	0.24 (0.01)	2.56 (0.06)	0	43.38 (1.41)	0.24 (0.01)	2.56 (0.07)	12.31
squirrel	2.72 (0.20)	0.28 (0.01)	1.67 (0.10)	75.59	2.81 (0.23)	0.26 (0.01)	2.01 (0.16)	0	2.40 (0.20)	0.27 (0.01)	2.01 (0.16)	122.78
mouse	1.05 (0.17)	0.17 (0.02)	1.40 (0.11)	5.29	1.05 (0.19)	0.17 (0.02)	1.40 (0.11)	0	1.05 (0.17)	0.17 (0.03)	1.40 (0.11)	12.22

5.8 Modelling angle to detection data

Another aspect of the detection zone in camera trapping is the detection angle, which varies in practice. In REM, this variation is likely to have an effect on the encounter rate, which is dependent on the angle of detection. Rather than using a fixed value of the detection angle as REM does, we consider fitting a circular model to the angle to detection data and estimating an effective detection angle. The angle to detection data is circular data (angular), which is in the range 0-360 degrees or 0-2π radians. Circular data are modelled by circular distributions such as a *von Mises* distribution, which models the entry directions of animals in camera traps detection zones.

5.8.1 The model

In this section we propose a model for estimating the effective detection angle. A *von Mises* model is an appropriate model for fitting angular data and an effective detection angle can be obtained by maximizing the likelihood. For a vector of detection angles $\boldsymbol{\theta} = \{\theta_1, \theta_2, ..., \theta_n\}$ with mean υ, we use a truncated von Mises distribution for $\boldsymbol{\theta}$ $(0 \leq \boldsymbol{\theta} \leq \frac{\pi}{2})$. The probability density function is defined as

$$\begin{aligned} p_{\upsilon M}(\boldsymbol{\theta};\ \upsilon,\ \eta,\ 0,\ \pi/2) &= \frac{\exp\{\eta\ cos(\boldsymbol{\theta} - \upsilon)\}}{(2\pi)I_0(\eta)\int_0^{\pi/2}\frac{\exp\{\eta\ cos(\boldsymbol{\theta}-\upsilon)\}}{(2\pi)I_0(\eta)}d\boldsymbol{\theta}} \\ &= \frac{\exp\{\eta\ cos(\boldsymbol{\theta} - \upsilon)\}}{\int_0^{\pi/2}\exp\{\eta\ cos(\boldsymbol{\theta} - \upsilon)\}d\boldsymbol{\theta}} \end{aligned} \tag{5.8.1}$$

where υ is the mean detection angle which is equivalent to θ in equation (5.5.3); η is the concentration parameter; and $I_0(\eta)$ is the modified Bessel function of order zero and the first kind, and given by:

$$I_0(\eta) = \frac{1}{2\pi}\int_{\theta=0}^{2\pi}\exp\{\eta\ cos(\boldsymbol{\theta})\}\ d\boldsymbol{\theta} \tag{5.8.2}$$

(Pewsey et al., 2013). The modified Bessel function I_P (of the first kind and order p) is defined by

$$I_p(\eta) = \sum_{r=0}^{\infty}\frac{1}{(r+p)!r!}\Big(\frac{\eta}{2}\Big)^{2r+p},\quad p = 0, 1, 2, ...$$

In particular

$$I_0(\eta) = \sum_{r=0}^{\infty} \frac{1}{r!}\left(\frac{\eta}{2}\right)^{2r}.$$

(see Zhao, 2009, pages 233-234).

We fit the truncated *von Mises* model described in equation (5.8.1) to the angular data from BCI, Panama. Estimates of the mean angle to detection and the concentration parameter are given in Table 5.8.1. Unlike the results obtained for the detection distance which showed more variation between species (section 5.8) the results in Table 5.8.1 show that there is not much variation in mean direction between species.

Table 5.8.1: Estimated mean angle to detection, $\hat{\upsilon}$ and concentration parameter $\hat{\eta}$

Species	**Parameter estimates**	
	$\hat{\upsilon}$	$\hat{\eta}$
ocelot	0.24 (0.02)	26.13 (4.53)
coati	0.20 (0.02)	35.98 (6.86)
rat	0.24 (0.02)	23.86 (3.29)
peccary	0.27 (0.02)	23.38 (3.11)
brocket	0.24 (0.02)	30.06 (4.89)
paca	0.23 (0.02)	27.33 (3.44)
agouti	0.23 (0.01)	29.79 (1.62)
squirrel	0.24 (0.03)	26.95 (5.37)
mouse	0.20 (0.03)	35.75 (8.12)

5.9 iREM with detection zone dimensions (iREM-D)

The iREM-dd model in Section 5.5 can be expanded to incorporate the angle to detection data, forming a single integrated likelihood to estimate abundance of unmarked species. Suppose the encounter data, a_i has probability mass function $h(a_i \mid \lambda,\ t_i)$ where $i = 1, 2, ..., c$ is the ith camera trap, t_i the camera trap time period for the ith camera trap

and where

$$\lambda = \left(\frac{2+\upsilon}{\pi}\right)\gamma\mu D. \tag{5.9.1}$$

We assume the encounters between animals and camera traps are independent, and animals move randomly and encounter camera traps independently of each other. The mean angle to detection is denoted by υ; γ is the effective detection radius; μ is the expected animal speed; and D is the density. Also, suppose m animal speed observations such that $x = \{x_1, ..., x_m)$ have probability density function $f(x_l \mid \mu,\ \nu)$ where $l = 1, 2, ..., m$, μ is the expected animal speed and ν represents any additional parameters in the model. And, suppose the detection distances, z_{ij}, have probability density function $g(z_{ij} \mid \sigma)$ where $j = 1, 2, ..., s_i$ is the jth observed detection distance, on the ith camera trap, and σ represents any additional parameters; and the angles to detections, θ_{ij}, have probability density function $p(\theta_{ij} \mid \upsilon,\ \eta)$ where $j = 1, 2, ..., s_i$ is the jth observed angle to detection on the ith camera trap, υ is the expected angle to detection, and η represents any additional parameters in the model. Under the assumption that the data sets are independent, the joint log-likelihood can be constructed as

$$\begin{aligned}\ell = &\sum_{i=1}^{c} \log\ h(a_i \mid \lambda,\ t_i,\ \tau) + \sum_{l=1}^{m} \log\ f(x_l \mid \mu,\ \nu) + \\ &\sum_{i=1}^{c}\sum_{j=1}^{s_i} \log\ g(z_{ij} \mid \sigma) + \sum_{i=1}^{c}\sum_{j=1}^{s_i} \log\ p(\theta_{ij} \mid \upsilon,\ \eta).\end{aligned} \tag{5.9.2}$$

5.10 Simulation Study

In this section we investigate the performance of iREM-D, under the assumption that the data sets are independent, for the two following cases.

(i) Firstly, we investigate the performance of the Poisson iREM-dd and the importance of accounting for the sampling variability in animal speed, detection distance and angle to detection. To do this we fit a Poisson iREM-D, a Poisson iREM-dd, a Poisson iREM and a Poisson REM to data simulated from a Poisson REM given in Section 5.5 above (Section 5.10.1).

(ii) Secondly, we investigate the importance of accounting for the variation in the encounter data. To do this we fit a Poisson iREM-D, a Poisson iREM-dd, a NB

iREM-D, and a NB iREM-dd to encounter data simulated from a NB REM given in Section 5.5 above (Section 5.10.2).

For the simulations we use parameter values and sample sizes that are plausible ecologically for real species at BCI, Panama. For illustration we assume a lognormal for the speed data. As done in Section 5.6 we simulate N_A individuals uniformly and select individuals that are detected based on their distance and angle from the centre point of a circle. We assume a halfnormal probability detection function for the distances and a von Mises model for the angular data. To investigate the importance of accounting for the sampling variability in distance and angle to detection we simulate data with low and high variance. For the angle to detection $1/\eta$ is analogous to the variance from a normal distribution, therefore, the smaller the value of η the larger the variance. The average parameter estimates (average standard error in parentheses), Standard deviation (Sd) and Root Mean Square Error (RMSE) are computed for 100 simulation runs.

5.10.1 Investigating the importance of accounting for the sampling variability in speed, detection distance and angle to detection angle data

In this Section we compare estimates from iREM, iREM-dd, iREM-D methods and the REM method for simulated data. The purpose is to investigate the importance of accounting for the variation in data for animal speed, detection distance and angle to detection. We set the expected speed, $\mu = 0.156$ ms^{-1} with variance, $\text{Var}(x_l) = 0.11$. In all cases we fit a Poisson REM to encounters simulated from a Poisson REM. A lognormal model is fitted to the speed data; a halfnormal model is fitted to the distance data, and von Mises model is fitted to the angular data. For illustration we simulate camera trap time data from a gamma model, which is used to estimate to encounter rate, λt_i. The methods are tested using 100 simulation runs and the parameter estimates, their associated standard errors, the standard deviation (Sd) and the Root Mean Square Error are computed.

Table 5.10.1 gives the average estimates of the density (standard error in parentheses) for increasing density, D and increasing variability, $\text{Var}(z_{ij})$ in the distance data. We set a low value of the concentration parameter, $\eta = 10$, so that the variability in the angular

data would be minimal. We set the number of animals in the population available for detection to $N_A = 100$; the number of camera trapping days is set to $n = 1500$; and the number of speed observations is set to $m = 100$. The results (Table 5.10.1) show that the Poisson REM performs poorly for small values of the density but gave estimates of the density that are in good agreement with the true values of the density when D increases. However, the Poisson REM and Poisson iREM achieved smaller estimates of the standard error compared with a Poisson iREM-dd and a Poisson iREM-D. It is worth reiterating that REM does not account for the sampling variability in the data entering the model, while iREM accounts only for the variability in the speed data. When the sampling variability in the distance data increases the bias in the density from a Poisson REM and a Poisson iREM increases. The Poisson iREM-dd and a Poisson iREM-D gave similar estimates but the standard error from the Poisson iREM-dd is smaller compared with estimates from the Poisson iREM-D. Also, the evidence shows that the Poisson iREM-D gave larger RMSE compared with the other models. The reason for this is because the Poisson iREM-D accounts for the sampling variability in the all parameters, and hence, gave larger estimates of the variance (note, the RMSE $= \sqrt{\text{Var}(\hat{\theta}) + \text{Bias}(\hat{\theta}, \theta)^2}$ as given in Section 1.4 in Chapter 1). We also give estimates of the density for increasing variability in the angular data, and increasing survey effort (Table 5.10.2). For all methods estimates of the density are in good agreement with the true density value. But as expected, increasing variability in the angular data has an effect on the estimated standard error from the simpler models (Poisson REM and Poisson iREM). However, there is minimal difference in estimated density between the Poisson iREMs, and a Poisson REM gave a larger bias compared with the Poisson iREMs. Increasing survey effort improves the estimates and precision but again.

Based on these results we can conclude that accounting for the variability in the data entering REM is highly relevant, since ignoring this variation can introduce estimation bias and can result in an underestimation of the standard error of the density. iREM is a simpler model to fit compared with iREM-dd and iREM-D and can provide an adequate description of the data, but as shown in the simulations, iREM-dd and iREM-D are more appropriate models for data with high variability in detection distance and angle as they provide more precise estimates and correctly estimates the standard error.

Table 5.10.1: Simulation results from a Poisson iRM-D. Detection distance is assumed to follow a halfnormal distribution and animal speed is assumed to follow a lognormal model. Density, D is estimated in km^2, expected speed, μ is in ms^{-1}. The effective radius is denoted by γ, and the effective detection angle is denoted by η. The number of animals in the population available for detection is $N_A = 100$; $n = 1500$ is the number of camera trapping days, and $m = 100$ is the number of speed observations. The standard errors (in parentheses), Standard deviation (Sd) and Root Mean Square Error (RMSE) are also given.

	REM	iREM	iREM-dd	iREM-D	REM	iREM	iREM-dd	iREM-D
	Increasing D				**Increasing variability in distance data (Increasing** $\mathrm{Var}(z_{ij})$**)**			
	$D = 0.96$; $\mu = 0.156$; $\gamma = 1.22$; $\eta = 10$; $\mathrm{Var}(z_{ij})$ =0.27				$D = 15.96$; $\mu = 0.156$; $\gamma = 2.63$; $\eta = 10$; $\mathrm{Var}(z_{ij})$ =1.26			
$\hat{D}$	2.72 (0.01)	1.09 (0.23)	1.09 (0.28)	1.02 (0.31)	18.22 (0.02)	17.35 (3.79)	17.00 (4.22)	16.83 (4.36)
Sd	0.01	0.32	0.31	0.34	4.49	3.49	4.19	4.41
RMSE	1.76	0.27	0.31	0.35	2.26	4.04	4.34	4.45
	$D = 15.96$; $\mu = 0.156$; $\gamma = 1.22$; $\eta = 10$; $\mathrm{Var}(z_{ij})$ =0.27				$D = 15.96$; $\mu = 0.156$; $\gamma = 4.47$; $\eta = 10$; $\mathrm{Var}(z_{ij})$ =3.63			
$\hat{D}$	17.80 (0.03)	16.70 (3.65)	16.59 (4.41)	16.47 (4.61)	18.65 (0.01)	17.55 (3.94)	17.16 (4.34)	16.92 (4.46)
Sd	5.22	4.67	4.61	4.62	4.63	3.28	4.41	4.59
RMSE	1.84	3.73	4.53	4.69	2.69	4.25	4.50	4.57
	$D = 30.96$; $\mu = 0.156$; $\gamma = 1.22$; $\eta = 10$; $\mathrm{Var}(z_{ij})$ =0.27				$D = 15.96$; $\mu = 0.156$; $\gamma = 5.88$; $\eta = 10$; $\mathrm{Var}(z_{ij})$ =6.29			
$\hat{D}$	34.92 (0.04)	33.28 (7.31)	33.02 (8.75)	32.79 (8.95)	18.75 (0.01)	17.68 (3.98)	17.18 (4.31)	16.81 (4.41)
Sd	10.56	8.06	8.31	8.37	5.14	3.24	4.18	4.28
RMSE	3.96	7.67	9.13	9.14	2.79	4.33	4.47	4.49

Table 5.10.2: Simulation results from a Poisson iRM-D. Detection distance is assumed to follow a halfnormal distribution and animal speed is assumed to follow a lognormal model. Density, D is estimated in km^2. We set the expected speed, $\mu = 0.156$ (ms^{-1}). The effective radius is denoted by γ, and the effective detection angle is denoted by η. The number of animals in the population available for detection is denoted by N_A; n is the number of camera trapping days, and m is the number of speed observations. The standard errors (in parentheses), Standard deviation (Sd) and Root Mean Square Error (RMSE) are also given.

	REM	iREM	iREM-dd	iREM-D	REM	iREM	iREM-dd	iREM-D
	Increasing η				**Increasing survey effort**			
	$D = 15.96$; $\gamma = 1.22$; $\eta = 15$; Var(z_{ij}) =0.27; $N_A = 100$; $n = 1500$; $m = 100$				$D = 15.96$; $\gamma = 4.47$; $\eta = 45$; Var(z_{ij}) =3.63 $N_A = 500$; $n = 2500$; $m = 200$			
$\hat{D}$	17.91 (0.03)	16.35 (3.65)	16.46 (4.43)	16.80 (4.50)	17.39 (0.01)	16.71 (2.62)	16.67 (2.75)	16.38 (2.75)
Sd	4.95	4.11	4.43	4.49	3.21	2.54	2.71	2.72
RMSE	1.95	3.67	4.54	4.57	1.43	2.72	2.84	2.79
	$D = 15.96$; $\gamma = 1.22$; $\eta = 30$; Var(z_{ij}) =0.27; $N_A = 100$; $n = 1500$; $m = 100$				$D = 15.96$; $\gamma = 4.47$; $\eta = 45$; Var(z_{ij}) =3.63; $N_A = 1000$; $n = 3500$; $m = 300$			
$\hat{D}$	17.40 (0.03)	16.77 (3.67)	16.78 (4.41)	16.82 (4.50)	16.24 (0.01)	16.60 (2.05)	16.22 (2.15)	16.01 (2.17)
Sd	4.98	4.16	4.11	4.60	2.58	2.01	2.13	2.14
RMSE	1.44	3.76	4.49	4.57	0.28	2.15	2.17	2.17
	$D = 15.96$; $\gamma = 1.22$; $\eta = 45$; Var(z_{ij}) =0.27; $N_A = 100$; $n = 1500$; $m = 100$				$D = 15.96$; $\gamma = 4.47$; $\eta = 45$; Var(z_{ij}) =3.63; $N_A = 1500$; $n = 4500$; $m = 400$			
$\hat{D}$	17.40 (0.03)	16.68 (3.67)	16.69 (4.40)	16.79 (4.42)	16.13 (0.01)	16.20 (1.76)	16.03 (1.89)	15.96 (1.90)
Sd	4.95	4.28	4.09	4.56	2.19	1.77	1.88	1.91
RMSE	1.44	3.74	4.46	4.50	0.17	1.78	1.89	1.90

5.10.2 *Investigating the importance of accounting for variation in encounter data*

In this section we investigate the importance of accounting for the sampling variability in the encounter data. We fit a Poisson iREM-dd and a Poisson iREM-D to encounters simulated from a NB REM described in Section 5.5 above. For comparison we a fit a NB iREM-dd and a NB iREM-D to the same encounter data. For illustration, we fit a lognormal model to the speed data, which is simulated from the relevant fitted model. As in Section 5.6 and Section 5.10.1 we simulate N_A individuals uniformly and select individuals that are detected based on their distance and angle to the centre point of a circle. We assume a halfnormal model for the distance data and a von Mises model for the angular data. We set $D = 15.96$ (km^2); $\mu = 0.156$ (ms^{-1}); and $\kappa = 0.76$. We test the performance of the models for low and high variability in the detection distance and angular data, and increasing sample sizes.

Table 5.10.3 compares density estimates from a Poisson iREM-dd, a Poisson iREM-D, a NB iREM-dd and a NB iREM-D. As expected, the Poisson iREMs gave a larger bias and/or smaller estimates of the standard error of the density since the variability in the encounter data is not accounted for. The evidence also suggest that not accounting for the sampling variability in the detection distance and angle to detection data would induce bias in the density estimator, and smaller estimates of the standard error.

Based on these results, we would recommend NB-iREMs, as the variability in the data is accounted for as shown by the estimated standard errors. A Poisson iREM-D would also provide reasonably accurate estimates of the density but care must be taken since it would give smaller estimates of the standard error of the density.

Table 5.10.3: Simulation results from fitting a Poisson iREM-dd, a Poisson iREM-D, a NB iREM-dd and a NB iREM-D to data simulated from a NB REM. Detection distance is assumed to follow a Weibull distribution, animal speed is assumed to follow a gamma model, and angular data is assumed to follow a von Mises model. We set $\kappa = 0.76$; $D = 15.96$ (km^2); $\mu = 0.156$ (ms^{-1}); $\upsilon = 0.28$ (radians). The standard errors (in parentheses), Standard deviation (Sd) and Root Mean Square Error (RMSE) are also given. Estimated densities from REM, iREM, iREM-dd and iREM-D are compared.

	Results for small sample sizes				**Results for large sample sizes**			
	iREM-dd		**iREM-D**		**iREM-dd**		**iREM-D**	
	Poisson	NB	Poisson	NB	Poisson	NB	Poisson	NB
	$\text{Var}(z_{ij}) = 0.27$; $\gamma = 1.22$; $\eta = 15$; $N_A = 100$; $m = 40$; $n = 1000$				$\text{Var}(z_{ij}) = 0.27$; $\gamma = 1.22$; $\eta = 15$; $N_A = 1000$; $m = 200$; $n = 2500$			
$\hat{D}$	18.39 (7.08)	16.46 (8.11)	17.84 (8.08)	16.84 (8.21)	16.73 (3.79)	16.32 (4.10)	16.57 (3.91)	16.81 (4.10)
Sd	7.57	8.16	8.21	8.14	4.19	4.41	4.09	4.42
RMSE	7.49	8.13	8.24	8.26	3.86	4.12	3.96	4.18
	$\text{Var}(z_{ij}) = 3.63$; $\eta = 15$; $N_A = 100$; $m = 40$; $n = 1000$				$\text{Var}(z_{ij}) = 3.63$; $\eta = 15$; $N_A = 1000$; $m = 200$; $n = 1500$			
$\hat{D}$	17.21 (6.76)	16.09 (7.83)	16.36 (7.09)	16.50 (8.00)	16.39 (2.64)	15.14 (3.00)	16.44 (2.84)	15.63 (3.55)
Sd	6.89	8.03	7.12	8.28	2.46	3.29	2.55	3.51
RMSE	6.88	7.83	7.10	8.02	2.67	3.11	2.88	3.56
	$\text{Var}(z_{ij}) = 0.27$; $\gamma = 1.22$; $\eta = 45$; $N_A = 100$; $m = 40$; $n = 1000$				$\text{Var}(z_{ij}) = 0.27$; $\gamma = 1.22$; $\eta = 45$; $N_A = 1000$; $m = 200$; $n = 2500$			
$\hat{D}$	18.72 (6.67)	18.91 (7.40)	18.75 (6.92)	15.92 (7.76)	16.78 (3.05)	16.10 (3.21)	16.85 (3.11)	15.59 (3.27)
Sd	7.09	7.76	6.40	7.25	3.49	3.57	3.33	3.28
RMSE	7.22	7.96	7.46	7.76	3.16	3.21	3.24	3.29
	$\text{Var}(z_{ij}) = 3.63$; $\eta = 45$; $N_A = 100$; $m = 40$; $n = 1000$				$\text{Var}(z_{ij}) = 3.63$; $\eta = 45$; $N_A = 1000$; $m = 200$; $n = 2500$			
$\hat{D}$	18.50 (7.19)	17.61 (8.05)	16.87 (7.40)	16.48 (8.29)	16.38 (2.62)	15.07 (3.01)	15.97 (2.66)	15.74 (3.13)
Sd	7.48	8.65	6.20	7.99	2.42	3.40	2.63	3.34
RMSE	7.63	8.22	7.46	8.31	2.66	3.14	2.66	3.14

5.11 Application of iREM-D to real data at BCI

In this section we analyse the data at BCI, Panama, for nine terrestrial mammal species using an iREM-D. For the speed data we fit the three alternative models (gamma, lognormal, Weibull) used throughout this thesis. We assume a halfnormal model for distance data (Section 5.4), and a von Mises model (Section 5.8) for the angular data.

First, we assume the encounters follow a Poisson REM. So we fit a Poisson iREM-D to the data, and compare the estimates from a Poisson iREM-dd, a Poisson REM and a REM. We assume independence between the data sets. Rowcliffe et al. (2011) investigated the independence between the detection distance and angle to detection data by looking at the correlation between the two data sets for the most abundant species (agouti) at BCI, Panama, and a weak correlation was found to exists between the data sets (Section 5.2.1). As such an assumption of independence was made. The correlation between the detection distance and angle is given in Table 5.11.1 for each of the nine species. A weak negative correlation exists between the data sets for all species, except mouse, which has a weak positive correlation.

The results (Table 5.11.1) show minimal differences in the estimates from the speed data models. The estimated effective radius, $\hat{\gamma}$, is quite low for the rat species and mouse species. Note that these animals are typically small and are more likely to be missed if they pass under the field of view of the camera traps; also detectability declines the further away they are from the trap. The number of detections, particularly for mouse, was relatively low (43) compared with the other species (see Table 5.3.2 above). For larger species such as peccary, brocket, paca and ocelot estimated effective detection radius is larger. We note that detections very close to the trap are low (see Figure 5.4.1 above), which may be because animals are either unwilling or physically unable to pass very close to the tree to which the trap is attached. This is particularly true for larger animals. Therefore, a larger number of animals would be detected further away from the trap, and hence, larger estimates of the effective detection radius (see Buckland et al., 2001; Rowcliffe et al., 2011).

In Table 5.11.2 we compare estimates of the density from all four methods: REM, iREM, iREM-dd and iREM-D. We fit the nonparameteric REM method (Rowcliffe et al., 2008), given in equation (5.5.1), using the sample averages for the detection distance data, animal speed data, and angle to detection data. Estimates of the density from fitting a Poisson iREM-D are compared with estimates from a Poisson REM, Poisson iREM and a Poisson iREM-dd. The results (Table 5.11.2) show that estimates of the density from the four methods are similar for all species. However, REM and iREM gave smaller estimates of the standard error. It is worth reiterating that REM does not consider the sampling variability in the data, and iREM only considers the variability in the speed data and not the distance and angular data, hence, the underestimation of the standard error compared with iREM-dd and iREM-D. The difference in estimates of the density and its standard error between iREM-dd and iREM-D is minimal since the variability in the angle to detection data is low (see Table 5.8.1). We also note that either of the speed data models can be used in the model to speed data as these gave similar estimates of the parameters.

We also fit a NB iREM-D to the the BCI data and the results are given in Table 5.11.3. Like the Poisson iREM-D the difference in the parameter estimates between the three speed data models is minimal. We also compare estimates from REM (using equation (5.5.1)), a NB iREM-D, a NB iREM-dd and a NB iREM and the results are shown in Table 5.11.4. Again, the four methods gave similar estimates of the density but the estimated standard errors from REM and iREM are smaller compared with estimates from iREM-dd and iREM-D. Note that the Poisson REM (in Table 5.11.1 above) and a NB REM in Table 5.11.3 below can be used to provide estimates of the density but the estimated standard error of the density from a Poisson REM is smaller compared with that from a NB REM. This result is confirmed by the simulation study in Section 5.10.2 above. We compare a Poisson iREM-D with a NB iREM-D and the ΔAIC values are given in Table 5.11.5. The results show that the a NB iREM-D is selected for all species with the difference in AICs of at least 100 units between the two models. For the speed data models, the lognormal model is more supported, followed by the gamma model.

Table 5.11.1: Estimates of density (D, km^2), animal speed (μ, ms^{-1}) effective detection distance (ρ, m) and effective detection angle (υ, radians) from fitting a Poisson iREM-D (standard error in parentheses).

Species		**speed data models**											
		gamma				lognormal				Weibull			
	corr(z_{ij}, θ_{ij})	$\hat{D}$	$\hat{\mu}$	$\hat{\gamma}$	$\hat{\upsilon}$	$\hat{D}$	$\hat{\mu}$	$\hat{\gamma}$	$\hat{\upsilon}$	$\hat{D}$	$\hat{\mu}$	$\hat{\gamma}$	$\hat{\upsilon}$
ocelot	-0.12	0.96 (0.09)	0.41 (0.02)	2.15 (0.13)	0.24 (0.02)	0.96 (0.09)	0.41 (0.03)	2.15 (0.19)	0.24 (0.02)	0.96 (0.09)	0.41 (0.02)	2.15 (0.19)	0.24 (0.02)
coati	-0.15	2.68 (0.27)	0.30 (0.02)	2.41 (0.15)	0.17 (0.02)	2.60 (0.30)	0.33 (0.03)	2.49 (0.26)	0.20 (0.02)	2.68 (0.25)	0.32 (0.02)	2.22 (0.25)	0.17 (0.02)
rat	-0.21	6.73 (0.58)	0.23 (0.01)	1.47 (0.07)	0.24 (0.07)	6.73 (0.67)	0.23 (0.02)	1.47 (0.07)	0.23 (0.02)	6.73 (0.68)	0.23 (0.02)	1.47 (0.07)	0.23 (0.02)
peccary	-0.15	11.65 (0.80)	0.31 (0.02)	3.47 (0.15)	0.27 (0.02)	11.99 (0.82)	0.30 (0.02)	3.53 (0.34)	0.29 (0.02)	11.05 (0.78)	0.27 (0.01)	3.65 (0.32)	0.27 (0.02)
brocket	-0.06	2.80 (0.25)	0.23 (0.01)	3.26 (0.18)	0.23 (0.02)	2.80 (0.29)	0.23 (0.02)	3.26 (0.41)	0.24 (0.02)	2.80 (0.29)	0.23 (0.02)	3.26 (0.41)	0.24 (0.02)
paca	-0.17	3.83 (0.26)	0.27 (0.01)	2.55 (0.11)	0.23 (0.02)	3.83 (0.28)	0.27 (0.01)	2.55 (0.20)	0.23 (0.02)	3.83 (0.30)	0.27 (0.02)	2.55 (0.20)	0.23 (0.02)
agouti	-0.13	43.94 (1.44)	0.24 (0.01)	2.32 (0.04)	0.23 (0.01)	42.83 (1.54)	0.24 (0.01)	2.48 (0.07)	0.24 (0.01)	44.06 (1.38)	0.24 (0.01)	2.35 (0.08)	0.24 (0.01)
squirrel	-0.18	2.81 (0.23)	0.26 (0.01)	2.01 (0.13)	0.24 (0.03)	2.81 (0.24)	0.26 (0.03)	2.01 (0.19)	0.24 (0.03)	2.24 (0.18)	0.29 (0.01)	1.93 (0.18)	0.30 (0.02)
mouse	0.05	1.05 (0.17)	0.17 (0.02)	1.30 (0.08)	0.20 (0.03)	1.05 (0.18)	0.17 (0.02)	1.30 (0.08)	0.20 (0.03)	1.04 (0.18)	0.17 (0.02)	1.30 (0.08)	0.20 (0.03)

Table 5.11.2: Estimates of density (D, km^2) from the four methods (REM, Poisson iREM, Poisson iREM-dd and Poisson iREM-D) where detection distance is assumed to follow a Weibull distribution (standard error in parentheses).

Species	**speed data models**									
			gamma			lognormal			Weibull	
	REM	iREM-D	iREM-dd	iREM	iREM-D	iREM-dd	iREM	iREM-D	iREM-dd	iREM
	$\hat{D}$	$\hat{D}$	$\hat{D}$	$\hat{D}$	$\hat{D}$	$\hat{D}$	$\hat{D}$	$\hat{D}$	$\hat{D}$	$\hat{D}$
ocelot	0.96 (0.05)	0.96 (0.09)	0.96 (0.09)	0.96 (0.08)	0.96 (0.10)	0.96 (0.10)	0.94 (0.08)	0.96 (0.09)	0.96 (0.09)	0.96 (0.07)
coati	2.60 (0.09)	2.68 (0.27)	2.36 (0.24)	2.60 (0.20)	2.60 (0.30)	2.60 (0.30)	2.52 (0.24)	2.60 (0.28)	2.57 (0.27)	2.59 (0.21)
rat	6.73 (0.22)	6.73 (0.58)	6.20 (0.54)	6.73 (0.51)	6.73 (0.67)	6.73 (0.67)	6.71 (0.60)	6.73 (0.68)	6.73 (0.67)	6.68 (0.53)
peccary	11.65 (0.16)	11.65 (0.80)	11.37 (0.69)	11.65 (0.61)	11.99 (0.82)	1.65 (0.86)	11.65 (0.69)	11.05 (0.78)	11.70 (0.74)	11.55 (0.62)
brocket	2.31 (0.08)	2.80 (0.25)	2.80 (0.25)	2.80 (0.21)	2.80 (0.29)	2.80 (0.29)	2.79 (0.25)	2.80 (0.29)	2.80 (0.29)	2.77 (0.22)
paca	3.83 (0.12)	3.83 (0.26)	3.83 (0.26)	3.83 (0.21)	3.83 (0.28)	3.83 (0.28)	3.87 (0.23)	3.83 (0.30)	3.83 (0.30)	3.79 (0.23)
agouti	44.43 (0.42)	43.94 (1.44)	43.05 (1.39)	44.44 (1.17)	42.83 (1.54)	43.38 (1.72)	43.99 (1.39)	44.06 (1.38)	43.39 (1.41)	44.06 (1.21)
squirrel	2.87 (0.12)	2.81 (0.23)	2.72 (0.21)	2.81 (0.13)	2.81 (0.24)	2.81 (0.23)	2.78 (0.14)	2.24 (0.18)	2.40 (0.20)	2.79 (0.13)
mouse	1.05 (0.10)	1.05 (0.17)	1.05 (0.17)	1.05 (0.16)	1.05 (0.19)	1.05 (0.19)	1.05 (0.17)	1.04 (0.18)	1.05 (0.20)	1.04 (0.16)

Table 5.11.3: Estimates of density (D, km^2), animal speed (μ, ms^{-1}), effective detection distance (ρ, m) and effective detection angle (υ, radians) from fitting a NB iREM-D (standard error in parentheses).

Species		**speed data models**											
		gamma				lognormal				Weibull			
	corr(z_{ij}, θ_{ij})	$\hat{D}$	$\hat{\mu}$	$\hat{\gamma}$	$\hat{\upsilon}$	$\hat{D}$	$\hat{\mu}$	$\hat{\gamma}$	$\hat{\upsilon}$	$\hat{D}$	$\hat{\mu}$	$\hat{\gamma}$	$\hat{\upsilon}$
ocelot	-0.12	0.96 (0.11)	0.41 (0.02)	2.15 (0.13)	0.24 (0.02)	0.96 (0.12)	0.41 (0.03)	2.15 (0.19)	0.24 (0.02)	0.96 (0.11)	0.41 (0.02)	2.15 (0.19)	0.24 (0.02)
coati	-0.15	2.60 (0.36)	0.33 (0.02)	2.49 (0.16)	0.20 (0.02)	2.60 (0.40)	0.33 (0.03)	2.49 (0.28)	0.20 (0.02)	2.60 (0.39)	0.33 (0.03)	2.49 (0.28)	0.20 (0.02)
rat	-0.21	6.73 (0.75)	0.23 (0.01)	1.47 (0.07)	0.24 (0.02)	6.73 (0.83)	0.23 (0.02)	1.47 (0.07)	0.24 (0.02)	6.73 (0.83)	0.23 (0.02)	1.47 (0.07)	0.24 (0.02)
peccary	-0.15	11.65 (1.23)	0.31 (0.02)	3.47 (0.15)	0.27 (0.02)	11.65 (1.27)	0.31 (0.02)	3.47 (0.39)	0.27 (0.02)	11.65 (1.31)	0.31 (0.02)	3.47 (0.38)	0.27 (0.02)
brocket	-0.06	2.31 (0.23)	0.28 (0.01)	3.26 (0.18)	0.24 (0.02)	2.31 (0.25)	0.28 (0.02)	3.26 (0.41)	0.24 (0.02)	2.31 (0.25)	0.28 (0.02)	3.26 (0.41)	0.24 (0.02)
paca	-0.17	3.74 (0.34)	0.27 (0.01)	2.55 (0.11)	0.23 (0.02)	3.87 (0.29)	0.27 (0.01)	2.25 (0.10)	0.23 (0.02)	3.78 (0.28)	0.28 (0.01)	2.25 (0.10)	0.23 (0.02)
agouti	-0.13	44.43 (2.20)	0.26 (0.01)	2.43 (0.03)	0.23 (0.01)	44.43 (2.31)	0.26 (0.01)	2.35 (0.09)	0.23 (0.01)	44.43 (2.31)	0.26 (0.01)	2.43 (0.03)	0.23 (0.01)
squirrel	-0.18	2.87 (0.45)	0.26 (0.03)	2.01 (0.13)	0.24 (0.03)	2.87 (0.44)	0.26 (0.03)	2.01 (0.19)	0.24 (0.03)	2.87 (0.45)	0.26 (0.03)	2.01 (0.19)	0.24 (0.03)
mouse	0.05	1.05 (0.25)	0.17 (0.02)	1.40 (0.11)	0.20 (0.03)	1.05 (0.26)	0.17 (0.02)	1.40 (0.11)	0.20 (0.03)	1.05 (0.27)	0.17 (0.03)	1.40 (0.11)	0.20 (0.03)

Table 5.11.4: Estimates of density (D, km^2) from the four methods (REM, NB iREM, NB iREM-dd and NB iREM-D) where detection distance is assumed to follow a Weibull distribution (standard error in parentheses).

Species	**speed data models**									
			gamma			lognormal			Weibull	
	REM	iREM-D	iREM-dd	iREM	iREM-D	iREM-dd	iREM	iREM-D	iREM-dd	iREM
	$\hat{D}$	$\hat{D}$	$\hat{D}$	$\hat{D}$	$\hat{D}$	$\hat{D}$	$\hat{D}$	$\hat{D}$	$\hat{D}$	$\hat{D}$
ocelot	0.96 (0.05)	0.96 (0.11)	0.96 (0.11)	0.96 (0.09)	0.96 (0.12)	0.96 (0.12)	0.96 (0.10)	0.96 (0.11)	0.96 (0.11)	0.96 (0.09)
coati	2.60 (0.09)	2.60 (0.36)	2.60 (0.36)	2.60 (0.31)	2.60 (0.40)	2.60 (0.40)	2.60 (0.35)	2.60 (0.39)	2.60 (0.39)	2.60 (0.34)
rat	6.73 (0.22)	6.73 (0.75)	6.73 (0.75)	6.73 (0.66)	6.73 (0.83)	6.73 (0.82)	6.73 (0.75)	6.73 (0.83)	6.73 (0.82)	6.73 (0.75)
peccary	11.65 (0.16)	11.65 (1.23)	11.65 (1.26)	11.65 (0.71)	11.65 (1.27)	11.65 (1.26)	11.65 (1.11)	11.65 (1.31)	11.65 (1.30)	11.65 (1.16)
brocket	2.31 (0.08)	2.31 (0.23)	2.31 (0.23)	2.31 (0.19)	2.31 (0.25)	2.31 (0.25)	2.31 (0.21)	2.31 (0.25)	2.31 (0.25)	2.31 (0.21)
paca	3.83 (0.12)	3.89 (0.35)	3.80 (0.35)	3.81 (0.30)	3.89 (0.36)	3.84 (0.36)	3.85 (0.31)	3.75 (0.35)	3.77 (0.35)	3.77 (0.31)
agouti	44.43 (0.42)	44.43 (2.20)	44.43 (2.21)	44.65 (1.85)	44.43 (2.31)	44.43 (2.15)	44.38 (1.97)	44.43 (2.31)	44.43 (2.15)	44.43 (1.92)
squirrel	2.87 (0.12)	2.87 (0.45)	2.87 (0.44)	2.87 (0.40)	2.87 (0.44)	2.87 (0.44)	2.87 (0.39)	2.87 (0.45)	2.87 (0.45)	2.87 (0.41)
mouse	1.05 (0.10)	1.05 (0.25)	1.05 (0.25)	1.05 (0.23)	1.05 (0.26)	1.05 (0.26)	1.05 (0.24)	1.05 (0.27)	1.05 (0.27)	1.05 (0.25)

Table 5.11.5: ΔAIC values from fitting a Poisson iREM-D and a NB iREM-D to the BCI data.

Species	**Poisson iREM-D**			**NB iREM-D**		
	gamma	lognormal	Weibull	gamma	lognormal	Weibull
ocelot	122.18	129.14	122.26	0	6.98	0.08
coati	1685.20	1681.06	1690.36	7.00	0	9.30
rat	957.52	944.20	969.32	13.31	0	25.63
peccary	10205.28	10253.16	10238.94	15.02	0	10.48
brocket	212.29	196.91	221.95	15.38	0	25.63
paca	1366.58	1342.46	1402.88	0	85.08	145.48
agouti	7432.56	7434.26	7495.04	0	1.69	62.47
squirrel	7522.94	7437.44	7586.14	85.52	0	148.72
mouse	222.34	217.04	229.28	5.29	0	12.22

5.12 Discussion

The iREM method and its extensions developed in Chapters 3, 4, and in this Chapter combines multiple data sets into a comprehensive framework (iREM-D). Consequently, estimates of the density and additional parameters in the model, in particular effective detection distance are made with correct estimations of the standard error of the density.

As shown by the simulations, accounting for the variation in the parameters required to estimate abundance in REM is important. For high variability in the data, REM is a poor fit for small density values. However, REM can provide stable and relatively accurate estimates of the density for large density values but ignoring the variability in detection zone dimensions would result in smaller estimates of the standard error for scenarios that can be considered ecologically realistic. While iREM is an improvement over REM in accounting for the sampling variability in animal speed, it does not account for the sampling variability in the detection distance and angle to detection, and as shown by the simulation results estimation bias and underestimation of the standard error can be introduced.

The simulation results also revealed that not accounting for the variability in the rate of encounter would have some consequences on the standard error, particularly if the sampling variability in the detection distance is not accounted for. The NB REM accounts for the variability in the encounter rate. This is not included in a Poisson REM but comparable results to the NB REM are produced despite the simplicity of the model.

An analysis of the BCI data shows obvious variation in detection distance data, and clear differences in estimated standard error of density between iREM and iREM-dd. However, we have shown that iREM can obtain relatively accurate estimates of the density if there is large variation in the detection distance and angle. Since the process of collecting detection distance and angular data is tedious as described in this Chapter and Rowcliffe et al. (2011), iREM, which is a simpler model can provide relatively accurate estimates of the density but it would give smaller estimates of the standard error.

The single integrated model (iREM-D) developed in this Chapter is flexible and has the advantage of incorporating other factors as covariates such as activity level and habitat type/land-cover (as shown in Chapter 4), which may describe additional variation in model parameters, particularly, density. But as shown in the real data at BCI, Panama, the variation in angle to detection is minimal resulting in similar estimates of the density from iREM-dd. We have shown that a weak correlation exists between the distance and angle data. Hence, the correlation is not important in practice and we can assume independence between the data sets and their contribution to the likelihood could be multiplied. While the iREM-dd is a simpler model to fit compared with iREM-D and can provide an adequate description of our BCI data set, in general, the iREM-D would be more appropriate for modelling this type of data as it accounts for the potential variation in the direction of entry of animals into the detection zone, which is relevant if the variation within species is large.

In REM or an iREM method density is dependent on expected animal speed for the population. In most cases in practice, a sample of the population is used and an average of the speed observations are used in the estimation process. However, faster moving animals are more likely to contact camera traps, and we therefore expect our sample to be biased towards faster speeds. To correct for this bias, we adopt a size biased sampling approach in estimating expected animal speed of movement. Chapter 6 describes in detail, a size biased sampling method that can reduce the bias in expected animal speed.

Chapter 6

Measuring Animal Speed From Camera Trap Data

The motivation for this Chapter is the animal speed of movement data collected from a survey of terrestrial mammals at Barro Colorado Island (BCI), Panama, which was discussed in Chapter 5, Section 5.3. The speed of movement of animals is an important behavioural metric that influences processes such as foraging success, and disease transmission, and which can be applied to a range of questions in ecology and consevation (Rowcliffe et al., 2016). For example, the estimation of animal abundance in the Random Encounter Model (REM) developed by Rowcliffe et al. (2008) requires an estimate of average animal speed of movement. Also, Rowcliffe et al. (2011) used camera traps, recording animals to measure movement paths of animals at a very fine scale. The speed observation was obtained by dividing the length of a passage travelled by an animal by its duration. From a sample of speeds, an average speed of movement was computed. The motivation for Rowcliffe et al. (2011) was to provide a means of estimating detection zone dimensions for application to REM estimation of abundance.

In camera trapping analysis faster moving animals are more likely to encounter camera traps than slower moving animals (Hutchinson and Waser, 2007). As such, the probability of encounter may differ within species. To correct for this bias in speed of movement of animals we can use the method of size biased sampling (Patil and Rao, 1978). Size biased sampling arises when a positive-valued outcome variable is sampled

with a selection probability proportional to its size (Chen, 2010). According to Arratia and Goldstein (2010), a random variable, X, can be size biased if and only if it is nonnegative. Rowcliffe et al. (2016), for example, used size biased sampling methods to estimate average speed from camera trap data by fitting size biased models to speed data and maximizing the likelihood. The purpose is to correct for the bias in average speed required in Rowcliffe et al. (2008) REM to estimate density. In Chapter 2 we showed that REM describes a linear relationship between speed and encounter rate, therefore the probability of sampling a given speed is expected to be proportional to its speed, and the observed distribution of the speeds is therefore size biased.

Rowcliffe et al. (2016) considered three non-negative probability density functions: gamma, lognormal, Weibull, to fit the speed data and estimate average speed. The variance was approximated by inverting the Hessian matrix at the maximum likelihood estimates, and the AIC method was used to select the best distribution. To test the models Rowcliffe et al. (2016) generated a sample of d_i $(i = 1, 2, ..., n)$ distances, and passage durations, t_i $(i = 1, 2, ..., n)$ and computed the ith speed observation, s_i, as the ratio of distance, d_i to passage duration t_i, that is $s_i = d_i/t_i$, and using these speeds a size biased gamma model, a size biased lognormal model and a size biased Weibull model were fitted and the likelihoods maximized. An application of the size biased models to animal speed data of 12 terrestrial mammal species at Barro Colorado Island, BCI, Panama suggested that the method is acceptably accurate with the lognormal model most strongly supported, particularly for those species with large sample sizes.

In this Chapter, we adopt the method of size biased sampling to correct for the bias in speed of faster moving animals, and estimate average speed, which is an input parameter in REM. In particular, we derive the probability of encountering an animal with a given speed in REM showing that it is approximately proportional to its speed. We assume four probability density functions for which the probability of encounter could be derived and test the approximation via simulations. We also test the true probability of encountering an animal via simulations. The Chapter begins with an introduction to size biased sampling in Section 6.1. It then goes on, in Section 6.2, to describe size biased sampling in REM, and the proposed probability models for animal encounters

used to derive the probability of encounter. To obtain estimates of average speed we adopt the method of maximum likelihood estimation, which is discussed in Section 6.3. The harmonic mean method, which is a nonparametric alternative approach that can be used to estimate the average speed, is described in Section 6.4. The simulation study is given in given in Section 6.5. The Chapter concludes with an application of the size biased sampling method to BCI data in Section 6.6.

6.1 Introduction to size biased sampling

The concept of size biased sampling has resulted from the concept of weighted distributions. Suppose that an original observation X has $f(x)$ as the probability density function, and that the probability of recording the observation x is $0 \leq w(x) \leq 1$, then the probability density function of X_w, the recorded observation is

$$g_w(x) = \frac{w(x)f(x)}{w}, \tag{6.1.1}$$

where $w(x)$ is an arbitrary non-negative weight function; and w is the normalising factor needed to make the probability density integrate to unity. Distributions with arbitrary $w(x)$ are called weighted distributions. The weighted distribution with $w(x) = x$ is called a size biased distribution with probability density function

$$g_w(x) = \frac{xf(x)}{\mathbb{E}(X)} \tag{6.1.2}$$

(see Patil and Rao, 1978). The prove this let us suppose that m animals with speeds $x = \{x_1, x_2, ..., x_m\}$ such that $l = 1, 2, ..., m$ have probability $w(x)$ of being recorded, and expected speed of movement, $\mathbb{E}(X) = \mu$. If the distribution is size biased then $w(x) = x$, and the probability density function is

$$g_w(x) = \frac{xf(x)}{w} = kxf(x),$$

where $f(x)$ is the true probability density function of x and k is a constant. Integrating over the entire space of x, we have

$$\int_x g_w(x)dx = 1.$$

Since the random variable x can be size biased if and only if it is nonnegative, the constant k can be found as follows,

$$\begin{aligned}\int_0^\infty kxf(x)dx &= k\int_0^\infty xf(x)dx \\ 1 &= k\int_0^\infty xf(x)dx = k\mathbb{E}(x) \\ 1 &= k\mu \\ k &= \frac{1}{\mu}.\end{aligned}$$

Therefore,

$$g_w(x) = \frac{xf(x)}{\mu}.$$

6.2 Size biased sampling in REM

The Random Encounter Model (REM) describes the rate of encounter between animals and camera traps, λ, as a function of the detection zone dimensions (r, θ), average animal speed of movement ($\bar{v}$) and density (D) such that

$$\lambda = \frac{2+\theta}{\pi} rt\bar{v}D. \tag{6.2.1}$$

Suppose the population size is 1, and the area considered is 1, in appropriate units, this gives a density, D, of 1 since density = population/area. In this case we then consider the expected encounter rate of an individual with speed x_l to be

$$\lambda = \frac{2+\theta}{\pi} rtx_l. \tag{6.2.2}$$

We can simplify this to

$$\lambda = Cx_l$$

where $C = \{(2+\theta)/\pi\}rt$. A linear relationship exists between speed and the encounter rate, λ, therefore, we would expect that the probability of sampling an animal with a given speed should be proportional to its speed. If the probability of encountering an animal is proportional to λ we can use size biased sampling described above in Section 6.1. An alternative method is the harmonic mean, which is a nonparametric method and this is discussed in Section 6.4.

Here we propose the probability models used in Chapter 3 for encountering animals, which include: 1) a Poisson REM, 2)ZIP REM, 3) a NB REM, and 4) a ZINB REM, and we assume the speeds of movement follow 1) a gamma model, 2) a lognormal model or 3) a Weibull model described in Chapter 3. If we let

$$A_l = \begin{cases} 1 & \text{if individual } l \text{ is encountered,} \\ 0 & \text{if individual } l \text{ is not encountered,} \end{cases}$$

then for animals moving at speeds x_l (for $l = 1, 2, ..., m$) we can show that the probability of encountering an animal with a given speed is approximately proportional to the expected encounter rate, λ, defined in equation (6.2.2) for the four proposed probability models. We also provide examples and give the conditions for which the approximation holds.

6.2.1 Poisson REM

A Poisson model has probability density defined as

$$P(x_l;\ \lambda) = \frac{e^{-\lambda}\lambda^{x_l}}{x_l!}.$$

The probability that an individual animal is not encountered is

$$\mathrm{P}(A_l = 0) = e^{-\lambda},$$

and therefore,

$$\begin{aligned} P(A_l = 1) &= 1 - \mathrm{P}(A_l = 0) \\ &= 1 - e^{-\lambda} \\ &= 1 - \left(1 - \lambda + \frac{\lambda^2}{2!} - \frac{\lambda^3}{3!} + ... + \frac{(-1^m)\lambda^m}{m!}\right) \\ &\approx \lambda = \mathrm{C}x_l \propto x_l, \end{aligned}$$

where $\mathrm{C} = \{(2+\theta)/\pi\} rtD$. As $P(A_l = 1) \propto$ (approximately proportional to) x_l, the distribution is size biased and we can apply equation (6.1.2) with $w(x) = x$, and normalising constant $w = \mu$, where μ is the expected animal speed of movement. For the approximation to hold, λ needs to be small. Note that λ is a function of several parameters and the random variable x_l, that is,

$$\lambda = \frac{2+\theta}{\pi} rtx_l. \tag{6.2.3}$$

We note that for λ to be small, t needs to be small. As a demonstration we compare the true probability of encounter with the approximation for different parameter values for: detection distance r (in km), detection angle θ (in radians), and camera trap time, t (in days); and animal speed x_l (in km/day^{-1}), which we give in Table 6.2.1 below. Table 6.2.1 shows the true probability of encounter and the approximation from a Poisson REM. For small values of t, estimates of λ are small and the approximation holds; the smaller the value of t, the smaller λ gets. The approximation also holds for different values of the detection zone dimensions and animal speed of movement for a short time frame where the encounters occur. It is worth noting that the smaller the input parameters, the smaller the difference between the approximation, λ, and the true probability.

Table 6.2.1: True probability of encounter and its approximation from a Poisson REM.

θ (radians)	r (km)	x_l (km/day^{-1})	t in days (hours)	Approximate probability λ	Exact probability $1 - e^{-\lambda}$
0.175	0.012	30.7584	0.0416667 (1)	0.01091657	0.0108572
0.175	0.012	30.7584	0.104167 (2.5)	0.0272915	0.02692245
0.175	0.012	30.7584	0.22916 (5.5)	0.06003936	0.05827253
0.175	0.012	30.7584	0.4375 (10.5)	0.1146239	0.1082956
0.175	0.012	30.7584	0.854117 (20.5)	0.2237766	0.2005063
0.175	0.012	30.7584	0.0416667 (1)	0.0106474	0.01059091
0.175	0.012	30.7584	0.4375 (10.5)	0.1117976	0.1057748
0.175	0.012	9.849	0.0146667 (1)	0.003400956	0.003403754
0.175	0.012	9.849	0.4375 (10.5)	0.03580035	0.0351671
0.175	0.012	2.7648	0.041667 (1)	0.0009570695	0.0009566116
0.175	0.012	2.7648	0.4375 (10.5)	0.01004922	0.009998897
0.271	0.012	9.849	0.0416667 (1)	0.003560051	0.003553721
0.271	0.012	9.849	0.4375 (10.5)	0.0373805	0.03669048
0.492	0.012	9.849	0.0146667 (1)	0.003906494	0.003898873
0.492	0.012	9.849	0.4375 (10.5)	0.04101815	0.04018829
0.175	0.001	9.849	0.0416667 (1)	0.0003255411	0.0003254881
0.175	0.001	9.849	0.4375 (10.5)	0.003418179	0.003412344
0.175	0.024	9.849	0.0146667 (1)	0.007812987	0.007782545
0.175	0.024	9.849	0.4375 (10.5)	0.0820363	0.07876148

6.2.2 ZIP REM

To account for zero-inflation in the data set we assume encounters of animals with speeds x_l follow a ZIP REM with probability density function

$$f_{zip}(x_l \mid \rho,\ \lambda) = \begin{cases} \rho + (1-\rho)e^{-\lambda}, & \text{for } x_l = 0, \\ \\ \dfrac{(1-\rho)e^{-\lambda}\lambda^{x_l}}{x_l!}, & \text{for } x_l > 0, \end{cases}$$

where $0 \leq \rho \leq 1$ is the probability which inflates the zero response category. Therefore, the probability of not encountering an individual animal is

$$\mathrm{P}(A_l = 0) = \rho + (1-\rho)e^{-\lambda}.$$

From this we can compute the probability of encountering an animal, such that

$$\begin{aligned} \mathrm{P}(A_l = 1) &= 1 - \left\{\rho + (1-\rho)e^{-\lambda}\right\} \\ &= 1 - \left\{\rho + (1-\rho)\left(1 - \lambda + \frac{\lambda^2}{2!} - \frac{\lambda^3}{3!} + ... + \frac{(-1^m)\lambda^m}{m!}\right)\right\} \\ &\approx 1 - \{\rho + (1-\rho)(1-\lambda)\} = \lambda(1-\rho), \end{aligned}$$

and we know that $\lambda = Cx_l$. Therefore,

$$P(A_l = 1) = \lambda(1-\rho) = C(1-\rho)x_l \propto x_l,$$

where $\mathrm{C} = \{(2+\theta)/\pi\}\, rtD$. As $P(A_l = 1) \propto x_l$, then the distribution is size biased and we can apply equation (6.1.2) with $w(x) = x$, and normalising constant $w = \mu$. The approximation will hold for small values of t. Table 6.2.2 gives an example of the approximation and the exact probability of encounter from a ZIP REM. For faster speeds the approximation holds when λ and camera trap time period, t are small. The difference between the approximation and the exact probability increases when ρ and t increase. For slower speeds the approximation of the probability of encounter holds because, λ, which is dependent on x, is small.

Table 6.2.2: True probability of encounter and its approximation from a zero-inflated Poisson REM.

θ (radians)	r (km)	x_l (km/day^{-1})	t in days (hours)	ρ	Encounter rate λ	Approximate probability $\lambda(1-\rho)$	Exact probability $1-\left\{\rho+(1-\rho)e^{-\lambda}\right\}$
0.175	0.012	30.75	0.0417 (1)	0.10	0.011	0.010	0.010
0.175	0.012	30.76	0.1042 (2.5)	0.10	0.027	0.024	0.024
0.175	0.012	30.76	0.2292 (5.5)	0.10	0.059	0.053	0.051
0.175	0.012	30.76	0.0417 (1)	0.40	0.011	0.007	0.006
0.175	0.012	30.76	0.1042 (2.5)	0.40	0.027	0.016	0.016
0.175	0.012	30.76	0.2292 (5.5)	0.40	0.059	0.035	0.034
0.175	0.012	2.76	0.0417 (1)	0.40	0.0010	0.001	0.0006
0.175	0.012	2.76	0.1042 (2.5)	0.40	0.002	0.001	0.001
0.175	0.012	2.76	0.2292 (5.5)	0.40	0.005	0.003	0.003

6.2.3 NB REM

A negative binomial model with NB-2 form (Hilbe, 2011) has probability density

$$P(x_l \mid \lambda,\ \kappa) = \frac{\Gamma(\kappa + x)}{\Gamma(a_i + 1)\Gamma(\kappa)} \left(\frac{\kappa}{\lambda + \kappa}\right)^{\kappa} \left(\frac{\lambda}{\lambda + \kappa}\right)^{x_l}, \tag{6.2.4}$$

where $\kappa > 0$, and has a dispersion parameter $1/\kappa$, then the probability of not encountering an individual is

$$\mathrm{P}(A_l = 0) = \left(\frac{\kappa}{\lambda + \kappa}\right)^{\kappa}.$$

From this we can compute the probability of encountering an individual, which is

$$\begin{aligned}
P(A_l = 1) &= 1 - \left(\frac{\kappa}{\lambda + \kappa}\right)^{\kappa} \\
&= 1 - \left[\kappa^{\kappa}\left\{\kappa\left(1 + \frac{\lambda}{\kappa}\right)\right\}^{-\kappa}\right] \\
&= 1 - \left\{\left(1 + \frac{\lambda}{\kappa}\right)^{-\kappa}\right\} \\
&= 1 - \left(1 + (-\kappa)\left(\frac{\lambda}{\kappa}\right) + \frac{(-\kappa)(-\kappa - 1)}{2!}\left(\frac{\lambda}{\kappa}\right)^{2} + \ldots\right) \\
&= 1 - \left(1 - \lambda + \frac{\kappa + 1}{2\kappa}\lambda^2 + \ldots\right) \\
&\approx \lambda = Cx_l \propto x_l.
\end{aligned}$$

where C $= \{(2 + \theta)/\pi\}\, rtD$. As $P(A_l = 1) \overset{\propto}{\approx} x_l$, then the distribution is size biased and we can apply equation (6.1.2) with $w(x) = x$, and normalising constant $w = \mu$. Like the Poisson REM, λ needs to be small for the approximation to hold, and for λ to be small t should be small. We have checked the approximation for small and large values of x_l, and typical values of the dispersion parameter, κ.

Table 6.2.3 gives the values of the true probability of encounter and the approximation from a negative binomial REM. The estimations show that for faster speeds and when κ is small the approximation holds for small values of t, but as t increases to 5.5 hours the difference between the exact probability and the approximation increases. A similar

conclusion can be drawn for a larger value of κ. However, the approximation seems to hold for a larger value of κ at increasing camera trap time t for slower speeds.

Table 6.2.3: True probability of encounter and its approximation from a negative binomial REM.

θ (radians)	r (km)	x_l (km/day^{-1})	t in days (hours)	κ	Approximate probability λ	Exact probability $1-\left(\frac{\kappa}{\lambda+\kappa}\right)^{\kappa}$
0.175	0.012	30.75	0.0417 (1)	0.50	0.011	0.0105
0.175	0.012	30.76	0.1042 (2.5)	0.50	0.027	0.026
0.175	0.012	30.76	0.2292 (5.5)	0.50	0.059	0.054
0.175	0.012	30.76	0.0417 (1)	1.50	0.011	0.011
0.175	0.012	30.76	0.1042 (2.5)	1.50	0.027	0.026
0.175	0.012	30.76	0.2292 (5.5)	1.50	0.059	0.056
0.175	0.012	2.76	0.0417 (1)	1.50	0.005	0.005
0.175	0.012	2.76	0.1042 (2.5)	1.50	0.002	0.002
0.175	0.012	2.76	0.2292 (5.5)	1.50	0.005	0.005

6.2.4 ZINB REM

A ZINB probability density function has NB-2 form (Hilbe, 2011) defined as

$$f(x_l \mid \kappa,\ \lambda,\ \rho) = \begin{cases} \rho + (1-\rho)\left(\dfrac{\kappa}{\lambda+\kappa}\right)^{\kappa}, & \text{for } x_l = 0, \\ (1-\rho)\dfrac{\Gamma(\kappa+x_l)}{\Gamma(x_l+1)\Gamma(\kappa)}\left(\dfrac{\kappa}{\lambda+\kappa}\right)^{\kappa}\left(\dfrac{\lambda}{\lambda+\kappa}\right)^{x_l}, & \text{for } x_l > 0, \end{cases}$$

where $0 \leq \rho \leq 1$ is the probability which inflates the zero response category, and $1/\kappa$ is the dispersion parameter, and where $\kappa > 0$. The probability of not encountering an individual is given as

$$P(A_l = 0) = \rho + (1-\rho)\left(\frac{\kappa}{\lambda+\kappa}\right)^{\kappa},$$

therefore, the probability that an individual is encountered is given as

$$\begin{aligned}
P(A_l = 1) &= 1 - \left\{\rho + (1-\rho)\left(\frac{\kappa}{\lambda+\kappa}\right)^{\kappa}\right\} \\
&= 1 - \left[\rho + (1-\rho)\left\{\left(1+\frac{\lambda}{\kappa}\right)^{-\kappa}\right\}\right] \\
&= 1 - \left\{\rho + (1-\rho)\left(1 - \lambda + \frac{\kappa+1}{2\kappa}\lambda^2 + ...\right)\right\} \\
&\approx 1 - \{\rho + (1-\rho)(1-\lambda)\} \\
&= (1-\rho)\lambda,
\end{aligned}$$

and we know that $\lambda = Cx_l$, therefore,

$$P(A_l = 1) = (1-\rho)\lambda = C(1-\rho)x_l \propto x_l.$$

Here $\mathrm{C} = \{(2+\theta)/\pi\} rtD$ is a constant. As $P(A_l = 1) \propto x_l$, then the distribution is size biased and we can apply equation (6.1.2) with $w(x_l) = x_l$, and normalising constant $w = \mu$. The approximation will hold for small values of λ. As shown previously in Table 6.2.3 above, λ will be small when sampling effort, t is small.

In Table 6.2.4 we give a demonstration of the approximation of the probability of encounter under different parameter settings. For the approximation to hold, it is expected that the camera trap time period t and κ should be small. As t increases the difference between the approximation and the exact probability of encounter increases for faster speeds. But for slower speeds the approximation of the probability of encounter holds for higher values of κ and for increasing t.

Table 6.2.4: True probability of encounter and its approximation from a zero-inflated negative binomial REM.

θ (radians)	r (km)	x_l (km/day^{-1})	t in days (hours)	κ	ρ	Encounter rate λ	Approximate probability $\lambda(1-\rho)$	Exact probability $1-\left\{\rho+(1-\rho)\left(\frac{\kappa}{\lambda+\kappa}\right)^{\kappa}\right\}$
0.175	0.012	30.75	0.0417 (1)	0.50	0.10	0.011	0.010	0.009
0.175	0.012	30.76	0.1042 (2.5)	0.50	0.10	0.027	0.024	0.023
0.175	0.012	30.76	0.2292 (5.5)	0.50	0.10	0.059	0.053	0.048
0.175	0.012	30.76	0.0417 (1)	1.50	0.10	0.011	0.010	0.010
0.175	0.012	30.76	0.1042 (2.5)	1.50	0.10	0.027	0.024	0.023
0.175	0.012	30.76	0.2292 (5.5)	1.50	0.10	0.059	0.035	0.05
0.175	0.012	2.76	0.0417 (1)	1.50	0.10	0.0010	0.001	0.001
0.175	0.012	2.76	0.1042 (2.5)	1.50	0.10	0.002	0.002	0.002
0.175	0.012	2.76	0.2292 (5.5)	1.50	0.10	0.005	0.005	0.005

6.2.5 *Points to consider in size biased sampling in REM*

For all four probability models we show that the probability of encountering an animal is approximately proportional to its speed, that is $P(A_l = 1) \propto x_l$, when we consider an individual encountering a camera trap in a short time frame. But as discussed in Chapter 1 and as shown in Chapters 2 and 5, camera traps are usually set up for long time periods and information is usually extracted and aggregated on a daily basis, rather than in hours as shown in the examples above for which the approximation holds. An alternative approach would be to use the exact probability of encounter that can be derived by an analytical approach, which we show in Section 6.3.4. However, there are limitations with the analytical approach in obtaining the exact probability. For instance, it may not be possible to obtain the exact probability of encounter for some models such as a lognormal or Weibull for the speed data, or a NB probability of encounter model. We note that one possible solution to this problem would be to use a numerical approximation to obtain the exact probability. In Section 6.3 we give the size biased probability density functions and the standard probability density functions used for the speed data using the approximation method: a size biased lognormal and a standard lognormal (Section 6.3.1), a size biased gamma model and standard gamma model (Section 6.3.2) and a size biased Weibull model and a standard Weibull model (Section 6.3.3). We also give the size biased probability density function with the exact probability of encounter using an analytical approach to derive the exact probability (Section 6.3.4). We test the approximations for the four probability density functions using simulations, which does not use the approximation, in Section 6.5.

6.3 *Maximum likelihood estimation*

This section explores how to reduce the bias of faster moving animals using a parametric likelihood-based approach. Maximum likelihood estimation is used to work out an estimate for expected animal speed, μ_x. Suppose we have m speed observations such that $x = \{x_1, ..., x_m\}$ and where $l = 1, 2, ..., m$, then the likelihood function can be written as

$$L(\mu_x,\ \nu \mid x_l) = \prod_{l=1}^{m} \frac{x_l f(x_l \mid \mu_x,\ \nu)}{\mu_x}, \tag{6.3.1}$$

where μ_x is the mean speed, $f(\cdot)$ is the true probability density function of speed in the absence of sampling bias, and ν represents any additional parameters in the model. We use the three probability density functions discussed in Chapter 3 to model animal speed: a gamma, a lognormal, and a Weibull. The size biased and standard models are:

6.3.1 Lognormal and size biased lognormal models

The lognormal probability density is

$$f(x_l;\nu,\epsilon) = \frac{1}{x_l\nu\sqrt{2\pi}}\exp\left(-\frac{\{\log(x_l)-\epsilon\}^2}{2\nu^2}\right)$$

with mean $\mu_x = \exp\left(\epsilon + \frac{1}{2}\nu^2\right)$. The size biased lognormal probability density function is

$$\begin{aligned}
g_w(x_l;\nu,\epsilon) &= \frac{\frac{x_l}{x_l\nu\sqrt{2\pi}}\exp\left(-\frac{\{\log(x_l)-\epsilon\}^2}{2\nu^2}\right)}{\mu} = \frac{\frac{x_l}{x_l\nu\sqrt{2\pi}}\exp\left(-\frac{\{\log(x_l)-\epsilon\}^2}{2\nu^2}\right)}{\exp\left(\epsilon+\frac{1}{2}\nu^2\right)} \\
&= \frac{1}{x_l\nu\sqrt{2\pi}}\left[\exp\{\log(x_l)\}\times\exp\left(-\frac{\{\log(x_l)-\epsilon\}^2}{2\nu^2}\right)\times\exp\left(-\left\{\epsilon+\frac{1}{2}\nu^2\right\}\right)\right] \\
&= \frac{1}{x_l\nu\sqrt{2\pi}}\left[\exp\left(-\frac{(\log(x_l)-\epsilon)^2+2\nu^2\left(\epsilon+\frac{1}{2}\nu^2\right)-2\nu^2\log(x_l)}{2\nu^2}\right)\right], \\
&= \frac{1}{x_l\nu\sqrt{2\pi}}\exp\left\{-\frac{1}{2\nu^2}\left(\log(x_l)-(\epsilon+\nu^2)\right)^2\right\},
\end{aligned}$$

which is a lognormal distribution with expected value $\mathbb{E}_{g_w}(x_l) = \exp(\epsilon + \frac{3}{2}\nu^2)$ for $x_l > 0$ (Ratnaparkhi and Naik-Nimbalkar, 2012). Note that for a small ν^2 relative to ϵ, $\mathbb{E}_{f_x}(x_l) \approx \mathbb{E}_{g_w}(x_l)$.

6.3.2 Gamma and size biased gamma models

The probability density function of a gamma is

$$f(x_l;\ \nu,\ \alpha) = \frac{\nu^\alpha}{\Gamma(\alpha)}{x_l}^{\alpha-1}\exp(-\nu x_l)$$

with mean, $\mathbb{E}(x) = \alpha/\nu$. A size biased gamma model is given as

$$g_w(x) = \frac{\frac{x_l \nu^\alpha}{\Gamma(\alpha)} {x_l}^{\alpha-1} \exp(-\nu x_l)}{\frac{\alpha}{\nu}} = \frac{\nu^{\alpha+1}}{\alpha\Gamma(\alpha)} {x_l}^{(\alpha+1)-1} \exp(-\nu x_l)$$
$$= \frac{\nu^{\alpha+1}}{\Gamma(\alpha+1)} {x_l}^{(\alpha+1)-1} \exp(-\nu x_l).$$

Therefore, $g_w(x_l;\ \alpha+1,\ \nu)$ is a gamma distribution with mean $\mathbb{E}_{g_w}(x_l) = (\alpha+1)/\nu$ and variance, $\text{Var}(x_l) = (\alpha+1)/\nu^2$ (Mir et al., 2013). Note that the difference between the mean of a size biased gamma model and the mean of a standard gamma model is $1/\nu$. If ν is large then the difference between the two means will be small, and if ν is large enough $\mathbb{E}_{g_w}(x_l) \approx \mathbb{E}_f(x_l)$; the variance will also be small. If ν is small (the variance will be large) the difference between the two means will be large.

6.3.3 *Weibull and size biased Weibull models*

The probability density function of a Weibull distribution is

$$f(x_l;\ \nu,\ \beta) = \frac{\nu}{\beta}\left(\frac{x_l}{\beta}\right)^{\nu-1} e^{-\left(\frac{x_l}{\beta}\right)^\nu}, \tag{6.3.2}$$

with mean $\mu = \beta\Gamma\left(1+\frac{1}{\nu}\right)$. A size biased Weibull distribution is

$$f(x_l;\ \nu,\ \beta) = \frac{x_l}{\beta\Gamma\left(1+\frac{1}{\nu}\right)}\left(\frac{\nu}{\beta}\right)\left(\frac{x_l}{\beta}\right)^{\nu-1} e^{-\left(\frac{x_l}{\beta}\right)^\nu}$$
$$= \frac{1}{\Gamma\left(1+\frac{1}{\nu}\right)}\left(\frac{\nu}{\beta}\right)\left(\frac{x_l}{\beta}\right)^{\nu} \mathrm{e}^{-\left(\frac{x_l}{\nu}\right)^\nu}$$

(Dey et al., 2015). Note that a size biased Weibull distribution is not a Weibull distribution.

6.3.4 *Size biased gamma function with the true probability of encounter from a Poisson REM*

So far we have considered an approximation for the probability of encounter for size biased sampling. Here we use the true probability of encounter from a Poisson REM

where animal speed is assumed to follow a gamma model. For a Poisson REM, the probability of encounter is

$$\mathrm{P}(A_l = 1) = 1 - e^{-\lambda} = 1 - e^{-Cx_l},$$

where $\mathrm{C} = \{(2+\theta)/\pi\} rtD$, and $x_l = \{x_1, x_2, ..., x_m\}$ is the animal speed observations. The size biased probability density function is given as

$$g_w \propto (1 - e^{-Cx_l}) f(x_l \mid \mu_x,\ \nu)$$

where $f(x_l \mid \mu_x,\ \nu)$ is the true probability density function for the speed observations. In this case, we assume a gamma distribution. Integrating g_w over the entire space of x gives

$$\int_{x_l} g_w(x_l) dx_l = 1.$$

So,

$$\begin{aligned}\int_{x_l} k(1 - e^{-Cx}) f(x_l \mid \mu_x,\ \nu) dx_l &= \int_{x_l} k f(x_l \mid \mu_x,\ \nu) dx_l - \int_{x_l} k e^{-Cx_l} f(x \mid \mu_x,\ \nu) dx_l \\ &= k\left(1 - \int_{x_l} e^{-Cx} f(x_l \mid \mu_x,\ \nu) dx_l\right),\end{aligned}$$

where k is a constant. We assume animal speed follows a gamma model with shape parameter ν, and scale parameter α, therefore, the constant k can be found as follows

$$\begin{aligned}&\int_0^\infty e^{-Cx_l} \frac{\nu^\alpha}{\Gamma(\alpha)} {x_l}^{\alpha-1} e^{-\nu x_l} dx_l \\ &= \nu^\alpha \int_0^\infty \frac{{x_l}^{\alpha-1} e^{-\nu+C}}{\Gamma(\alpha)} dx_l \\ &= \frac{\nu^\alpha}{(\nu+C)^\alpha} \int_0^\infty \frac{(\nu+C)^\alpha}{\Gamma(\alpha)} {x_l}^{\alpha-1} e^{-(\nu+C)} dx_l \\ &= \frac{\nu^\alpha}{(\nu+C)^\alpha}.\end{aligned}$$

Therefore,

$$k\left(1-\frac{\nu^{\alpha}}{(\nu+C)^{\alpha}}\right)=1.$$

Solving for k we have

$$k=\frac{(\nu+C)^{\alpha}}{(\nu+C)^{\alpha}-\nu^{\alpha}}.$$

The size biased distribution would be

$$\begin{aligned} g_w &\propto k(1-e^{-Cx_l})f(x_l \mid \mu_x,\ \nu) \\ &= \frac{(\nu+C)^{\alpha}}{(\nu+C)^{\alpha}-\nu^{\alpha}}(1-e^{-Cx_l})\frac{\nu^{\alpha}}{\Gamma(\alpha)}{x_l}^{\alpha-1}\exp(-\nu x_l). \end{aligned}$$

6.4 Harmonic Mean

The nonparametric alternative to the size biased distribution is the harmonic mean method, which can be used to estimate averages. Dixon and Chapman (1980) for example used the harmonic mean to calculate centres and areas of animal activity, which was found to have certain advantages such as being insensitive to grid size over previous methods (such as arithmetic mean) of calculating home ranges.

Let $x=\{x_1,\ x_2,...,x_m\}$ be a random sample of speed observations of sample size m. The harmonic mean speed can be expressed as

$$h_{\bar{x}}=\frac{m}{\sum_{l=1}^{m}\left(\frac{1}{x_l}\right)}.$$

(see Lam et al., 1985). We give an example of the harmonic mean of a data set. Suppose we have $m=6$ animals with speeds $X=\{0.126, 1.562, 0.36, 2.465, 0.228, 1.036\}$. The harmonic mean speed is computed as

$$h_{\bar{x}}=\frac{m}{\sum_{l=1}^{m}\frac{1}{X_i}}=\frac{6}{(17.11139)}=0.3506437.$$

Table 6.4.1 gives the sample sizes, the harmonic mean speed, the standard mean speed and the coefficient of variation, expressed as a percentage ($C_v(\%)$) of the 9 terrestrial mammals at BCI, Panama. In parentheses are an approximate standard error of the mean speed of movement computed from the formula $\sigma_v = \sigma/\sqrt{m}$. The coefficient of variation of estimated harmonic mean speed is $C_v(\bar{v}) = \sigma_v/\bar{v}_h$ and standard (arithmetic) mean speed $C_v(\bar{v}) = \sigma_v/\bar{v}$. The C_v for most species is less than 10% and is generally smaller for the estimated harmonic mean. For smaller species such as squirrel and mouse the C_v is 10% or more despite there are 40 recorded observations.

Table 6.4.1: Sample size m, harmonic mean speed (in ms^{-1}) and standard mean speed (in ms^{-1}), and the coefficient of variation, expressed as a percentage ($C_v(\%)$), for nine Panamanian forest mammal species (approximate standard errors in pararentheses).

Species	sample size	**Averages from real data** Harmonic mean speed	Standard mean speed	Coefficient of Variation ($C_v(\%)$)	
	m	$\bar{v}_h$	$\bar{v}$	$C_v(\bar{v}_h)$	$C_v(\bar{v})$
ocelot	93	0.27 (0.03)	0.40 (0.04)	10%	10%
coati	125	0.15 (0.01)	0.32 (0.03)	6%	9%
rat	132	0.11 (0.01)	0.22 (0.02)	9%	9%
peccary	265	0.15 (0.01)	0.30 (0.02)	6%	6%
brocket	181	0.15 (0.01)	0.27 (0.02)	6%	7%
paca	195	0.17 (0.01)	0.26 (0.02)	6%	8%
agouti	953	0.14 (0.004)	0.25 (0.01)	3%	4%
squirrel	66	0.12 (0.02)	0.25 (0.03)	17%	12%
mouse	43	0.09 (0.01)	0.17 (0.03)	10%	18%

6.5 Simulation study

In this section we test the size biased sampling method and the approximation of the probability of encounter via simulations for the five following cases:

(i) Firstly, we explore the approximation of the probability for size biased sampling

where encounters are simulated from a Poisson REM. We also compare the performance of the size biased models with the standard models and the nonparameteric method (harmonic mean) used for size biased samples (Section 6.5.1).

(ii) Secondly, we explore the approximation of the probability for size biased sampling where encounters are simulated from a NB REM (Section 6.5.2).

(iii) Thirdly, we explore the approximation of the probability for size biased sampling where encounters are simulated from a ZIP REM (Section 6.5.3).

(iv) Fourthly, we explore the approximation of the probability for size biased sampling where encounters are simulated from a ZINB REM (Section 6.5.4).

For all four cases we fit a gamma model, a lognormal model or a Weibull model to animal speed data simulated from the relevant fitted models.

(v) Finally, we explore the size biased sampling method by simulating data from a Poisson REM with the true probability of encounter. We fit a gamma model to animal speed data simulated from the relevant fitted model (Section 6.5.5).

For the simulations, we test the approximation of the probability of encounter and the true probability of encounter for parameters and sample sizes that are plausible for ecological studies. For 100 simulation runs, the average parameter values (standard error in parentheses), Standard deviation (Sd) and Root Mean Square Error (RMSE) are given. The parameters required for estimation are $r = 1.25$ (m), $\theta = 0.27$ (radians) and $t = 1$ (day). The simulation procedure is outlined below.

Step 1: First we start with M animals, with speeds $x = \{x_1, ..., x_M\}$. The speeds are random numbers generated from: a gamma distribution $X \sim \text{Ga}(\alpha, \nu)$, a lognormal distribution $X \sim \ln N(\mu, \nu)$ or a Weibull distribution $X \sim \text{Wei}(\beta, \nu)$. Then we calculate the mean encounter $\Lambda = \{\lambda_1, \lambda_2, ..., \lambda_M\}$ where $\lambda_l = \{(2 + \theta)/\pi\} \times rtx_l$ for an individual with speed x_l where $l = 1, 2, ..., M$. The symbol t is the time, e.g, 1 day or 5 days; r (m) and θ (radians) are the detection zone dimensions, which are held fixed. As a demonstration we use a gamma distribution to simulate animal speed, which is shown below.

Example

This example consists of a population size of $M = 5$ with speeds following a gamma distribution with parameters $\alpha = 1.35$; $\nu = 3$. The parameters required to estimate Λ are $t = 1$; $r = 1.25$ (m) and $\theta = 0.27$. The R-code and R-output for this example are:

```
###generating random speed from a gamma distribution
   M <- 5; alpha <- 1.35; nu <- 3
   X <- rgamma(M,alpha, nu)
   0.0345 0.1460 0.1505 0.0611 0.1655


###computing expected encounter rate, Lambda
   r       <- 1.25; theta <- 0.27; D <- 1; t <- 1
   Lambda  <- ((2+ theta)/ pi)rtDX
   0.0311 0.1319 0.1359 0.0552 0.1495
```

Step 2: Next, for the M animals with speeds $x = \{x_1, ..., x_M\}$ generated in *Step* 1, we generate the number of encounters $a = \{a_1, a_2, ..., a_M\}$ from a Poisson REM $\mathbf{a} \sim$ Poi(Λ), a NB REM $\mathbf{a} \sim$ NB(Λ, κ), a ZIP REM $\mathbf{a} \sim$ ZIP(Λ, ω) or a ZINB REM $\mathbf{a} \sim$ ZINB(Λ, κ, ω). The R-code and R-output for this example are given below. In this example the first animal with speed $x_1 = 0.0345$ was seen twice, the second animal with speed $x_2 = 0.1460$ was not seen and so on.

Example continued

```
###generating the number of encounters from a Poisson model
   a  <- rpois(M, Lambda)
   2 0 2 1 2
```

Step 3: Using the number of encounters generated in *Step* 2 we then generate the actual individual speeds recorded by the camera traps, $X^* = \{x^*_1, x^*_2, ..., x^*_{M^*}\}$ where M^* is the sample size of the speeds actually recorded by the camera traps. The R-code and R-output are shown below. The example shows that the actual sample size of animals is 7 since $\sum_{i=1}^{5} a_i = 7$.

Example continued

```
###generating actual speed data observed
   X* <-  rep(X, a)
   0.0345 0.03449 0.1460 0.1505 0.0611 0.1655 0.1655
```

Using the actual speeds, X^* generated in *Step* 3 we fit a size biased gamma model and estimate the mean speed of movement, $\hat{\mu}_x$. We also fit the standard model to X^* and estimate the mean speed. We estimate mean speed, $\hat{\mu}_x$ from the average of r simulation runs (for example 100 simulation runs), the standard deviations (Sd), and Root Mean Square Errors (RMSE). From the actual data X^*, we compute the harmonic mean and the standard mean, and an estimate of the average means from r simulation runs.

6.5.1 Simulation results for encounters drawn from a Poisson REM

Table 6.5.1 gives the true values used in the simulation process. We test the model for different population sizes, M, under different parameter settings.

Table 6.5.1: True values, camera trap time t and the number of individuals, M

Camera trap time	Population size	True values and variance					
t	M	μ_x	Var(x_l)	μ_x	Var(x_l)	μ_x	Var(x_l)
1	40	0.15	0.01	0.30	0.11	0.45	0.15
1	160	0.15	0.01	0.30	0.11	0.45	0.15
1	280	0.15	0.01	0.30	0.11	0.45	0.15
5	40	0.15	0.01	0.30	0.11	0.45	0.15
5	160	0.15	0.01	0.30	0.11	0.45	0.15
5	280	0.15	0.01	0.30	0.11	0.45	0.15

For each simulation run the actual sample size, M^*, of animal with speeds $x = \{x_1, ..., x_M\}$ varies. Therefore, we give an example for the first 2 simulation runs for the different parameter settings, and the number of individuals in Table 6.5.2. As expected, the actual sample size of animal speeds, M^* recorded by the camera traps increases as the number of individuals simulated increases. Also, M^* increases for larger values of μ_x and t.

Table 6.5.2: Actual speed data sample sizes, M^*, for the first 2 simulation runs. We set $\mu_x = 0.15$, or $\mu_x = 0.30$, or $\mu_x = 0.45$, and the speed data population sizes we set $M = 40$, $M = 160$ and $M = 280$, for $t = 1$ or $t = 5$ camera trapping days.

	$M = 40$			$M = 160$			$M = 280$	
gamma	lognormal	Weibull	gamma	lognormal	Weibull	gamma	lognormal	Weibull
$t = 1$								
$\mu_x = 0.15$								
7	9	7	20	14	19	33	30	54
9	8	7	25	27	23	45	39	38
$\mu_x = 0.30$								
13	18	8	38	35	33	72	96	79
12	12	10	46	41	42	88	67	63
$\mu_x = 0.45$								
10	17	19	57	53	61	123	99	124
19	18	18	67	55	73	118	108	111
$t = 5$								
$\mu_x = 0.15$								
29	31	28	107	97	105	193	177	220
34	26	30	112	116	114	192	194	176
$\mu_x = 0.30$								
52	76	56	217	212	210	381	402	394
58	61	41	189	235	209	421	366	325
$\mu = 0.45$								
65	87	91	293	334	322	611	558	603
87	88	86	334	327	316	599	553	555

6.5.1.1 Results where size biased models perform better than standard models

A comparison of the average estimated speed, μ_x, from fitting a size biased gamma model and a gamma model to actual sample sizes M^* is given in Table 6.5.3. The Standard deviation (Sd) and Root Mean Square Error (RMSE) are also given. As expected, the size biased gamma model performs better than a gamma model in estimating mean animal speed. In this case ν is small (variance is large), which makes the difference between the two means, $1/\nu$, large. There is also a slight increase in RMSE as μ_x increases. Increasing μ_x increases λ, and therefore, the approximation gets slightly worse. But this slight bias is minimal for larger population sizes. Table 6.5.4 compares

estimations of average expected speed from a size biased lognormal model with estimations from a lognormal model. A size biased lognormal model is a lognormal model, and if ν^2 is large relative to ϵ, then the difference in estimations between the means will be large. The results show that the size biased lognormal model performs better than the lognormal model. Also, as μ_x increases RMSE increases, but this increase in the bias is minimal as population size increases. Note that as μ_x increases, λ also increases, which will cause the approximation to get slightly worse. The results from fitting a size biased Weibull model and a Weibull model to M^* are given in Table 6.5.5. A size biased Weibull model is not a Weibull model, and therefore, as expected a size biased Weibull performs better than a Weibull model. There is also a slight increase in RMSE as μ_x increases but this slight bias is minimal for larger population sizes.

Table 6.5.3: Estimates of average expected animal speed μ_x (in ms^{-1}) from a size biased gamma distribution compared with estimates from a gamma distribution.

Population size	True value		size biased gamma			gamma		
M	μ_x	ν	$\hat{\mu}_x$	Sd	RMSE	$\hat{\mu}_x$	Sd	RMSE
$t=1$								
40	0.15	15	0.16 (0.05)	0.05	0.05	0.22 (0.04)	0.05	0.08
160	0.15	15	0.16 (0.03)	0.03	0.03	0.22 (0.03)	0.03	0.08
280	0.15	15	0.15 (0.02)	0.02	0.02	0.21 (0.02)	0.02	0.07
40	0.30	2.73	0.36 (0.17)	0.14	0.16	0.66 (0.12)	0.20	0.41
160	0.30	2.73	0.30 (0.10)	0.09	0.09	0.67 (0.08)	0.11	0.39
280	0.30	2.73	0.32 (0.06)	0.07	0.02	0.66 (0.05)	0.08	0.37
40	0.45	3	0.48 (0.14)	0.15	0.16	0.78 (0.12)	0.21	0.39
160	0.45	3	0.45 (0.07)	0.07	0.07	0.78 (0.06)	0.10	0.34
280	0.45	3	0.45 (0.05)	0.06	0.06	0.78 (0.05)	0.08	0.34
$t=5$								
40	0.15	15	0.15 (0.02)	0.03	0.03	0.22 (0.02)	0.04	0.08
160	0.15	15	0.15 (0.01)	0.01	0.01	0.22 (0.01)	0.01	0.07
280	0.15	15	0.15 (0.01)	0.01	0.01	0.22 (0.01)	0.01	0.07
40	0.30	2.73	0.32 (0.01)	0.10	0.11	0.66 (0.06)	0.18	0.40
160	0.30	2.73	0.30 (0.10)	0.09	0.09	0.67 (0.08)	0.11	0.39
280	0.30	2.73	0.31 (0.06)	0.06	0.07	0.66 (0.05)	0.08	0.37
40	0.45	3	0.46 (0.06)	0.09	0.09	0.77 (0.05)	0.05	0.36
160	0.45	3	0.45 (0.03)	0.06	0.05	0.78 (0.03)	0.08	0.34
280	0.45	3	0.45 (0.02)	0.04	0.04	0.79 (0.02)	0.06	0.34

Table 6.5.4: Estimates of average expected animal speed μ_x (in ms^{-1}) from a size biased lognormal distribution compared with estimates from a lognormal distribution.

Population size	True value		size biased lognormal			lognormal		
M	μ_x	ν	$\hat{\mu}_x$	Sd	RMSE	$\hat{\mu}_x$	Sd	RMSE
$t = 1$								
40	0.15	15	0.15 (0.01)	0.01	0.01	0.16 (0.01)	0.01	0.01
160	0.15	15	0.15 (0.01)	0.01	0.01	0.16 (0.01)	0.01	0.01
280	0.15	15	0.15 (0.01)	0.01	0.01	0.16 (0.01)	0.01	0.01
40	0.30	2.73	0.30 (0.03)	0.04	0.04	0.34 (0.03)	0.04	0.06
160	0.30	2.73	0.30 (0.02)	0.02	0.02	0.34 (0.02)	0.02	0.04
280	0.30	2.73	0.30 (0.01)	0.01	0.04	0.33 (0.01)	0.02	0.04
40	0.45	3	0.45 (0.04)	0.05	0.05	0.52 (0.05)	0.06	0.08
160	0.45	3	0.45 (0.02)	0.03	0.03	0.52 (0.03)	0.03	0.08
280	0.45	3	0.45 (0.02)	0.02	0.02	0.52 (0.02)	0.03	0.08
$t = 5$								
40	0.15	15	0.15 (0.01)	0.01	0.01	0.16 (0.01)	0.01	0.01
160	0.15	15	0.15 (0.003)	0.004	0.004	0.16 (0.003)	0.005	0.01
280	0.15	15	0.15 (0.002)	0.003	0.003	0.16 (0.003)	0.004	0.01
40	0.30	2.73	0.30 (0.01)	0.02	0.02	0.33 (0.01)	0.03	0.04
160	0.30	2.73	0.30 (0.01)	0.01	0.01	0.33 (0.01)	0.01	0.03
280	0.30	2.73	0.30 (0.01)	0.01	0.01	0.33 (0.01)	0.01	0.03
40	0.45	3	0.45 (0.02)	0.03	0.03	0.52 (0.02)	0.05	0.08
160	0.45	3	0.45 (0.01)	0.02	0.02	0.52 (0.01)	0.03	0.08
280	0.45	3	0.45 (0.01)	0.01	0.01	0.52 (0.01)	0.02	0.08

Table 6.5.5: Estimates of average expected animal speed μ_x (in ms^{-1}) from a size biased Weibull distribution compared with estimates from a Weibull distribution.

Population size	True value		size biased Weibull			Weibull		
M	μ_x	ν	$\hat{\mu}_x$	Sd	RMSE	$\hat{\mu}_x$	Sd	RMSE
$t = 1$								
40	0.15	15	0.16 (0.04)	0.04	0.04	0.21 (0.03)	0.04	0.07
160	0.15	15	0.15 (0.03)	0.02	0.03	0.21 (0.02)	0.02	0.07
280	0.15	15	0.15 (0.02)	0.02	0.02	0.21 (0.02)	0.02	0.07
40	0.30	2.73	0.33 (0.08)	0.10	0.10	0.68 (0.08)	0.13	0.40
160	0.30	2.73	0.32 (0.07)	0.08	0.09	0.67 (0.07)	0.12	0.39
280	0.30	2.73	0.31 (0.05)	0.05	0.05	0.66 (0.05)	0.08	0.37
40	0.45	3	0.49 (0.11)	0.14	0.14	0.75 (0.10)	0.17	0.35
160	0.45	3	0.46 (0.06)	0.07	0.08	0.78 (0.06)	0.10	0.35
280	0.45	3	0.46 (0.04)	0.05	0.05	0.77 (0.04)	0.06	0.33
$t = 5$								
40	0.15	15	0.16 (0.02)	0.03	0.03	0.22 (0.02)	0.03	0.07
160	0.15	15	0.15 (0.01)	0.01	0.01	0.21 (0.01)	0.01	0.06
280	0.15	15	0.15 (0.01)	0.01	0.01	0.21 (0.01)	0.01	0.06
40	0.30	2.73	0.34 (0.06)	0.09	0.10	0.67 (0.06)	0.18	0.41
160	0.30	2.73	0.31 (0.03)	0.03	0.05	0.67 (0.03)	0.08	0.38
280	0.30	2.73	0.30 (0.02)	0.03	0.03	0.67 (0.02)	0.07	0.37
40	0.45	3	0.46 (0.05)	0.09	0.09	0.76 (0.05)	0.05	0.33
160	0.45	3	0.46 (0.03)	0.05	0.05	0.78 (0.03)	0.07	0.34
280	0.45	3	0.46 (0.02)	0.02	0.03	0.77 (0.02)	0.05	0.33

6.5.1.2 Results where estimations from size biased models and standard models are approximately equal

As shown in Section 6.3, a size biased gamma model and a size biased lognormal model are a gamma model and a lognormal model, respectively. And, under certain conditions estimations from size biased models are approximately equal to estimations from the standard models. In this section we test the approximation of the probability of encounter where estimations from size biased models are approximately equal to estimations from standard models. Table 6.5.6 compares estimations from a size biased gamma model and a gamma model. If the shape parameter, ν is large enough the two means are approximately equal. For a higher mean speed the difference in estimates

from a size biased gamma model and a gamma model is much smaller compared with a lower expected speed.

Table 6.5.7 gives estimated mean speeds from a size biased Weibull model and a Weibull model. Whilst a size biased Weibull model is not a Weibull model there are similarities in estimations under certain conditions. To compare estimations from the three speed data models (gamma, lognormal, Weibull), the same variance is used in the simulation process. For small variances estimations from a size biased Weibull model are similar to estimations from a Weibull model. Figures 6.5.1 and 6.5.2 shows the probability density for fitted size biased Weibull models and fitted Weibull models.

Table 6.5.8 shows the results from a size biased lognormal model and lognormal model where estimated mean speed are approximately equal. If ν^2 is small relative to ϵ then the means from the two models would be approximately equal (Table 6.5.8).

Table 6.5.6: Estimates of average expected animal speed μ_x (in ms^{-1}) from a size biased gamma distribution compared with estimates from a gamma distribution for $t = 1$ day.

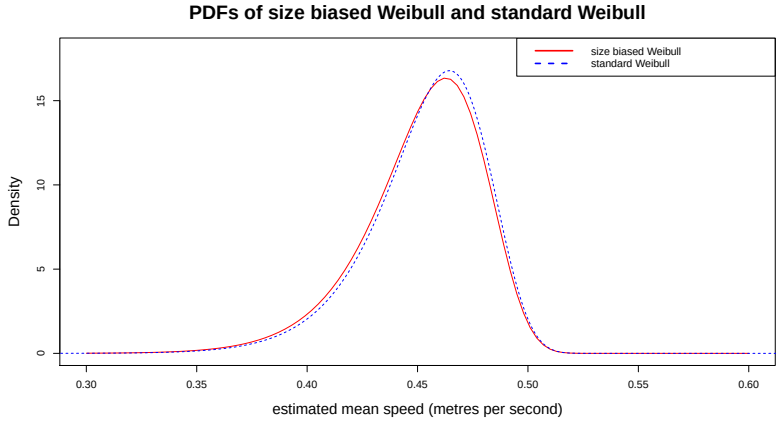
Population size	True value			size biased gamma			gamma		
M	μ_x	ν	α	$\hat{\mu}_x$	Sd	RMSE	$\hat{\mu}_x$	Sd	RMSE
40	0.45	450	202.5	0.449 (0.01)	0.01	0.01	0.451 (0.01)	0.01	0.01
160	0.45	450	202.5	0.450 (0.004)	0.005	0.005	0.452 (0.004)	0.005	0.005
280	0.45	450	202.5	0.450 (0.003)	0.004	0.004	0.452 (0.003)	0.004	0.004
40	0.15	150	22.5	0.149 (0.01)	0.01	0.01	0.155 (0.01)	0.01	0.01
160	0.15	150	22.5	0.151 (0.01)	0.01	0.01	0.158 (0.01)	0.01	0.01
280	0.15	150	22.5	0.149 (0.005)	0.006	0.01	0.156 (0.005)	0.006	0.01

Table 6.5.7: Estimates of average expected animal speed μ_x (in ms^{-1}) from a size biased Weibull distribution compared with estimates from a Weibull distribution for $t = 1$ day.

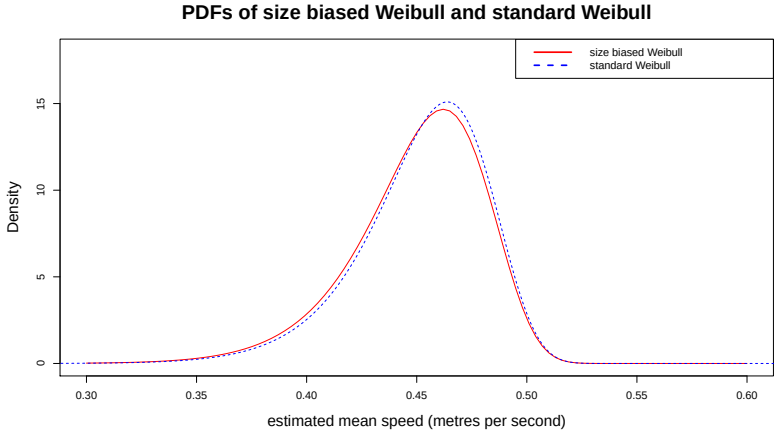
Population size	True value			size biased Weibull			Weibull		
M	μ_x	ν	β	$\hat{\mu}_x$	Sd	RMSE	$\hat{\mu}_x$	Sd	RMSE
40	0.45	17.88	0.464	0.451 (0.01)	0.01	0.01	0.453 (0.01)	0.01	0.01
160	0.45	17.88	0.464	0.450 (0.004)	0.005	0.005	0.452 (0.004)	0.005	0.005
280	0.45	17.88	0.464	0.450 (0.003)	0.003	0.003	0.452 (0.003)	0.003	0.004
40	0.15	5.42	0.163	0.151 (0.01)	0.01	0.01	0.157 (0.01)	0.01	0.01
160	0.15	5.42	0.163	0.150 (0.01)	0.01	0.01	0.157 (0.01)	0.01	0.01
280	0.15	5.42	0.163	0.151 (0.01)	0.01	0.01	0.157 (0.005)	0.006	0.01

Table 6.5.8: Estimates of average expected animal speed μ_x (in ms^{-1}) from a size biased lognormal distribution compared with estimates from a lognormal distribution for a small ν relative to ϵ for $t = 1$ day.

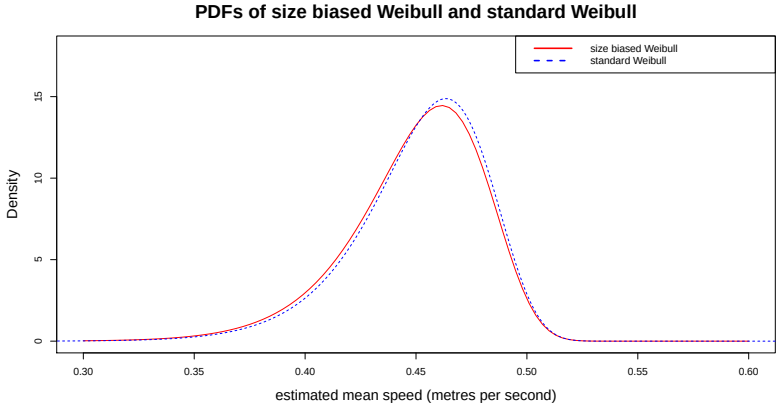
Population size	True value			size biased lognormal			lognormal		
M	μ_x	ν	ϵ	$\hat{\mu}_x$	Sd	RMSE	$\hat{\mu}_x$	Sd	RMSE
40	0.45	0.005	-0.799	0.450 (0.01)	0.01	0.001	0.450 (0.01)	0.01	0.001
160	0.45	0.005	-0.799	0.450 (0.0001)	0.0002	0.0002	0.450 (0.0001)	0.0002	0.0002
280	0.45	0.005	-0.799	0.450 (0.0001)	0.0001	0.0001	0.450 (0.0001)	0.0001	0.0001
40	0.15	0.0025	-1.897	0.150 (0.0002)	0.0003	0.0003	0.150 (0.0002)	0.0003	0.0003
160	0.15	0.0025	-1.897	0.150 (0.0001)	0.0001	0.0001	0.150 (0.0001)	0.0001	0.0001
280	0.15	0.0025	-1.897	0.150 (0.00004)	0.00005	0.00005	0.150 (0.00004)	0.00005	0.00005



(a) m=40; $\mu_x = 0.45$; $\nu = 17.88$

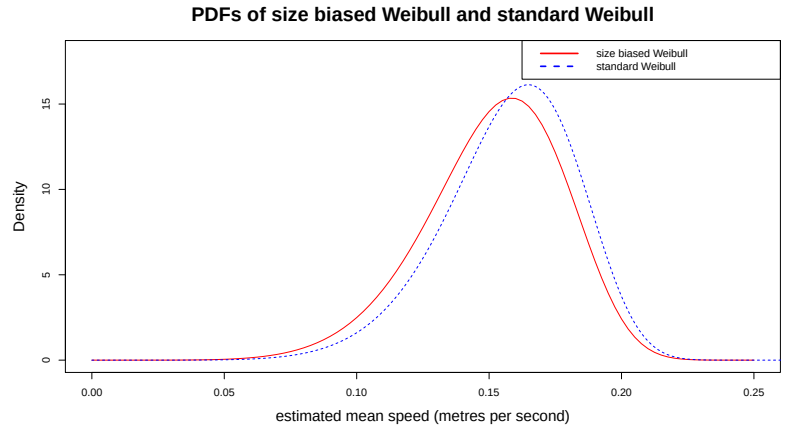


(b) m=160; $\mu_x = 0.45$; $\nu = 17.88$

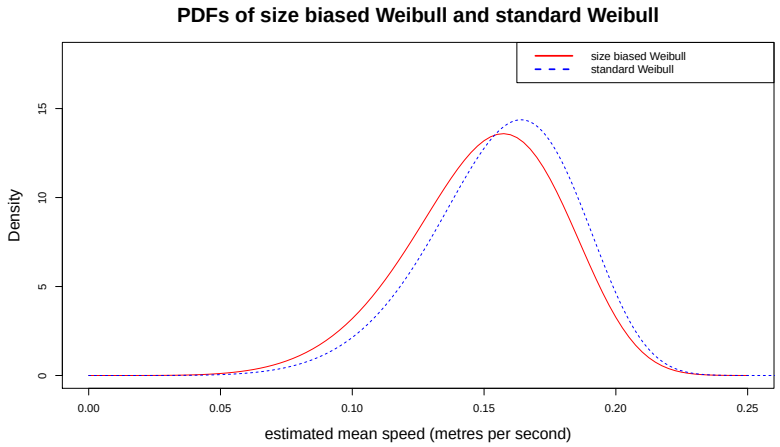


(c) m=280; $\mu_x = 0.45$; $\nu = 17.88$

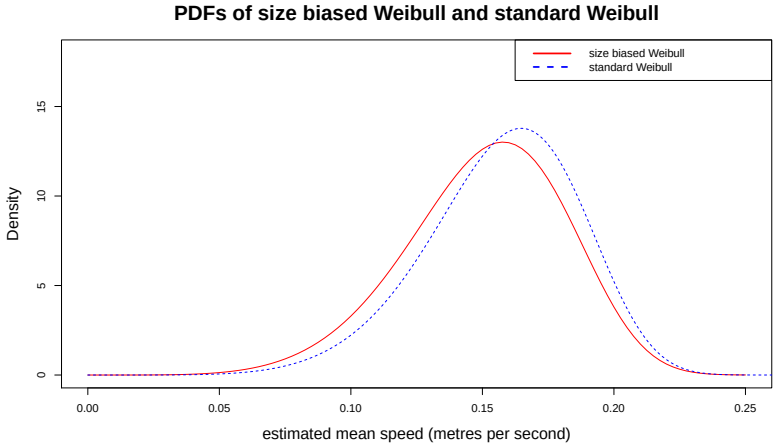
Figure 6.5.1: Probability density of the fitted size biased Weibull model (red) and fitted Weibull model (blue).



(a) m=40; $\mu_x = 0.15;\ \nu = 5.42$



(b) m=160; $\mu_x = 0.15;\ \nu = 5.42$



(c) m=280; $\mu_x = 0.15;\ \nu = 5.42$

Figure 6.5.2: Probability density of the fitted size biased Weibull model (red) and fitted Weibull model (blue).

6.5.1.3 *Comparing estimations from parametric models with estimations from non-parametric methods*

In this section we simulate encounters from a Poisson REM to test the approximation method and compare estimations from the size biased likelihood model, the standard likelihood model, the harmonic mean method and the standard (arithmetic) mean method. For 100 simulation runs, we compute the average parameter estimates (average standard errors in parentheses), Standard deviation (Sd) and Root Mean Square Error (RMSE). For illustration, we fit a lognormal model to the speed data and the results are given in Table 6.5.9. The simulation results (Table 6.5.9) reveal that a size biased lognormal model and the nonparameteric harmonic mean method and a lognormal model can provide similar estimates when the difference between ϵ and ν^2 is small. But when this difference increases it would be more appropriate to use a size biased lognormal model since it would provide more accurate estimates of the average speed.

We also investigate the performance of the gamma models (Table D.1.1) and Weibull models (Table D.1.2), with the same mean and variance, and the results are given in appendix D.1. For the chosen parameters the simulation results (Table D.1.1) show that the size biased gamma model does better than the standard gamma model and the harmonic mean method under all scenarios. The bias from the gamma model and the harmonic mean method is much larger when ν is small and, hence a bigger difference between the means of the two models, that is, $1/\nu$ is large. Particularly, for lower μ_x, the harmonic mean method performs poorly. As expected, the size biased Weibull model (Table D.1.2) performs better than the Weibull model, and the nonparametric method under all scenarios. For lower expected speeds, the bias from the harmonic mean method is much larger compared with higher values of expected speeds. Based on these results, if the data follows a lognormal model, we would recommend fitting size biased models when the sampling variability in the data is high (the difference between ϵ and ν^2 is large). However, if the sampling variability is low the harmonic mean method can be used, which is a simpler method and easier to implement. But, if the data follows a gamma model or Weibull model, and under these scenarios, we recommend fitting size biased models, which would provide more accurate estimates of the average speed.

Table 6.5.9: Estimates of average expected animal speed μ_x (ms^{-1}) from a size biased lognormal distribution compared with estimates from a lognormal distribution, and nonparametric methods. The standard deviation (Sd) and Root Mean Square Error (RMSES) are also given.

				Parametric Methods						**Nonparametric Methods**					
Population size	True value			size biased lognormal			lognormal			Harmonic			Standard		
M	μ_x	ν	ϵ	$\hat{\mu}_x$	Sd	RMSE	$\hat{\mu}_x$	Sd	RMSE	$\hat{\mu}_x$	Sd	RMSE	$\hat{\mu}_x$	Sd	RMSE
$t = 1$															
40	0.45	0.10	-0.80	0.449 (0.01)	0.01	0.01	0.453 (0.01)	0.01	0.01	0.446	0.006	0.008	0.450	0.006	0.006
160	0.45	0.10	-0.80	0.448 (0.006)	0.007	0.007	0.453 (0.006)	0.007	0.008	0.445	0.003	0.006	0.449	0.003	0.003
280	0.45	0.10	-0.80	0.451 (0.004)	0.005	0.005	0.455 (0.004)	0.004	0.007	0.445	0.003	0.006	0.450	0.003	0.003
40	0.45	0.39	-0.87	0.45 (0.04)	0.05	0.05	0.52 (0.05)	0.07	0.09	0.39	0.02	0.07	0.45	0.06	0.03
160	0.45	0.39	-0.87	0.45 (0.02)	0.03	0.03	0.52 (0.03)	0.03	0.09	0.39	0.01	0.06	0.45	0.01	0.01
280	0.45	0.39	-0.87	0.45 (0.02)	0.02	0.02	0.52 (0.02)	0.02	0.08	0.39	0.009	0.06	0.45	0.01	0.01
40	0.15	0.10	-1.90	0.149 (0.005)	0.006	0.006	0.151 (0.005)	0.006	0.006	0.149	0.002	0.003	0.150	0.002	0.002
160	0.15	0.10	-1.90	0.151 (0.003)	0.004	0.004	0.151 (0.003)	0.004	0.005	0.149	0.001	0.002	0.150	0.001	0.001
280	0.15	0.10	-1.90	0.150 (0.002)	0.002	0.0002	0.151 (0.002)	0.002	0.001	0.149	0.001	0.002	0.150	0.001	0.001
40	0.15	0.24	-1.93	0.153 (0.01)	0.01	0.01	0.162 (0.01)	0.02	0.02	0.141	0.006	0.01	0.149	0.006	0.006
160	0.15	0.24	-1.93	0.150 (0.008)	0.009	0.009	0.160 (0.009)	0.01	0.01	0.141	0.003	0.009	0.150	0.003	0.003
40	0.15	0.24	-1.93	0.150 (0.006)	0.009	0.007	0.160 (0.006)	0.006	0.01	0.142	0.002	0.009	0.150	0.002	0.002

6.5.2 *Simulation results for encounters drawn from a negative binomial REM*

In this section we test the approximation of the probability of encounter from a NB REM via simulations. For the simulations we set the camera trap time, $t = 1$ or $t = 5$; expected speed, $\mu_x = 0.15$ (ms^{-1}) with variance $\text{Var}(x_l) = 0.01$ or $\mu = 0.45$ (ms^{-1}) with variance $\text{Var}(x_l) = 0.15$; population size, $M = 40$ or $M = 280$; and dispersion parameter, $\kappa = 0.56$.

Table 6.5.10 compares estimations from a size biased gamma with estimations from a gamma model. Overall, stable estimations are obtained, and as expected the size biased model provides better estimations than the standard model. For this scenario the shape parameter, ν, is small resulting in a large difference in estimated mean between a size biased gamma model and a gamma model. Also, as μ_x increases the RMSEs increase slightly, but this slight increase in minimal for larger population sizes. The expected encounter rate, λ, depends on μ_x, therefore, as μ_x increases λ increases, which causes the size biased approximation to perform slightly worse.

We also provide estimations from a size biased lognormal model and a size biased Weibull model, which can be found in Table D.2.2, and Table D.2.3, respectively in appendix D.2. Accurate estimations are obtained from a size biased lognormal distribution and a lognormal distribution for all parameter values and for all population sizes, M. In this case the variance from the normal distribution is small relative to the mean of the logarithm, which results in estimations from the two models being approximately equal. Also, the size biased approximation works well for small values of μ, but as μ increases, there is a slight increase in RMSE but this slight increase in the bias is minimal for larger population sizes (see Table D.2.2 in appendix D.2).

A size biased Weibull distribution also performs better than the Weibull distribution. For a small μ_x $(= 0.15)$ (ms^{-1}) accurate estimations are obtained for $M = 160$ or larger. The bias is also smaller for larger population sizes. However, as μ_x increases, the bias increases slightly but this increase is minimal for larger population sizes. Note that the expected encounter rate, λ, increases for larger values of expected speed, μ_x, causing

the approximation of the probability for size biased to be slightly worse (see Table D.2.3 in appendix D.2).

For each simulation run, the actual sample size of animal speed, M^*, generated varies, so as an example we give the results for the first 3 simulation runs in Table D.2.1 in appendix D.2. The results show M^* increases for larger values of t and μ_x since λ_l is dependent on both t and μ_x, as expected. Also, as population size, M, increases M^* increases.

Table 6.5.10: Estimates of expected animal speed μ_x (in ms^{-1}) from a size biased gamma distribution compared with estimates from a gamma distribution.

Population size	True value		size biased gamma			gamma		
M	μ_x	ν	$\hat{\mu}_x$	Sd	RMSE	$\hat{\mu}_x$	Sd	RMSE
$t=1$								
40	0.15	15	0.17 (0.05)	0.06	0.06	0.22 (0.04)	0.06	0.09
160	0.15	15	0.15 (0.03)	0.03	0.03	0.22 (0.03)	0.04	0.08
280	0.15	15	0.15 (0.02)	0.02	0.02	0.21 (0.02)	0.03	0.07
40	0.45	3	0.45 (0.16)	0.15	0.15	0.74 (0.11)	0.19	0.35
160	0.45	3	0.45 (0.07)	0.10	0.10	0.78 (0.06)	0.13	0.36
280	0.45	3	0.46 (0.04)	0.04	0.05	0.77 (0.04)	0.06	0.33
$t=5$								
40	0.15	15	0.15 (0.02)	0.03	0.03	0.21 (0.02)	0.04	0.07
160	0.15	15	0.15 (0.01)	0.02	0.02	0.22 (0.01)	0.03	0.07
280	0.15	15	0.15 (0.01)	0.02	0.02	0.22 (0.01)	0.02	0.07
40	0.45	3	0.48 (0.06)	0.14	0.14	0.78 (0.05)	0.21	0.39
160	0.45	3	0.45 (0.03)	0.07	0.07	0.78 (0.03)	0.12	0.39
280	0.45	3	0.45 (0.02)	0.06	0.06	0.80 (0.02)	0.10	0.36

6.5.3 Simulation results for encounters drawn from a ZIP REM

In this section we test the approximation of the probability of encounter for size biased sampling from a ZIP REM via simulations. We set a small value of $\rho = 0.10$ to test the

model, since the approximation will hold for small values of ρ (see Table 6.2.2 in Section 6.2.2). We set the camera trap time, $t = 1$ or $t = 5$; population size, $M = 40$ or $M = 280$; expected speed, $\mu_x = 0.15$ (ms^{-1}) with variance, $\text{Var}(x) = 0.01$ or $\mu_x = 0.45$ (ms^{-1}) with variance, $\text{Var}(x) = 0.15$; and probability of the zero-response category, $\rho = 0.10$.

The results from fitting a size biased gamma distribution and a standard gamma distribution to the speed data are given in Table 6.5.11. The Standard deviation (Sd) and Root Mean Square Error (RMSE) are also given. The simulation results (Table 6.5.11) show that under the scenario, small shape parameter, the size biased gamma model performs better than the gamma model, and as expected, larger sample size improves estimation and precision. But as μ_x increases the approximation of the probability of encounter for size biased sampling is slightly worse with increasing RMSEs, but this increase is minimal for larger population sizes.

The results from fitting lognormal models (Table D.3.2) and Weibull models (Table D.3.3) are shown in appendix D.3. Simulation results from the size biased logormal model and lognormal model are approximately equal (Table D.3.2) but the size biased Weibull model provides better estimates than the Weibull model (Table D.3.3). But as expected, as μ_x increases λ increases, and the approximation method performs slightly worse. The simulation results show that the lognormal models obtained smaller RMSEs compared with the RMSEs from the gamma models and Weibull models.

Table D.3.1 in appendix D.3 gives the sample sizes of the speeds (M^*) captured by the camera traps for the first 3 simulation runs. The actual sample sizes of animal speed, M^* changes for each simulation run and for larger population sizes, M. For larger population size, M^* increases. Also for larger values of μ_x and t, M^* increases.

Table 6.5.11: Estimates of expected animal speed μ_x (in ms^{-1}) from a size biased gamma distribution compared with estimates from the gamma distribution.

Population size	True value		size biased gamma			gamma		
M	μ_x	ν	$\hat{\mu}_x$	Sd	RMSE	$\hat{\mu}_x$	Sd	RMSE
$t = 1$								
40	0.15	15	0.15 (0.04)	0.05	0.05	0.21 (0.04)	0.06	0.08
160	0.15	15	0.16 (0.03)	0.03	0.03	0.22 (0.03)	0.03	0.08
280	0.15	15	0.15 (0.03)	0.03	0.03	0.21 (0.03)	0.03	0.07
40	0.45	3	0.50 (0.14)	0.18	0.17	0.78 (0.12)	0.21	0.39
160	0.45	3	0.46 (0.07)	0.09	0.09	0.78 (0.07)	0.12	0.35
280	0.45	3	0.45 (0.05)	0.07	0.07	0.78 (0.05)	0.10	0.34
$t = 5$								
40	0.15	15	0.15 (0.03)	0.03	0.03	0.21 (0.02)	0.05	0.08
160	0.15	15	0.15 (0.01)	0.01	0.02	0.22 (0.01)	0.02	0.07
280	0.15	15	0.15 (0.01)	0.01	0.01	0.22 (0.01)	0.02	0.07
40	0.45	3	0.45 (0.06)	0.11	0.11	0.78 (0.06)	0.21	0.39
160	0.45	3	0.45 (0.03)	0.06	0.06	0.78 (0.03)	0.10	0.34
280	0.45	3	0.45 (0.02)	0.05	0.05	0.78 (0.02)	0.08	0.34

6.5.4 Simulation results for encounters drawn from a ZINB REM

In this section we investigate the performance of the approximation of the probability of encounter from a ZINB REM via simulations. Population sizes and true values used in the simulation process are given in Table 6.5.12. A typical value of the dispersion parameter, κ, and a small value of ρ for the which the approximation holds are chosen.

Table 6.5.12: True values, population size m and camera trap time t used in the simulation process from a ZINB REM.

Camera trap time	Population size	True values and variance					
t	M	μ_x	Var(x)	μ_x	Var(x)	κ	ρ
1	40	0.15	0.01	0.45	0.15	0.56	0.10
1	160	0.15	0.01	0.45	0.15	0.56	0.10
1	280	0.15	0.01	0.45	0.15	0.56	0.10
5	40	0.15	0.01	0.45	0.15	0.56	0.10
5	160	0.15	0.01	0.45	0.15	0.56	0.10
5	280	0.15	0.01	0.45	0.15	0.56	0.10

Table 6.5.13 compares estimations from a size biased gamma model and a gamma model. The simulation results show that the size biased gamma model performs better than the gamma model but obtains a positive bias for small population sizes but as expected the bias reduces for large population sizes. However, as μ_x increases RMSE increases and the approximation of the probability method performs slightly worse, but the bias is minimal for larger population sizes.

Table D.4.2, in appendix D.4, gives the results from fitting a size biased lognormal model and a lognormal model to animal speed data. The results show that the approximation method works well and accurate estimations are obtained for both models with smaller bias compared with the gamma models.

Table D.4.3, in appendix D.4, compares estimations from a size biased Weibull model with estimations from a Weibull model. The size biased Weibull model performs better than the Weibull model with lower RMSEs. As expected, as t increases estimations improve.

For each simulation run the actual sample size, M^* varies. In Table D.4.1, in appendix D.4, we give the actual sample size, M^* used in the simulation process for the first 3 simulation runs. As expected, M^* increases as population size, M and the number of camera trap days, t increase.

Table 6.5.13: Estimates of expected mean speed μ_x (in ms^{-1}) from a size biased gamma distribution compared with estimates from the standard gamma distribution.

Population size	True value		size biased gamma			gamma		
M	μ_x	ν	$\hat{\mu}_x$	Sd	RMSE	$\hat{\mu}_x$	Sd	RMSE
$t = 1$								
40	0.15	15	0.16 (0.04)	0.05	0.05	0.22 (0.04)	0.05	0.08
160	0.15	15	0.15 (0.03)	0.03	0.03	0.22 (0.03)	0.04	0.08
280	0.15	15	0.15 (0.02)	0.02	0.02	0.21 (0.02)	0.02	0.07
40	0.45	3	0.49 (0.18)	0.21	0.21	0.75 (0.13)	0.30	0.42
160	0.45	3	0.46 (0.10)	0.12	0.12	0.76 (0.09)	0.15	0.35
280	0.45	3	0.45 (0.09)	0.10	0.10	0.78 (0.07)	0.13	0.35
$t = 5$								
40	0.15	15	0.15 (0.03)	0.04	0.04	0.20 (0.03)	0.05	0.07
160	0.15	15	0.15 (0.02)	0.02	0.02	0.22 (0.02)	0.03	0.07
280	0.15	15	0.15 (0.01)	0.02	0.02	0.21 (0.01)	0.02	0.07
40	0.45	3	0.48 (0.08)	0.16	0.16	0.77 (0.07)	0.28	0.42
160	0.45	3	0.45 (0.04)	0.07	0.07	0.76 (0.04)	0.13	0.34
280	0.45	3	0.45 (0.03)	0.06	0.06	0.78 (0.03)	0.11	0.34

6.5.5 Results from size biased method with true probability and the approximation

In this section we investigate the performance of the size biased method using the true probability of encounter and the approximation of the probability of encounter from a Poisson REM. As a demonstration we assume animal speed follows a gamma model. Table 6.5.14 compares estimated mean speed from using the true probability of encounter and the approximation of the probability of encounter from a Poisson REM. The average parameter estimates from the two methods show minimal differences but as expected, the true probability of encounter method does slightly better that the approximation method. For low sampling variability the bias introduced from both methods is inconsequential, and there is no obvious difference in estimates between the two methods. Increasing the sampling variability introduces a larger bias, particularly for the approximation method with low survey effort but as expected, increasing survey effort reduces the bias and improves precision. Based on these results we can conclude that the approximation method for the probability of encounter from the underlying Poisson model in REM works well to estimate average speed of movement.

Table 6.5.14: Average estimated mean animal speed $\hat{\mu}_x$ (ms^{-1}) from a size biased gamma distribution using the true probability of encounter and the approximation of the probability of encounter from a Poisson REM. We set $t = 5$ camera trap days, and test the methods under low and high sampling variability. The Standard deviation (Sd) and Root Mean Square Error (RMSE) are also given.

				Size biased gamma models					
Population size	True value			True probability of encounter method			Approximation method		
M	μ_x	ν	α	$\hat{\mu}_x$	Sd	RMSE	$\hat{\mu}_x$	Sd	RMSE
Low sampling variability									
40	0.45	45	20.25	0.450 (0.01)	0.02	0.02	0.451 (0.01)	0.02	0.02
160	0.45	45	20.25	0.449 (0.01)	0.01	0.01	0.450 (0.01)	0.01	0.01
280	0.45	45	20.25	0.450 (0.004)	0.01	0.01	0.451 (0.004)	0.01	0.01
High sampling variability									
40	0.45	1.013	2.25	0.457 (0.07)	0.11	0.12	0.478 (0.07)	0.11	0.12
160	0.45	1.013	2.25	0.454 (0.04)	0.05	0.05	0.453 (0.04)	0.06	0.06
280	0.45	1.013	2.25	0.452 (0.03)	0.04	0.04	0.452 (0.03)	0.04	0.04

6.6 Application of size biased method to real data at BCI, Panama

In this section we applied the size biased method to a community of terrestrial animals at BCI, Panama. For comparison, estimates from the nonparametric harmonic mean method and parametric models are given.

Table 6.6.1 compares the harmonic means and standard means with estimations from a size biased Weibull model and a standard Weibull model. The performance of the size biased Weibull model is better than the Weibull model as shown by the ΔAIC values. However, the size biased Weibull model is not always a lot better than the standard Weibull model, as in the case of the ocelot species. The ΔAIC values suggest that the size biased Weibull model is the best model for all species.

Table 6.6.2 compares all size biased models using the AIC method for 9 species at BCI, Panama. The results suggest that the lognormal model is the best model that fits the data for most of the species. Figure 6.6.1 shows the histograms of the speed data and fitted lognormal model (black), fitted gamma model (red) and fitted Weibull model (blue), which also confirms that the lognormal model is the model that best explains the data for most species. Also, estimates of the standard error from the size biased lognormal distribution are generally smaller than the estimates from the size biased gamma model and the size biased Weibull model.

We give the results from a size biased gamma model (Table D.5.1), and a lognormal model (Table D.5.2) in appendix D.5.

Table D.5.1, in appendix D.5, compares the harmonic mean and standard (arithmetic) sample mean with estimates from a size biased gamma distribution, and estimates from a standard gamma distribution. A size biased gamma model is a gamma model, and therefore, the likelihoods will always be the same. Therefore, model choice using AIC method would not distinguish between the two models. From the evidence in Section 6.3.2 and the simulation results in Section 6.5.1.1 above, a size biased gamma model will obtain better estimations than a standard gamma model when the shape parameter

ν is small. The results in Table D.5.1 show that the difference in estimations between the two methods is large suggesting that the shape parameter ν is small, and hence, variance is large. Also, the harmonic means are similar to estimations from the size biased parametric gamma model.

Table D.5.2, in appendix D.5, compares estimations from a size biased lognormal model with estimations from a lognormal model (standard errors are in parentheses). Similarly, a size biased lognormal model is a lognormal model and the likelihoods will always be the same, and the AIC method would not distinguish between the two models. Also, estimations of the variance, ν^2, from the normal model will always be the same. The results show that estimations between a size biased lognormal model and a standard lognormal is large suggesting that ν^2 is large relative to ϵ (Table D.5.2). Also, the harmonic mean and the estimates from the size biased model are the similar.

Table 6.6.1: Estimates (in ms^{-1}) from a size biased Weibull distibution compared with estimates from a Weibull distribution. The Standard error are in parentheses and the negative loglikelihood $(-\ell)$ is also given.

Species	**Nonparametric Method**		**Parametric Method**							
	Harmonic mean	Standard mean	size biased Weibull				Weibull			
	$\hat{\mu}_x$	$\hat{\mu}_x$	$\hat{\mu}_x$	$\hat{\nu}$	$(-\ell)$	ΔAIC	$\hat{\mu}_x$	$\hat{\nu}$	$(-\ell)$	ΔAIC
ocelot	0.27 (0.03)	0.40 (0.04)	0.26 (0.02)	1.44 (0.15)	-18.60129	0	0.40 (0.02)	1.95 (0.16)	-18.24274	0.72
coati	0.15 (0.01)	0.32 (0.03)	0.13 (0.02)	0.82 (0.07)	-2.603436	0	0.32 (0.02)	1.25 (0.08)	-22.54353	6.98
rat	0.11 (0.01)	0.22 (0.02)	0.09 (0.01)	0.81 (0.07)	-76.02399	0	0.23 (0.02)	1.23 (0.08)	-70.42447	11.20
peccary	0.15 (0.01)	0.30 (0.02)	0.12 (0.02)	0.81 (0.05)	-78.1969	0	0.30 (0.02)	1.21 (0.05)	-65.30716	25.78
brocket	0.15 (0.01)	0.27 (0.02)	0.12 (0.01)	0.92 (0.07)	-74.84183	0	0.27 (0.02)	1.34 (0.07)	-67.84225	14.00
paca	0.17 (0.01)	0.26 (0.02)	0.13 (0.01)	1.03 (0.07)	-97.85421	0	0.27 (0.01)	1.43 (0.07)	-86.91111	21.89
agouti	0.14 (0.004)	0.25 (0.01)	0.10 (0.005)	0.84 (0.03)	-438.4727	0	0.26 (0.01)	1.27 (0.03)	-401.3331	74.28
squirrel	0.12 (0.02)	0.25 (0.03)	0.08 (0.02)	0.74 (0.09)	-30.35628	0	0.25 (0.03)	1.15 (0.10)	-27.25814	6.20
mouse	0.09 (0.01)	0.17 (0.03)	0.07 (0.01)	0.87 (0.13)	-38.63665	0	0.17 (0.02)	1.28 (0.13)	-36.33189	4.61

Table 6.6.2: Parameter estimates (in ms^{-1}) and ΔAIC values for the three size biased distributions.

Species	**Sample size**	**gamma**			**lognormal**			**Weibull**		
	n	$\hat{\mu}_x$	$\hat{\nu}$	ΔAIC	$\hat{\mu}_x$	$\hat{\nu}$	ΔAIC	$\hat{\mu}_x$	$\hat{\nu}$	ΔAIC
ocelot	93	0.27 (0.02)	8.03 (0.59)	0.46	0.28 (0.02)	0.61 (0.04)	7.47	0.26 (0.02)	1.44 (0.15)	0
coati	125	0.12 (0.02)	5.08 (0.69)	2.04	0.16 (0.01)	0.86 (0.05)	1.23	0.13 (0.02)	0.82 (0.07)	0
rat	132	0.09 (0.02)	7.37 (0.97)	12.57	0.11 (0.01)	0.82 (0.05)	0	0.09 (0.01)	0.81 (0.07)	8.79
peccary	265	0.12 (0.02)	5.59 (0.52)	23.04	0.15 (0.01)	0.82 (0.04)	0	0.12 (0.01)	0.81 (0.05)	14.65
brocket	181	0.13 (0.02)	7.13 (0.79)	7.97	0.15 (0.01)	0.76 (0.04)	0	0.12 (0.01)	0.92 (0.07)	6.61
paca	195	0.15 (0.01)	9.18 (0.97)	24.12	0.17 (0.01)	0.65 (0.03)	0	0.13 (0.01)	1.03 (0.07)	27.61
agouti	953	0.11 (0.01)	6.84 (0.33)	65.49	0.13 (0.004)	0.81 (0.02)	0	0.10 (0.005)	0.84 (0.03)	46.69
squirrel	66	0.08 (0.03)	5.93 (1.12)	8.94	0.12 (0.01)	0.86 (0.07)	0	0.08 (0.02)	0.74 (0.09)	5.73
mouse	43	0.08 (0.02)	11.22 (2.56)	4.93	0.09 (0.01)	0.75 (0.08)	0	0.07 (0.01)	0.87 (0.13)	4.09

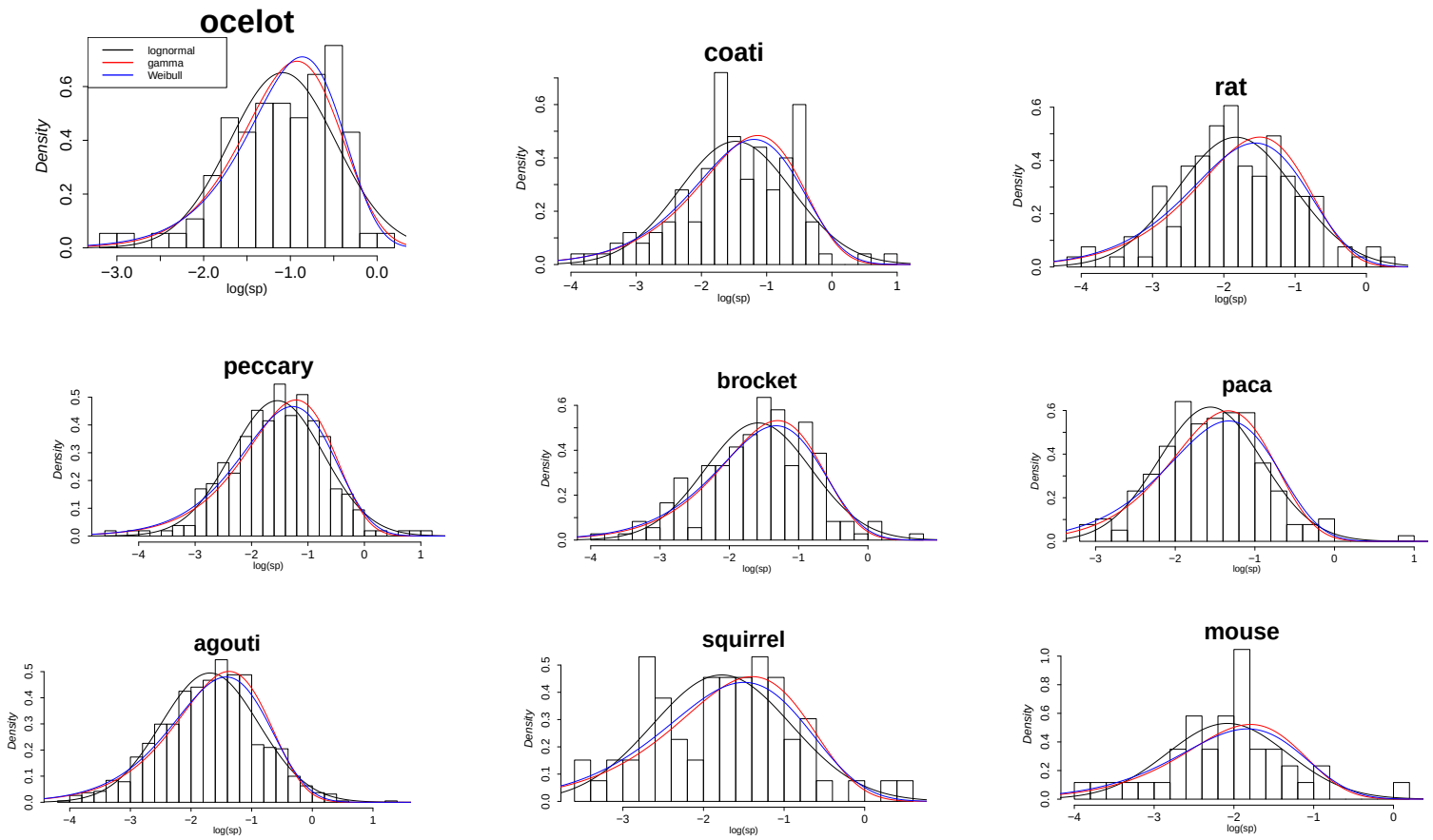


Figure 6.6.1: Histograms of log speed observations and fitted size biased distributions assuming three parametric distributions: lognormal (green), gamma(red), Weibull (blue) of nine species.

6.7 Discussion

In this Chapter we have explored the method of size biased sampling to estimate average speed of movement. A distribution is size biased if the probability of inclusion of a unit is proportional to a predetermined weight which is equal to the length of that unit (Patil and Rao, 1978). The main focus of the Chapter has been to reduce the bias in speed caused by encountering faster moving animals more frequently. Hutchinson and Waser (2007) found that faster moving animals are more likely to encounter camera traps, and as such expected animal speed would be biased towards these animals.

We started by considering the method of size biased sampling adopted by Rowcliffe et al. (2016) to estimate average speed. Rowcliffe et al. (2016) generated a sample of distances travelled and passage duration, and have computed the speed of movement by taking the ratio of distance moved to passage duration. Using the derived speeds Rowcliffe et al. (2016) fitted size biased models (gamma, lognormal, Weibull) and maximized the likelihood to estimate average speed. We have however derived the probability of encountering an animal in REM showing that it is approximately proportional its speed. We assumed four probability density functions for which the probability of encounter could be derived. These included a Poisson REM, a negative binomial model REM, a zero-inflated Poisson REM and a zero-inflated negative binomial model REM. We have illustrated how the approximation of the probability of encountering an animal holds by simulating encounters from the true probability density function (Poisson REM, NB REM, ZIP REM, ZINB REM) and maximizing the likelihood of the approximation probability density function of size biased sampling. For parameters and sample sizes that are realistic for ecological studies the simulations highlighted the value of accounting for the bias in speeds of faster moving animals, and of using simulations as a tool for design. But under certain conditions the simulations revealed that size biased models and standard models are approximately equal.

We have also considered how the probability density function for size biased sampling works for the true probability of encountering an animal. For illustration, we considered the underlying Poisson model in REM for the encounters and a gamma model for the

speed data. The simulations revealed that the performance of the approximation of the probability of encounter method and the true probability of encounter method are similar for parameters and sample sizes that are realistic in ecological studies. The simulations have also shown that deciding on the use of a parametric likelihood-based approach or a nonparmetric method for estimation of average speed from size biased sampling will depend on the distribution of the data and sampling variability. In the case where the data is lognormally distributed the nonparametric method has been found to provide accurate estimates as the parametric model when sampling variability in the data is low (or when the difference between mean of the logarithm and the variance from the normal distribution is small). But in case where the data is assumed to follow a gamma model or a Weibull model we would recommend a parametric likelihood-based approach accounting for the bias in the data.

The application of the size biased sampling method using the approximation method showed that the bias in the speed of faster moving animals can be reduced, and if this bias is not accounted for the average speed would be incorrectly estimated as shown by the large difference in estimations from the real data. While a size biased gamma model and a size biased lognormal model remain a gamma model and lognormal model, respectively, the results indicate that it is crucial to consider size biased sampling in ecological studies as shown in the example where animal speed is assumed to follow a Weibull model. The results showed that there was more support for a size biased lognormal model, which generally obtained more precise estimates compared with the gamma model and Weibull model.

Chapter 7

iREM With Size Biased Sampling

The objective of this Chapter is to improve estimation of the density by correcting for the bias in expected animal speed of movement. Size biased sampling, discussed in detail in Chapter 6, corrects for the bias in average speeds induced by faster moving animals, which are more likely to encounter the camera traps. An assumption of iREM is that the encounter data (photographs) and speed data are independent. Whilst the Whipsnade Wild Animal Park (WWAP) data are independent (see Rowcliffe et al., 2008, and Section 2.7.1) this may not be the case for the data collected at Barro Colorado Island (BCI), Panama, which is used in this Chapter, and which was discussed in detail in Chapter 5. For the BCI data the encounter records and the speed data were extracted from the same camera trap record. To ensure that the speed data and encounter data are independent it would be required to conduct two independent surveys simultaneously, where the encounter data would come from one survey and the speed data would come from the other survey. Since the encounter records and speed data from BCI, Pananma were collected from the same camera trap record, we investigate whether this assumption of independence, if violated, would affect estimated density. As animals captured by camera traps are likely to be faster moving suggest that there is some dependence, so we adopt the method of size biased sampling to model this dependence.

In this Chapter we develop an integrated Random Encounter Model that includes size biased sampling (iREM-SB) to correct for the bias in speed, and hence, estimated density. The Chapter begins with a description of iREM-SB in Section 7.1, which we test

via simulations in Section 7.2. Simulations were conducted to investigate the importance of correcting for the bias in speed of faster moving animals in density estimation, and to test the assumption of independence between data sets. An analysis of the data for the 9 Panamanian species at BCI rainforest using an iREM-D developed in Chapter 5 correcting for the bias in the speed of movement is given in Section 7.3.

7.1 The Model

This section describes the integrated likelihood that includes size biased sampling of animal speed to estimate density. The joint likelihood of the encounter data and size biased animal speed of movement data is maximized to estimate density and expected animal speed. As discussed in Chapter 5, the encounters, a_i, are assumed to follow a Poisson model, such that $a_i \sim \text{Pois}(\lambda t_i)$ where

$$\lambda = \frac{(2+\theta)}{\pi} r \mu_x D, \tag{7.1.1}$$

and t_i is the camera trap time in days for $i = 1, 2, ..., c$ camera trap. The detection zone dimensions, r and θ are held fixed; μ_x is the expected speed, and D is the density.

Suppose the encounter data a_i, has probability mass function $h(a_i \mid \lambda,\ t_i,\ \tau)$ where $i = 1, 2, .., c$ is the number of encounters on the ith camera trap, λ is the expected encounter rate, t_i is the camera trap time period (in days) for the ith camera, and τ represents any additional parameter in the model. And suppose the animal speed data, x_l, has probability density function $f(x_l \mid \mu_x,\ \nu)$ where $l = 1, 2, ..., m$ is the number of animal speed data, μ_x is expected speed and ν represents any additional parameter in the model. Assuming the following: 1) the encounters between animals and camera traps are independent, 2) the speeds of movement are independent and identically distributed, 3) animals move randomly, and 4) the speed data and encounter data are statistically independent, the log-likelihood function can be constructed as

$$\ell = \sum_{i=1}^{c} \log\, h(a_i \mid \lambda,\ t_i,\ \tau) + \sum_{l=1}^{m} \log\left(\frac{x_l f(x_l \mid \mu_x,\ \nu)}{\mu_x}\right). \tag{7.1.2}$$

As in the previous Chapters the speed data is assumed to follow a gamma model, a

lognormal model or a Weibull model and the encounters are discrete data, which are assumed to follow a Poisson REM as discussed in Chapter 5. We could also use a a negative binomial REM for the encounter data, as shown in Chapter 5 but this model is not considered here.

7.2 Simulation Study

In this section we investigate the performance of iREM-SB. The main assumption of iREM-SB is that the speed data and the encounter data are independent of each other. As the data sets are generally collected from the encounter records from camera traps, it would be required to conduct parallel surveys, one for each data set, in order to obtain independent information from camera traps. However, this would be expensive and time consuming in ecological surveys, therefore, the same data source is used to extract information and the assumption of independence is made. Also, as discussed in Section 5.3 Chapter 5, extracting the data, particularly the speed data from the camera trap records is tedious and time consuming, which involves the reconstruction of the movement path of animals using a measuring tap or hip chain. This reconstruction is done by viewing images in the field before removing the camera traps and measuring distances travelled relative to nearby landmarks such as trees and rocks (Rowcliffe et al., 2011). Therefore, in this Section the independence assumption and random selection of a sample of the speed data are motivating examples for the simulations. We explore the performance of iREM-SB, looking at the four following aspects

(i) Firstly, we investigate the independence assumption of the data in iREM-SB using all possible speed data. To do this we first simulate speed data from the alternative speed data models, and using these data we generate encounters from a Poisson REM. We fit a Poisson REM to the encounter data and a size biased model to the speed data. Next we generate a separate set of speed data from the alternative speed data models and fit a size biased model to these speed data and a Poisson REM model to the same encounters generated before (Section 7.2.1).

(ii) Secondly, we investigate the independence assumption in iREM-SB as in the first case but instead we use a random sample of the speed data (Section 7.2.2).

For these first two cases we compare estimations from Rowcliffe et al. (2008) REM with size biased sampling (REM-SB) with estimations from iREM-SB.

(iii) Thirdly, we investigate the importance of accounting for the bias in the speed of faster moving animals using all of the speed data. To do this we fit size biased speed data models and standard speed data models to data simulated from the speed data models and compare estimations (Section 7.2.3).

(iv) Fourthly, we investigate the importance of accounting for the bias in the speed of faster moving animals as in the third case but instead using a random sample of the speed data (Section 7.2.4).

For the last two cases we compare estimations obtained from iREM-SB and iREM with estimations from REM-SB and REM. The nonparametric REM formula of the density discussed in Chapters 2 and 5 is defined as

$$D = \frac{\sum_{i=1}^{c} a_i}{\sum_{i=1}^{c} t_i} \times \frac{\pi}{(2+\theta) r \bar{v}}, \tag{7.2.1}$$

where a_i is the encounters on the ith camera trap for $i = 1, 2, ..., c$ camera trap, t_i is the camera trapping time period of the ith camera trap, the fixed parameters θ and r are the detection angle and detection distance, respectively, and $\bar{v}$ is the mean speed. The variance of the density is computed using the Fisher Information matrix as given in Section 5.5

For the simulations the parameters and sample sizes chosen are considered plausible for ecological surveys. Each simulation was repeated 100 times and the average of the parameter estimates was computed. The parameters required to estimate density are the detection distance $r = 0.0125$ (km); detection angle $\theta = 0.10$ (radians), and the mean camera trap time, 7.59 which are held fixed. From the camera trap time, $T = \{t_1, t_2, ..., t_c\}$, where $i = 1, 2, ...c$ is the ith camera, we can compute the mean camera trap time as $\sum_{i=1}^{c} T/c$. For each simulation we also compute the total camera trap time as $\sum_{i=1}^{c} t_i$.

Step 1: First we start with M animals, each moving at a speed of x_l $(l = 1, 2, ..., M)$. The speeds, x_l are random numbers generated from either a gamma distribution $X \sim \text{Ga}(\alpha, \nu)$, a lognormal distribution $X \sim \ln N(\mu_x, \nu)$ or a Weibull distribution $X \sim \text{Wei}(\beta, \nu)$. We also generate the camera trap days, $T = \{t_1, t_2, ..., t_c\}$ with size c from a gamma distribution $T \sim \ln N(\tau, \beta^2)$. Then we calculate the mean encounter $\mathbf{\Lambda} = \{(2 + \theta)/\pi\} r t_i x_l D$ where $\mathbf{\Lambda}$ is an $c \times M$ matrix of mean encounters of individuals with speed x_l for t_i camera trap days; r (km) and θ (radians) are the detection zone dimensions, which are held fixed. The symbol $D = 1$ is animal density. Suppose that the population size is 1; then the animals being considered results in a density of 1.

Note that the camera trap time, T, is not limited to a lognormal model. Other relevant distributions such as a Weibull are also applicable. In the simulation study, we use a gamma model as a demonstration to the simulate the speeds. This is shown in the example below.

Example

This example consists of a population size of $M = 4$ animals with speeds (in km/day^{-1}) following a gamma model with parameters $\alpha = 388.8$; $\nu = 30$, recorded on $c = 3$ camera traps. The camera trap days, T, are simulated from a lognormal model with parameters $\beta = 0.0577$ and $\tau^* = 2.0251$, where τ^* is the mean of the logarithm;. The parameters required to estimate the mean encounter rate are $r = 0.0125$ (km) and $\theta = 0.10$ (radians). The R-code is given below.

```
###generating speeds for 4 individual animals
   M    <- 4; alpha <- 388.8;  nu  <- 30
   X    <- rgamma(M,alpha,nu)
   12.47096 12.84364 13.57303 13.60024


###generating camera trap days from a gamma model
   c   <- 3; tau* <- 2.0251;  beta <- 0.0577
   T   <- rlnorm(c,tau*,beta)
   7.753769 8.315009 7.665734


###computing the constant term in the expected encounter rate
   D     <- 1; r  <- 0.012;  theta <- 0.10
   C     <- ((2+ theta)/ pi) x r x D


###computing the expected encounter rate
  Lambda <- C x (sapply(X, function(a) sapply(T,function(b) axb)))
        [,1]      [,2]      [,3]      [,4]
[1,] 0.7756456 0.7988251 0.8441906 0.8458830
[2,] 0.8317890 0.8566463 0.9052954 0.9071103
[3,] 0.7668390 0.7897553 0.8346057 0.8362789
```

Step 2: Next, for the M individuals moving at speed x_l generated in *Step* 1, we generate the number of encounters from a Poisson distribution $\mathbf{a} \sim \text{Poi}(\mathbf{\Lambda})$, where $\mathbf{a}$ is a $c \times M$ matrix of encounters. We then calculate total encounters, $\mathbf{a}^*$, on the ith camera by computing the sum over $i(j)$. We also compute the total number of times, $\mathbf{a}^+$, an animal was seen as the sum over $j(i)$. The R-code and the R-output are shown in the example below.

Example continued

```
###generating encounters from a Poisson distribution
for(k in 1 :c){
     for(j in 1 :M){
       a[k,j] = rpois(1,Lambda)
       a*=rowSums(a)}
       a+ =colSums(a)}
### c x M matrix of encounters of animals
     a
     [,1] [,2] [,3] [,4]
[1,]    0    2    2    1
[2,]    1    0    0    1
[3,]    0    1    0    0
###total encounters on the ith camera trap
  a*
  5 2 1
###total number of times an animal was recorded
  a+
  1 3 2 2
```

Step 3: Using the number of encounters generated in *Step* 2 we then generate the actual individual speeds recorded by the camera traps, $X^* = \{x^*_1, x^*_2, ..., x^*_{M^*}\}$ where M^* is the actual sample size. The R-code and R-output for this example are given below.

Example continued

```
###generating actual speed data
  X* <- rep(X,a+)
  12.47096  12.84364  12.84364  12.84364
  13.57303  13.57303  13.60024  13.60024
```

At this point we can either use all of the speed data generated or a random sample of the speed data as is usually done in practice. As discussed previously, extracting

the speed data from camera trap records is not a simple task, therefore, we generate a random sample of the speed data in our simulations to test the model. The example is continued below with the selection of random sample of speed data. We give the R-code and R-output.

Example continued

```
###selecting a random sample X1* of size M from X*
  X*
  12.47096  12.84364  12.84364  12.84364
  13.57303  13.57303  13.60024  13.60024

  X1* <- sample(X*, size=M,replace=TRUE)
   12.84364 13.57303 12.84364 12.84364
```

Step 4: Fit an iREM-SB with a Poisson distribution for the encounter data, a^*, and a gamma distribution for the speed data, using either X^* or ${X^*}_1$.

To investigate the assumption of independence between data sets we generate an independent random sample from a gamma distribution, which represents an independent set of animal speed data, Z (km/day^{-1}). Using this new data set, Z (km/day^{-1}) we compute the expected encounter rate, $\mathbf{\Lambda}$ in *Step* 1 and then repeat *Step* 2. The example for this is given in the R-code and R-output below.

Example continued

```
###generating new set of speed data from a gamma model
  M <- 4; alpha  <- 388.8 ; nu <- 30 ;
  Z <- rgamma(M,alpha,nu)
 13.07493 13.69886 12.98350 12.75520

###computing the constant term in the expected encounter rate
  D <- 1; r <- 0.012; theta <- 0.10
  C <- ((2+ theta)/ pi) x r x D

###computing the expected encounter rate
 Lambda1 <- C x (sapply(Z, function(a) sapply(T,function(b) axb)))
       [,1]      [,2]      [,3]      [,4]
[1,] 0.8132102 0.8520166 0.8075239 0.7933246
[2,] 0.8720726 0.9136879 0.8659747 0.8507477
[3,] 0.8039772 0.8423429 0.7983554 0.7843173

###generating encounters from a Poisson model
 for(k in 1 :c){for(j in 1 :M){a1[k,j] = rpois(1,Lambda1)
       a1*=rowSums(a1)}
       a1+ =colSums(a1)}
  a1
     [,1] [,2] [,3] [,4]
[1,]    1    0    0    1
[2,]    0    0    1    1
[3,]    2    3    1    2

###total encounters on the ith camera trap
  a1* =   2 2 8
###total number of times an animal was recorded
  a1+ =  3 3 2 4
```

Using $\mathbf{a_1}^+$, generated in the example above we then generate the actual speed data Z^* (km/day) to be used in the estimation process. At this stage we also select a random sample, ${Z_1}^*$, of animal speed from Z^*. The R-code and R-output for this is shown below.

Example continued

```
###generating actual speed data Z*
 Z* <- rep(Z,a1+)
 13.07493 13.07493 13.07493 13.69886 13.69886 13.69886
 12.98350 12.98350 12.75520 12.75520 12.75520 12.75520


###selecting a random sample of speed Z1* from Z*
 Z1* <- sample(Z*, size=M,replace=TRUE)
 Z1*
 13.69886 12.75520 13.69886 12.75520
```

We then fit a size biased gamma to the independent speed data Z^* or ${Z_1}^*$.

7.2.1 Investigating independence assumption with all possible speed data

In this section we investigate the independence assumption using the simulation algorithm described in Section 7.2. We test the performance of the model for small density, $D = 10$ (km^2), and low expected speed $\mu_x = 0.150$ (ms^{-1}) where the variance of animal speed is low, $\text{Var}(x_l) = 0.05$, or high, $\text{Var}(x_l) = 0.31$. We also test the performance of the model for a larger density value, $D = 100$ (km^2) and higher expected speed, $\mu_x = 0.466$ (ms^{-1}) where the variability in speed is low, $\text{Var}(x_l) = 0.005$, or high, $\text{Var}(x_l) = 0.35$. For each scenario we set the number of camera traps to $c = 10$ or $c = 100$.

In Table 7.2.1 we compare results from REM with size biased sampling (REM-SB) with an iREM-SB using independent data and dependent data. We generated scenarios in which the true density value, $D = 100$ (km^2), was such that animals moving at a fast rate have expected speed of $\mu_x = 0.466$ (ms^{-1}) with low or high sampling variability;

$\text{Var}(x_l) = 0.005$ or $\text{Var}(x_l) = 0.35$, respectively. The standard errors are in parentheses, and the Standard (Sd) and Root Mean Square Error (RMSE) are also given. The simulation results show that violating the independence assumption has minor consequences on estimated density and its precision. That is, the difference between the density estimates from the independent data and the dependent data is minimal. Also, it is worth reiterating here that the method of size biased sampling accounts for the dependence introduced by faster moving animals that are more likely to be captured on camera traps, hence, minimal differences in density estimates between dependent and independent data. The results also show that REM-SB gave similar estimates of the density as iREM-SB but the standard error estimated from REM-SB is smaller than the standard error estimated from iREM-SB, particularly when the sampling variability in the speed is high. Increasing the number camera traps, hence, the number of camera trapping days improves estimation precision.

In Table E.0.3, appendix E, we show alternative scenarios for smaller true density values, $D = 10$ (km^2) and low expected speed, $\mu_x = 0.156$ (ms^{-1}) with low variability, $\text{Var}(x_l) = 0.05$ or high variability, $\text{Var}(x_l) = 0.31$ in the speed data. Similarly, there are minor consequences on the density estimator and precision if the independence assumption is violated. While REM-SB and iREM-SB gave similar estimates of the density, the difference in estimated standard of the density between REM-SB and iREM-SB is more obvious. REM-SB consistently gave smaller estimates of the standard error, particularly when the variability in the speed data is high.

Table 7.2.1: Simulation results from using all the speed data with a large density D; small and large number of camera traps c; and a fixed value of expected speed $\mu_x = 0.466$ (ms^{-1}) with low and high variability in animal speed. The average parameter estimates (standard errors are in parentheses), the Standard deviation (Sd) and Root Mean Square Error (RMSE) are given.

	Dependent data			**Independent data**			
	REM-SB	iREM-SB		REM-SB	iREM-SB		Camera trap days
	$\hat{D}$	$\hat{D}$	$\hat{\mu}_x$	$\hat{D}$	$\hat{D}$	$\hat{\mu}_x$	$\bar{n}$
				Low variability in animal speed (= 0.005)			
$D = 100;\ c = 10; \mu_x = 0.466; n = 75.08$							
estimate	100.48 (6.37)	100.48 (6.40)	0.467 (0.001)	100.47 (6.37)	100.47 (6.40)	0.467 (0.0005)	76.13
Sd	6.32	6.99	0.001	6.33	6.99	0.001	1.73
RMSE	6.39	6.42	0.001	6.38	6.42	0.0005	2.03
$D = 100;\ c = 100; \mu_x = 0.466; n = 758.94$							
estimate	100.03 (2.01)	100.03 (2.02)	0.466 (0.0002)	100.03 (2.01)	100.03 (2.02)	0.466 (0.0002)	7758.94
Sd	2.26	2.26	0.001	2.25	2.25	0.001	1.88
RMSE	2.01	2.02	0.002	2.01	2.02	0.0002	1.93
				High variability in animal speed (= 0.35)			
$D = 100;\ c = 10; \mu_x = 0.466; n = 74.98$							
estimate	100.27 (6.41)	100.27 (6.58)	0.458 (0.004)	100.54 (6.43)	100.54 (6.59)	0.457 (0.004)	75.92
Sd	6.85	6.85	0.01	6.43	6.59	0.004	1.69
RMSE	6.42	6.58	0.01	6.45	6.61	0.01	1.93
$D = 100;\ c = 100; \mu_x = 0.466; n = 761.70$							
estimate	99.54 (2.03)	99.55 (2.07)	0.456 (0.001)	99.42 (2.02)	99.42 (2.07)	0.457 (0.001)	759.02
Sd	2.75	2.77	0.01	2.19	2.18	0.01	1.74
RMSE	2.08	2.12	0.01	2.11	2.14	0.01	3.20

7.2.2 *Investigating independence assumption using a random sample of speed data*

Generally, it is not practical to extract information regarding the speed from all camera traps over the survey period, particularly for large scale survey designs. Therefore, a random sample that is considered large enough to provide relevant information in ecological studies is selected (see Rowcliffe et al., 2011). In this section we investigate the independence assumption using the simulation algorithm outlined in Section 7.2 using a simple random sample of the speed data. We generated scenarios in which the true density values, $D = 10$ (km^2) or $D = 100$ (km^2), was such that animals moving at a slow rate has expected speed of $\mu_x = 0.150$ (ms^{-1}) and $\text{Var}(x_l) = 0.05$.

Table 7.2.2 gives the results from using a random sample of independent data and dependent data to estimate animal density from iREM-SB and REM-SB. Analysing the data assuming independence in encounter and speed of movement data provides relatively accurate estimates of the density for smaller sampling effort. Again, there is minimal difference in the density estimator between the dependent and independent data. Correcting for the bias in the speed of faster moving animals accounts for dependence of these animals being more likely to encounter the camera trap. Also, REM-SB gave similar estimates of the density as iREM-SB, but the estimated standard error of the density is smaller compared with the estimated standard error from iREM-SB.

Table 7.2.2: Simulation results from using a random sample of animal speed data with small and large density D, small and large number of camera traps c, and expected speed $\mu_x = 0.150$ (ms^{-1}). The average parameter estimates (standard error in parentheses), the Standard deviation (Sd) and Root Mean Square Error (RMSE) are given.

	Dependent data			**Independent data**			
	REM-SB	iREM-SB		REM-SB	iREM-SB		Camera trap days
	$\hat{D}$	$\hat{D}$	$\hat{\mu}_x$	$\hat{D}$	$\hat{D}$	$\hat{\mu}_x$	$\bar{n}$
$D = 10;\ c = 10; \mu_x = 0.156; n = 75.96$							
estimate	10.62 (3.57)	10.25 (3.91)	0.151 (0.01)	10.84 (3.64)	10.76 (3.99)	0.150 (0.01)	75.80
Sd	3.63	3.99	0.01	3.67	3.39	0.003	1.77
RMSE	3.62	3.92	0.01	3.74	4.07	0.004	1.77
$D = 100;\ c = 10; \mu_x = 0.156; n = 75.64$							
estimate	102.88 (11.27)	102.78 (11.63)	0.151 (0.002)	102.88 (11.26)	102.45 (11.62)	0.151 (0.003)	76.34
Sd	11.80	11.63	0.01	11.57	11.83	0.01	1.56
RMSE	11.63	11.96	0.01	11.62	11.88	0.01	1.71
$D = 10;\ c = 100; \mu_x = 0.156; n = 754.64$							
estimate	10.29 (1.13)	10.35 (1.27)	0.152 (0.01)	10.31 (1.13)	10.35 (1.28)	0.152 (0.01)	758.83
Sd	1.81	1.97	0.01	1.92	1.99	0.01	1.70
RMSE	1.17	1.32	0.01	1.17	1.33	0.01	4.52
$D = 100;\ c = 100; \mu_x = 0.156; n = 761.57$							
estimate	102.11 (3.55)	102.92 (3.97)	0.152 (0.002)	102.49 (3.57)	102.73 (4.01)	0.152 (0.003)	758.96
Sd	3.57	3.99	0.004	3.35	4.63	0.004	1.82
RMSE	4.13	4.41	0.005	4.35	4.85	0.004	3.17

7.2.3 *Investigating the importance of accounting for bias in the speed of faster moving animals using all possible speed data*

To assess the impact that unaccounted bias in animal speed of faster moving animals can have on the estimation of density in iREM we simulate scenarios with iREM-SB and iREM without size biased sampling. Table 7.2.3 gives the true values used in the simulation process, which are ecologically plausible for our motivating BCI data set. We test the size biased sampling method under different parameter settings and sample size conditions, looking at the effect on density when the variance is large or small. We generate independent data sets to test the size bias method.

As shown in Chapter 6, a size biased gamma model remains a gamma model but with mean $(\alpha+1)/\nu$ and variance $(\alpha+1)/\nu^2$ (note that the mean of a gamma model is α/ν and variance α/ν^2). The difference between the size biased mean and the standard mean is $1/\nu$, and therefore, it is the shape parameter ν that will determine how large the difference between a size biased mean and a standard mean is. So as ν gets larger, the difference between the two means gets smaller, and hence, the variance also gets smaller.

Table 7.2.3: True values used in the simulation process.

Density	Camera traps	Parameters and Variance			
		μ_x	ν	α	Var(x_l)
Low speed, increasing variability					
10	10	0.150	15	194.4	0.01
10	10	0.150	3	38.88	0.05
10	10	0.150	1.5	19.44	0.10
10	10	0.150	1	12.96	0.15
Higher speed, high variability					
10	10	0.466	0.78	31.37	0.60
100	10	0.466	0.78	31.37	0.60
10	100	0.466	0.78	31.37	0.60
100	100	0.466	0.78	31.37	0.60

Table 7.2.4 gives the results from fitting iREM-SB and iREM for expected speed, $\mu_x = 0.150$ (ms^{-1}) and increasing variability of the speed. As discussed above when the variability in the speed data is low, hence, ν is large, we expect the difference in estimates between a gamma model and size biased gamma model to be minimal. The simulation results show that this is true for $\text{Var}(x_l) = 0.01$. Increasing the variability in the speed results in a larger bias in the density estimate from a standard gamma model. Additionally, estimated standard error of the density is smaller compared with the estimates from a size biased gamma model. The results also show that REMs gave similar estimates of the density as iREMs but the estimated standard from REMs is smaller compared with iREMs. Note again that REM does not account for the sampling variability in the speed data but rather uses a fixed estimate of the expected speed.

We test the models using a higher expected speed, $\mu_x = 0.466$ (ms^{-1}) with high variability in the speed data and the results are shown in Table 7.2.5 . We expect a size biased model to perform better than a standard model when variance is large, and hence, large shape parameter, ν. The simulation results (Table 7.2.5) show that the size biased model performs better than the standard under all scenarios in terms of the parameter estimates and precision. Again, REMs can gave similar estimates of the density as iREM but the estimated standard error of the density is smaller compared with the estimated standard error from iREMs a

Based on these results we would recommend a size biased model for estimation of average animal speed, particularly if the sampling variability of the data is high, since this model reduces the bias in the density estimator as well as the mean speed estimator.

Table 7.2.4: Simulation results for small density D; small number of camera traps c; and a fixed value of expected speed $\mu_x = 0.150$ (ms^{-1}) and increasing variance of animal speed $\text{Var}(x_l)$. The average parameter estimates (standard errors in parentheses), the Standard deviation (Sd) and Root Mean Square Error (RMSE) are given.

	size biased model			**standard model**			
	REM-SB	iREM-SB		REM	iREM		Camera trap days
	$\hat{D}$	$\hat{D}$	$\hat{\mu}_x$	$\hat{D}$	$\hat{D}$	$\hat{\mu}_x$	$\bar{n}$
$D = 10;\ c = 10; n = 73.66; \text{Var}(\text{x}_\text{l}) = 0.01$							
	9.70 (3.14)	9.73 (3.65)	0.155 (0.01)	9.66 (3.13)	9.88 (3.68)	0.156 (0.005)	75.73
Sd	3.88	3.15	0.01	3.86	3.23	0.01	1.48
RMSE	3.16	3.66	0.01	3.15	3.68	0.005	2.54
$D = 10;\ c = 10; n = 73.92; \text{Var}(\text{x}_\text{l}) = 0.05$							
	10.10 (3.21)	10.29 (3.83)	0.155 (0.01)	9.88 (3.14)	9.99 (3.75)	0.158 (0.01)	75.85
Sd	3.55	3.83	0.01	3.48	3.75	0.01	1.75
RMSE	3.21	3.85	0.01	3.15	3.75	0.01	2.61
$D = 10;\ c = 10; n = 72.47; \text{Var}(\text{x}_\text{l}) = 0.10$							
	10.68 (3.40)	10.73 (4.28)	0.150 (0.02)	9.97 (3.17)	9.96 (3.98)	0.156 (0.02)	75.78
Sd	4.69	4.69	0.02	4.36	4.17	0.02	1.78
RMSE	3.47	4.34	0.02	3.18	3.98	0.02	3.76
$D = 10;\ c = 10; n = 76.18; \text{Var}(\text{x}_\text{l}) = 0.15$							
	10.47 (3.36)	10.64 (4.42)	0.150 (0.02)	9.75 (3.12)	9.98 (3.84)	0.23 (0.02)	75.98
Sd	5.25	3.41	0.02	4.78	3.07	0.02	1.66
RMSE	3.39	4.27	0.02	3.13	3.84	0.02	1.67

Table 7.2.5: Simulation results from using all of the animal speed data for $\mu_x = 0.466$ (ms^{-1}) with high variability in animal speed, $\text{Var}(x_l) = 0.60$. The average parameter estimates (standard errors are in parentheses), the Standard deviation (Sd) and Root Mean Square Error (RMSE) are given.

	size biased model			**standard model**			
	REM-SB	iREM-SB		REM	iREM		Camera trap days
	$\hat{D}$	$\hat{D}$	$\hat{\mu}_x$	$\hat{D}$	$\hat{D}$	$\hat{\mu}_x$	$\bar{n}$
$D = 10;\ c = 10; \mu_x = 0.466; n = 73.79$							
	10.28 (1.86)	10.43 (2.15)	0.472 (0.02)	9.99 (1.80)	9.85 (2.06)	0.487 (0.02)	75.58
Sd	2.91	2.90	0.03	2.85	2.86	0.04	1.77
RMSE	1.88	2.19	0.02	1.80	2.07	0.02	2.51
$D = 100;\ c = 10; \mu_x = 0.466; n = 72.91$							
	101.94 (5.86)	102.49 (6.67)	0.468 (0.01)	98.69 (5.66)	98.96 (6.44)	0.511 (0.01)	76.03
Sd	6.55	6.10	0.01	6.34	6.64	0.01	1.58
RMSE	6.17	7.12	0.01	5.81	6.52	0.01	3.50
$D = 10;\ c = 100; \mu_x = 0.466; n = 7557.32$							
	10.25 (0.59)	10.34 (0.67)	0.466 (0.01)	9.94 (0.57)	9.79 (0.64)	0.481 (0.01)	759.28
Sd	0.63	0.71	0.03	0.64	0.68	0.02	2.03
RMSE	0.64	0.75	0.02	0.57	0.67	0.05	2.82
$D = 100;\ c = 100; \mu_x = 0.466; n = 760.29$							
	101.99 (1.86)	101.79 (2.09)	0.466 (0.002)	98.69 (1.80)	98.56 (2.02)	0.483 (0.002)	758.74
Sd	1.65	2.65	0.01	1.90	2.58	0.01	1.79
RMSE	2.72	2.75	0.01	2.23	2.48	0.02	2.37

7.2.4 *Investigating the importance of accounting for bias in the speed of faster moving animals using a random sample of speed data*

In this section we investigate the importance of accounting for the bias in the speed of faster moving animals using a random sample of size $m = 70$ of the speed data as described in the simulation algorithm outlined above in Section 7.2. For illustration, we test the model for $\mu_x = 0.466$ (ms^{-1}) and increasing sampling variability. We set $\text{Var}(x_l) = 0.15$; $\text{Var}(x_l) = 0.30$; $\text{Var}(x_l) = 0.45$ or $\text{Var}(x_l) = 0.60$. The results are given in Table 7.2.6. We expect the size biased model to perform better than the standard model as the sampling variability increases. The results (Table 7.2.6) show that the bias from the size biased gamma is consistently smaller compared with the bias in estimated density from the standard gamma model. Also, the estimated standard error of the density from a standard gamma is smaller compared with that from the size biased gamma model. REMs and iREMs gave similar estimates of the density but REMs gave smaller estimates of the standard error of density since the variability in the speed data is unaccounted for in REM. Finally, using a random sample of the speed data gave relatively accurate estimates of the density, for example when $\text{Var}(x_l) = 0.60$ as is the case in Table 7.2.5 above. Therefore, we can conclude that using a random sample of the observed speed data in practice would have minimal effect on the density and its precision.

Table 7.2.6: Simulation results from using a random sample of the animal speed data, and for $\mu_x = 0.466$ (ms^{-1}) with increasing sampling variability of animal speed. The average parameter estimates (standard errors in parentheses), the Standard deviation (Sd) and Root Mean Square Error (RMSE) are given.

	size biased model			**standard model**			
	REM-SB	iREM-SB		REM	iREM		Camera trap days
	$\hat{D}$	$\hat{D}$	$\hat{\mu}_x$	$\hat{D}$	$\hat{D}$	$\hat{\mu}_x$	$\bar{n}$
$D = 100;\ c = 100; \mu_x = 0.466; n = 756.12;\ \text{Var}(x_l) = 0.15$							
	100.66 (1.82)	100.60 (2.27)	0.480 (0.01)	99.88 (1.81)	99.87 (2.25)	0.483 (0.01)	759.17
Sd	2.39	2.66	0.01	2.33	2.46	0.01	1.89
RMSE	1.93	2.35	0.01	1.81	2.264	0.01	3.59
$D = 100;\ c = 100; \mu_x = 0.466; n = 760.61;\ \text{Var}(x_l) = 0.30$							
	100.50 (1.82)	100.29 (2.51)	0.476 (0.01)	98.86 (1.80)	98.88 (2.48)	0.482 (0.01)	759.26
Sd	2.34	2.65	0.01	2.16	2.04	0.01	1.91
RMSE	1.89	2.53	0.01	2.08	2.72	0.01	2.34
$D = 100;\ c = 100; \mu_x = 0.466; n = 758.52\ ;\ \text{Var}(x_l) = 0.45$							
	101.02 (1.84)	100.99 (2.79)	0.473 (0.01)	98.65 (1.80)	98.91 (2.70)	0.483 (0.01)	759.00
Sd	1.56	2.52	0.01	2.17	2.31	0.02	1.65
RMSE	2.10	2.96	0.01	2.24	2.91	0.01	1.72
$D = 100;\ c = 100; \mu_x = 0.466; n = 759.51;\ \text{Var}(x_l) = 0.60$							
	102.06 (1.85)	102.68 (3.08)	0.469 (0.01)	98.82 (1.80)	99.44 (2.92)	0.483 (0.01)	758.90
Sd	3.05	3.81	0.02	2.94	2.39	0.02	1.69
RMSE	2.77	4.08	0.02	2.14	2.97	0.01	1.80

7.3 Application of iREM-D with size biased sampling model to BCI data

In this section we fit the iREM-D developed in Chapter 5 to the BCI data, accounting for the bias in the speed of faster moving animals. As discussed in Chapter 6, faster moving animals are more likely to encounter camera traps and the probability of encounter may differ within species. We have illustrated that the speed estimator can be biased towards these faster animals. To correct for the bias introduced by faster animals, we adopted the method of size biased sampling. We believe that the bias introduced from faster animals would bias the density estimator. We assume the models discussed in Chapter 5 for the speed data (gamma, lognormal or a Weibull), a halfnormal model is assumed for the distance data; a Poisson REM is assumed for the encounter data, and a von Mises model is assumed for the angle to detection data.

Suppose the encounter data, a_i, is assumed to follow a Poisson distribution, such that $a_i \sim \text{Pois}(\lambda t_i)$ where λ is the expected encounter rate defined in equation (7.1.1) above, and t_i is the camera trap time in days for $i = 1, 2, ..., c$ camera trap. Then the probability mass function can be defined as $h(a_i \mid \lambda,\ t_i,\ \tau)$ where τ represents any additional parameters in the model. Suppose also that m speed observations, with expected speed, μ_x, have probability density function $f(x_l \mid \mu_x,\ \nu)$ where $l = 1, 2, ..., m$, and ν represents any additional parameters in the model. So for s_i observed detection distances which have probability density function $g(z_{ij} \mid \sigma)$ where $j = 1, 2, ..., s_i$ is the jth observed detection distance, on the ith camera trap, and σ represents any additional parameter in the model, and the corresponding angles to detection θ_{ij} with probability density function $p(\theta_{ij} \mid \upsilon,\ \eta)$, where υ is the expected angle to detection and η represents any additional parameters in the model, the log-likelihood can constructed as follows

$$\begin{aligned}\ell = &\sum_{i=1}^{c} \log\ h(a_i \mid \lambda,\ t_i,\ \tau) + \sum_{l=1}^{m} \log\left(\frac{x_l f(x_l \mid \mu_x,\ \nu)}{\mu_x}\right) + \\ &\sum_{i=1}^{c}\sum_{j=1}^{s_i} \log\ g(z_{ij} \mid \sigma) + \sum_{i=1}^{c}\sum_{j=1}^{s_i} \log\ p(\theta_{ij} \mid \upsilon,\ \eta).\end{aligned} \tag{7.3.1}$$

The model assumes the following: 1) the encounters at each camera trap are independent, 2) the speed observations are independent and identically distributed, and 3) the

encounters, speeds, detection distance and angle to detection are independent so that their contribution to the likelihood could be multiplied. We showed in the simulations in Section 7.2 that a violation of the independence assumption between the speed data and encounter data has minimal effect on the density and its standard error. Also, as shown in Section 5.11 and as discussed in Rowcliffe et al. (2011) a weak correlation was found to exist between the detection distance and angle to detection data, hence, the assumption of independence is made. For comparison estimates from fitting the nonparametric REM formula, described in equation (7.2.1) above, to the data using an estimate of the mean speed computed from the harmonic mean formula described in equation 6.4 and the standard (arithmetic) mean formula are computed. Table 7.3.1 gives the estimates from lognormal models. Table 7.3.2 gives the results from gamma models, while Table 7.3.3 gives the results from Weibull models.

The three speed data models (Table 7.3.1, Table 7.3.2, Table 7.3.3) gave similar estimates of the parameters for all species. The evidence also shows that REM and iREM-D gave similar estimates of the density but the estimated standard errors from REM are smaller. It is worth reiterating here that REM lacks the potential to account for the sampling variability in the parameters required to estimate density, hence, we expect iREM-D to give better estimates of the standard error. The size biased models also perform better than the standard models, as expected, as the bias in animal speed is accounted for. Note that the density estimator is defined as

$$\hat{D} = \frac{\hat{\lambda}}{t} \times \frac{\pi}{(2+\hat{\upsilon})\hat{\mu}_x\hat{\gamma}},$$

where $\hat{\lambda}$ is the expected encounter rate, $\hat{\gamma}$ is the effective detection distance, $\hat{\upsilon}$ is the mean angle to detection and $\hat{\mu}_x$ is the expected animal speed. So for larger values of $\hat{\mu}_x$ we would expect the estimated density, $\hat{D}$, to decrease as shown by the results in Table 7.3.1, Table 7.3.2 and Table 7.3.3. According to the ΔAIC values in Table 7.3.4 there is more support for a lognormal model. As discussed in Chapter 6 a size biased gamma model or size biased lognormal model is a gamma model or lognormal model, respectively, but with different parameters, while a size biased Weibull model is not a Weibull model.

Table 7.3.1: Estimates from a size biased lognormal distribution compared with estimates from a lognormal distribution where detection distance is assumed to follow a halfnormal model (standard errors in parentheses).

Species	**size biased lognormal**					**lognormal**				
	REM	iREM-D				REM	iREM-D			
	$\hat{D}$	$\hat{D}$	$\hat{\mu}$	$\hat{\gamma}$	$\hat{v}$	$\hat{D}$	$\hat{D}$	$\hat{\mu}$	$\hat{\gamma}$	$\hat{v}$
ocelot	1.40 (0.08)	1.40 (0.15)	0.28 (0.02)	2.15 (0.19)	0.24 (0.02)	0.96 (0.05)	0.96 (0.10)	0.41 (0.03)	2.15 (0.19)	0.24 (0.02)
coati	5.52 (0.19)	5.52 (0.65)	0.16 (0.01)	2.49 (0.28)	0.20 (0.02)	2.60 (0.09)	2.60 (0.30)	0.33 (0.03)	2.49 (0.28)	0.20 (0.02)
rat	13.26 (0.44)	13.26 (1.34)	0.12 (0.01)	1.47 (0.07)	0.24 (0.04)	6.73 (0.22)	6.73 (0.67)	0.23 (0.02)	1.47 (0.07)	0.24 (0.02)
peccary	23.36 (0.33)	23.36 (1.77)	0.15 (0.01)	3.47 (0.38)	0.27 (0.02)	11.65 (0.16)	11.65 (0.97)	0.30 (0.02)	3.18 (0.34)	0.27 (0.02)
brocket	4.16 (0.14)	4.16 (0.38)	0.16 (0.01)	3.26 (0.41)	0.24 (0.10)	2.31 (0.08)	2.31 (0.21)	0.28 (0.02)	3.26 (0.41)	0.24 (0.10)
paca	5.86 (0.18)	5.86 (0.43)	0.18 (0.01)	2.55 (0.20)	0.23 (0.06)	3.83 (0.12)	3.83 (0.26)	0.27 (0.01)	2.55 (0.11)	0.23 (0.06)
agouti	85.12 (0.80)	85.12 (3.31)	0.14 (0.004)	2.43 (0.08)	0.23 (0.01)	44.43 (0.42)	44.43 (1.61)	0.26 (0.01)	2.43 (0.08)	0.23 (0.01)
squirrel	5.95 (0.24)	5.95 (0.87)	0.12 (0.02)	2.01 (0.19)	0.24 (0.03)	2.87 (0.12)	2.87 (0.43)	0.26 (0.03)	2.01 (0.19)	0.24 (0.03)
mouse	1.88 (0.18)	1.88 (0.34)	0.10 (0.01)	1.40 (0.11)	0.20 (0.03)	1.05 (0.10)	1.05 (0.19)	0.17 (0.02)	1.40 (0.11)	0.20 (0.03)

Table 7.3.2: Estimates from a size biased gamma distribution compared with estimates from a gamma distribution where detection distance is modelled by a halfnormal distribution (standard errors in parentheses).

Species	**size biased gamma**					**gamma**				
	REM	iREM				REM	iREM			
	$\hat{D}$	$\hat{D}$	$\hat{\mu}$	$\hat{\gamma}$	$\hat{\upsilon}$	$\hat{D}$	$\hat{D}$	$\hat{\mu}$	$\hat{\gamma}$	$\hat{\upsilon}$
ocelot	1.40 (0.08)	1.40 (0.16)	0.28 (0.03)	2.15 (0.27)	0.24 (0.02)	0.96 (0.05)	0.96 (0.09)	0.41 (0.02)	2.15 (0.13)	0.24 (0.02)
coati	5.52 (0.19)	5.52 (0.93)	0.16 (0.02)	2.49 (0.42)	0.20 (0.02)	2.60 (0.09)	2.60 (0.25)	0.33 (0.03)	2.49 (0.16)	0.20 (0.02)
rat	13.26 (0.44)	13.26 (2.01)	0.12 (0.02)	1.47 (0.22)	0.24 (0.02)	6.73 (0.22)	6.73 (0.58)	0.23 (0.01)	1.47 (0.07)	0.24 (0.04)
peccary	23.36 (0.33)	23.36 (2.73)	0.15 (0.02)	3.47 (0.40)	0.27 (0.02)	11.65 (0.16)	11.65 (0.80)	0.31 (0.02)	3.47 (0.15)	0.27 (0.02)
brocket	4.16 (0.14)	4.16 (0.50)	0.16 (0.02)	3.26 (0.39)	0.24 (0.10)	2.31 (0.08)	2.31 (0.18)	0.28 (0.01)	3.26 (0.18)	0.24 (0.10)
paca	5.86 (0.18)	5.86 (0.55)	0.18 (0.01)	2.55 (0.24)	0.23 (0.02)	3.83 (0.12)	3.83 (0.26)	0.27 (0.01)	2.55 (0.11)	0.23 (0.02)
agouti	85.12 (0.80)	85.12 (4.71)	0.14 (0.01)	2.43 (0.13)	0.23 (0.01)	44.43 (0.42)	44.43 (1.45)	0.26 (0.01)	2.43 (0.04)	0.23 (0.01)
squirrel	5.95 (0.24)	5.95 (1.34)	0.12 (0.03)	2.01 (0.45)	0.24 (0.03)	2.87 (0.12)	2.87 (0.45)	0.26 (0.02)	2.01 (0.13)	0.23 (0.03)
mouse	1.88 (0.18)	1.88 (0.46)	0.10 (0.02)	1.40 (0.35)	0.20 (0.03)	1.05 (0.10)	1.05 (0.17)	0.17 (0.02)	1.40 (0.11)	0.20 (0.03)

Table 7.3.3: Estimates from a size biased Weibull distribution compared with estimates from a Weibull distribution where detection distance is modelled by a halfnormal distribution (standard errors in parentheses).

Species	**size biased Weibull**					**Weibull**				
	REM	iREM				REM	iREM			
	$\hat{D}$	$\hat{D}$	$\hat{\mu}$	$\hat{\gamma}$	$\hat{\upsilon}$	$\hat{D}$	$\hat{D}$	$\hat{\mu}$	$\hat{\gamma}$	$\hat{\upsilon}$
ocelot	1.40 (0.08)	1.40 (0.17)	0.28 (0.03)	2.15 (0.19)	0.24 (0.02)	0.96 (0.05)	0.96 (0.09)	0.41 (0.02)	2.15 (0.19)	0.24 (0.02)
coati	5.52 (0.19)	5.52 (0.75)	0.16 (0.02)	2.49 (0.28)	0.20 (0.02)	2.60 (0.09)	2.60 (0.29)	0.33 (0.03)	2.49 (0.28)	0.20 (0.09)
rat	13.26 (0.44)	13.26 (1.60)	0.12 (0.01)	1.47 (0.07)	0.24 (0.04)	6.73 (0.22)	6.73 (0.68)	0.23 (0.02)	1.47 (0.07)	0.24 (0.04)
peccary	23.36 (0.33)	23.36 (3.10)	0.15 (0.01)	3.47 (0.38)	0.27 (0.02)	11.65 (0.16)	11.65 (0.92)	0.31 (0.02)	3.47 (0.32)	0.27 (0.02)
brocket	4.16 (0.14)	4.16 (0.42)	0.16 (0.01)	3.26 (0.41)	0.24 (0.10)	2.31 (0.08)	2.31 (0.21)	0.28 (0.02)	3.26 (0.41)	0.24 (0.10)
paca	5.86 (0.18)	5.86 (0.47)	0.18 (0.01)	2.55 (0.20)	0.23 (0.02)	3.83 (0.12)	3.83 (0.30)	0.27 (0.02)	2.55 (0.20)	0.23 (0.02)
agouti	85.12 (0.80)	85.12 (3.81)	0.14 (0.01)	2.43 (0.08)	0.23 (0.01)	44.43 (0.42)	44.43 (1.61)	0.26 (0.01)	2.43 (0.08)	0.23 (0.01)
squirrel	5.95 (0.24)	5.95 (1.05)	0.12 (0.02)	2.01 (0.19)	0.24 (0.03)	2.87 (0.12)	2.87 (0.43)	0.26 (0.03)	2.01 (0.19)	0.24 (0.03)
mouse	1.88 (0.18)	1.88 (0.37)	0.10 (0.02)	1.40 (0.11)	0.20 (0.03)	1.05 (0.10)	1.05 (0.20)	0.17 (0.03)	1.40 (0.11)	0.20 (0.03)

Table 7.3.4: ΔAIC values for all models where detection distance follows a halfnormal distribution.

Species	size biased gamma	gamma	size biased lognormal	lognormal	size biased Weibull	Weibull
ocelot	0	0	6.97	6.97	0.08	3.38
coati	30.24	30.24	0	0	3.88	9.31
rat	56.01	56.01	0	0	25.35	25.63
peccary	51.09	51.09	0	0	90.63	119.63
brocket	34.89	34.89	0	0	11.41	27.71
paca	52.66	52.66	0	0	41.02	60.43
agouti	67.61	67.61	0	0	163.37	224.89
squirrel	37.04	37.04	0	0	16.23	16.27
mouse	14.24	14.24	0	0	6.06	12.22

7.4 Discussion

The iREM with size biased sampling model proposed in this Chapter allows for novel investigation of the effect of the sampling variability in the speed of faster moving animals on density estimation of unmarked animals. The model deals with the problem of induced-bias in the density estimator by accounting for the bias in the speed of faster moving animals. In particular, the integrated population model developed in this Chapter uses a Poisson REM process to estimate density from data sets that can be assumed to be independent.

In practice, demographic data about a population used in analyses are generally collected from a single survey and independence between the information is usually assumed. We found that disregarding the independence assumption between data sets does not appear to have strong effects on the density estimator in REM despite the fact the degree of the dependence in the data sets was strong, with all the data coming from the same source. This result is promising for ecological studies and is aligned with conclusions reached in other integrated population modelling approaches, which investigated the independence assumption between data sets (see for example Abadi et al., 2010). We considered the cases for all the speed data generated in our hypothetical survey and for a random sample of the speed data, as is the case of the real data at BCI, Panama. We also showed that correcting for the bias introduced by faster moving animals is crucial when estimating the density. Note that the model structure developed and tested via simulations assumes fixed values of the detection zone dimensions of the camera trap, hence, it is an iREM framework as developed in Chapter 3 but corrects for size bias. So, explorations of the iREM-D developed in Chapter 5, which accounts for the bias in animal speed is an avenue for future research.

To deal with the problem of accounting for the bias in animal speed and its effects on estimated density we adopted the size biased sampling method (Patil and Rao, 1978), which we have discussed in Chapter 6. We have shown in Chapter 6 that under certain conditions (small difference between the two means) estimations of expected speed from a size biased gamma model and a gamma model are approximately equal. In this

Chapter we have also shown that the difference in estimated expected speeds as well as estimated densities between an iREM-SB and an iREM could be small or large under certain conditions. If the sampling variability of the speed data is large, the difference in the estimated parameters from the two models is more pronounced compared with low sampling variability. The simulation results suggested that it is crucial to account for the bias in speed of faster moving animals as this bias can introduce bias in the density estimator. Also, the bias from using a random sample of the speed data is larger compared with the bias from using all of the speed data. This was expected since higher survey effort would improve estimation and its precision.

In this Chapter we have considered the encounter data to arise only from a Poisson distribution. In Chapters 2, 3, and 4, a negative binomial model, a zero-inflated Poisson model and a zero-inflated negative binomial model were also considered, and in Chapter 5 we also considered a negative binomial model and likelihood approaches were derived, but these alternatives for describing variation in the encounter data could be explored, for example to account for overdispersion and zero-inflation. Also, further work is needed to explore the effect of the detection zone dimensions on density for variable camera trapping time period.

An application of an iREM-D with size biased sampling model to real data at BCI produced realistic estimates of the density and provided a nice illustration of how accounting for bias in the speed of faster moving animals and the sampling variability in the detection zone dimensions can change the conclusions regarding the density estimator. The model structure assumes independence between the data sets and as shown in the simulation results ignoring the independence assumption has minor consequences on the parameter estimates and their standard errors. An analysis of the data assuming independence showed strong support for models that account for the sampling variability of faster species as shown in the case where animal speed is assumed to follow a Weibull model.

Chapter 8

Conclusions

In this thesis we have developed new approaches for density estimation of unmarked animals from camera trap data. During a period of habitat loss, climate change and loss of biodiversity, the availability of accurate and efficient modelling techniques is crucial for monitoring species, particularly those that cannot be identified to the individual level, and estimating population abundance. The novel models described in this thesis provide a basis for new and exciting future studies in camera trapping analysis for density estimation of unmarked species. The models developed in this thesis build on REM (Rowcliffe et al., 2008) for density estimation of unmarked animals from camera trap data.

We have first introduced REM, which is based on the "ideal gas" model from physics and we have demonstrated how REM can be used to estimate density using a small data set from WWAP with known census. As described in Chapter 1 abundance and density estimation of species, particularly of unmarked species, from camera trap data has become increasingly important in ecology. Of particular note is that the methods that use camera trap data for abundance and density estimation are restricted to capture-recapture analysis. The REM method described in this thesis is a recently developed modelling framework that aids in monitoring of unmarked species, and estimation of population density using camera trapping analysis. The WWAP data set, which was one of the primary data sets considered in this thesis, was used as the test data for REM, and the analysis of this data set raised several questions:

(i) Can REM, which simply applies the formula linking density with encounter rate,

animal speed, and detection zone parameters, using bootstrapping to estimate variance, adopt a maximum likelihood framework that allow for more robust model-based inference?

(ii) Is it possible to develop a unified statistical modelling framework that integrates the independent steps required to derive all the necessary components of the analysis (encounter rate, speed, detection distance, detection angle, and activity)?

(iii) How can REM be extended to account for variation in abundance in heterogeneous habitats?

(iv) The current REM approach is based on a flawed approximation of the detection process. Can a more valid process model be developed to improve estimation of the detection zone dimensions?

(v) There are multiple sources of variance in the analysis in the REM formula, each of which contributes to the overall precision of the density estimate. How does variance partition between the different sources empirically, and how do sample sizes in the difference sources affect overall precision?

(vi) Given that animal speeds and camera detection zones generally vary, what are the implications for unbiased estimation of the density in this situation?

In this thesis we addressed these six issues. The key to our work is to model the data entering REM to obtain unbiased estimates of the parameters and to correctly estimate the standard errors in REM. Maximum likelihood estimation approach in this context is attractive as it provides a unified framework for modelling data to estimate parameters and their variance. This approach provides a means to combine multiple data sets to estimate additional demographic parameters that are otherwise inestimable if separate analyses are conducted. Regarding the first question, we have developed a maximum likelihood framework for REM, which serves as a basis for further extension to estimate density.

The extended REM models developed for density estimation of unmarked animals are summarized by Figure 8.0.1. We started with REM, which is simple nonparametric

model for density estimation of unmarked species, assuming fixed values of the detection zone dimension parameters and an estimate of average animal speed, and built up more complex models taking into account the sampling variability in animal speed, and the detection zone dimensions, and incorporating covariates such as habitat type and camera random effect, which explain additional variation in population abundance. The model that accounts for the sampling variability in animal speed, and detection zone dimensions (iREM-D) is the most general model among those studied, with the others being particular or limiting cases of this one.

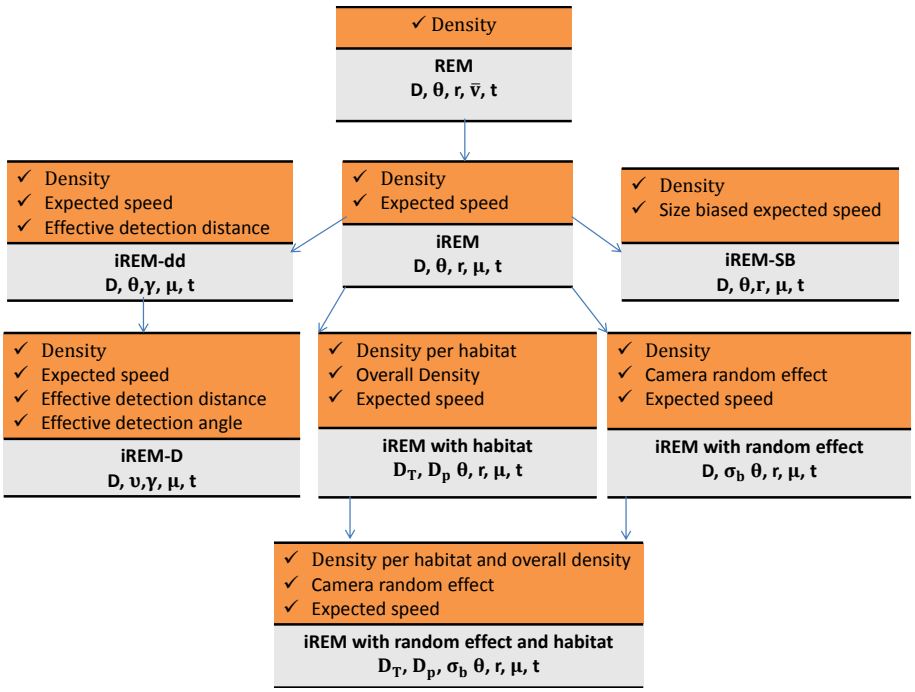


Figure 8.0.1: Summary diagram of the extended REM models developed in this thesis. Parameters are as defined in the corresponding Chapters. REM is given in Chapter 2, iREM is given in Chapter 3, whereas iREM with habitat, iREM with random effect, and iREM with random effect and habitat are given in Chapter 4. The models, iREM-dd and iREM-D are given in Chapter 5 and finally, iREM-SB is given in Chapter 7.

As described in Chapter 2 Section 2.7, WWAP data were collected from separate sources, therefore, they are considered statistically independent, so their contribution to the likelihood could be multiplied. We showed in Chapter 2 that REM, although a nonparamtric method, could be derived by assuming a Poisson distribution for the encounter data with a maximum likelihood estimate. We also showed that the maximum likelihood estimation method can provide meaningful results of the density. Of particular note is the

flexibility of the maximum likelihood framework to incorporate heterogeneous habitats where animals will use some habitats more than others. This variation in space could vary with time, particularly time of day, and we have shown that modelling this variation can be relevant when estimating animal density. There are some limitations of REM in estimating density and its precision, which were given in detail in the discussion of Chapter 2. We have pointed out that REM would give accurate estimates of the density but it would underestimate the standard error, since the sampling variability in the speed information is not accounted for. And, using the direct method and the inverse Hessian matrix for variance estimation may bias the estimated standard error for finite (small) sample sizes since these rely on asymptotic theory. However, a nonparametric bootstrap method or an adjusted variance method for ratio estimation would be useful in accounting for some of the variability that comes from using an estimate of the mean speed. These methods are advantageous since they do not assume any distribution for the data. However, there are some limitations of these methods. For the bootstrap method the most important limitation is the assumption that the distribution of the data represented by the sample is a reasonable estimate of the population function. If this assumption is violated the random sampling performed in the bootstrap procedure may add another level of sampling error, resulting in invalid statistical estimations (Haukoos and Lewis, 2005). In the WWAP survey the placement strategy of camera traps was random in all habitats, except in Central Park where the mara species inhabited, traps were not baited nor set on trails, and animals were arbitrarily selected to be followed around to gather the speed information. As discussed in Chapter 2, the case of the mara species is unique as the area surveyed (WWAP) is a park that is commonly traversed by humans, therefore the underestimation of the density is a direct result of the poor survey design adopted to avoid crowding the cameras with human photographs. Aside from this, there is no evidence to suggest that the data represented by the sample is not a reasonable estimate of the population function. Another limitation of the bootstrap is the smaller the original sample the less likely it is to represent the entire population, thus the more difficult it becomes to compute valid confidence intervals. Note that the bootstrap relies heavily on the tails of the estimated sampling distribution when computing confidence intervals, and using small samples may jeopardize the validity of this computation. In our WWAP data set the speed data, for example, is rather limited (10

observations for each species) and as shown by the results in Chapter 2, the confidence intervals from the direct method and the inverse of the negative Hessian is narrow and does not contain the density from the census, particularly when the variation in habitat type is not incorporated into the model. Finally, the sampling distribution of the bootstrapped statistics is frequently not symmetric as is the case of our estimates and the median value seemed to be a more valid estimate. But there is no simple method for calculating the 95% confidence interval for the median, and it is not valid to use a 95% confidence interval calculated from the standard error to represent the 95% confidence interval for the median value, unless the distribution of the underlying data is normal (see Haukoos and Lewis, 2005) and even so the mean and median would have different standard errors. For the adjusted variance method, the variance of the ratio is only asymptotically unbiased, also if the mean of the denominator is small then the variance of the ratio will be large. For the WWAP data set the average speed of movement for three of the four species is quite small, hence, a large estimate of the variance, for example the wallaby species. Also, it is important to note that when the population of the realizations of the random variables is heterogeneous, the variance estimator will mask such heterogeneity - for instance, the huge $C_v\%$ of the speed and encounter data for the species at WWAP (see Van Kempen and Van Vliet, 2000). Finally, it is difficult to adopt the adjusted variance method for variance estimation in more complex models, such as the REM with habitat or random effects. However, the maximum likelihood REM framework developed in this thesis is flexible and can be extended to incorporate multiple sources of data such as animal speed, detection distance and detection angle into one coherent framework, as we have shown in Chapters 3, 4, 5, 6, and 7 allowing for new analyses to be possible. It is worth reiterating here that the REM maximum likelihood framework developed in Chapter 2 does not account for the sampling variability in animal speed, and given the fact that the proposed variance estimation methods can be unreliable we developed an integrated Random Encounter Model (iREM) to account for the sampling variability in the data and variance estimation.

The integrated Random Model (iREM), which serves as a basis for further development of a more unified framework for density estimation, was developed in Chapter 3 to answer the second question. We have considered a variety of approaches for density

estimation in iREM. These different methods were designed to be broadly applicable to various types of species, each with its associated advantages and disadvantages. Given the variation between different animal species, we anticipate that the best modelling approach may vary according to individual species, as well as data available and study purpose. In this Chapter we have shown that integrating multiple sources of data allows the estimation of not only density but also parameters associated with the density. iREM has been shown to improve estimation accounting for the sampling variability in animal speed. The underlying Poisson model in REM has been shown to work well but the negative binomial REM, zero-inflated Poisson REM and zero-inflated negative binomial REM provide alternative approaches for model fitting for cases of zero-inflation and variation in the encounter data. It is hoped that the methods developed in this Chapter would provide attractive alternatives to modelling camera trap data for density estimation in iREM, and highlight potential issues associated with density estimation if certain features in the data are not accounted for.

In the case of the WWAP data set, our analysis revealed some limitations on the robustness of the abundance estimates from iREM. In this connection, we would like to highlight that, if the data set is large enough iREM can obtain meaningful parameter estimates. However, for species with limited data and if the sampling variability in the data is unaccounted for iREM would perform poorly as seen in Section 3.8.3 Chapter 3 for the muntjac species, but the development of more variable iREM models such as ZINB iREM can correct for this problem. However, if the real focus of the study is to obtain a precise estimation of population abundance, then an iREM that incorporates spatial covariates that can explain additional variation in the density would be more appropriate. In this Chapter we also extended iREM to model group size data but the model is limited to small family groups where counts of the number of individuals in each group is possible. The WWAP survey is a small survey where animals were moving in small family groups. In the discussions of Chapters 2 and 3 we described in detail the limitation of REM and iREM for density estimation of the individual found in groups. If all animals in a group are detected within the detection zone of a camera trap, and a count of the number of individuals in the observed group can be made, then it becomes possible to estimate average group size in the population (Buckland et al.,

2001). While larger groups tend to be more detectable than smaller groups, all animals within the observed group may not be detected given the limited view of a camera trap (i.e., maximum detection angle is $\pi/2$), and a narrow detection zone closer to the trap. Buckland et al. (2001) argued that estimation of average group size requires the location of the centre of observed groups and the number of individuals in the group to be recorded. In camera trapping analysis, however, the centre of the group is unobservable and, therefore, it becomes impossible to measure true group size. Also, if a group is replaced by objects of the same size as suggested by Buckland et al. (2001) the issue of violation of the independence assumption would arise, invalidating analytical variance estimation and model selection procedure. A recommendation by Buckland et al. (2001) if groups occur but are not well-defined is for the observer to record each individual animal, and its location, and use robust variance estimation methods. While this thesis offers some practical solutions, with rather strong assumptions, for modelling group size data and estimating average group size from camera trap data, the development of more robust modelling approaches is required. Chapter 3 also highlights the importance of accounting for observed zero speed of movement. We showed that including the observed zero speed would inflate the estimated density (hence, underestimation if unaccounted for), particularly for small sample sizes as is the case of our WWAP data set. We, therefore, recommend where possible to sample a large number of observations of the speed when there are observed zeros, especially when the density is large. The current iREM framework has the possibility of using zero inflated distributions to model the observed zero speeds, however, another possibility would be to actually model the process of animals not moving using a moving-resting process model as is done in Yan et al. (2014), for example. This approach requires further exploration, and also remains an avenue for future research.

In Chapter 4 we explored the flexibility of iREM developed in Chapter 3 by incorporating covariates such as habitat type and random effects of the camera location. This new model can provide insights into the effect of spatial variation and unobserved heterogeneity of camera location on abundance estimation. The estimation of abundance for species that cannot be identified to the individual level from camera trap data has received limited attention, hence, from the iREM with covariates model density can be

estimated for various species, and also compared within and across habitats. Our model tackles the problem posed in question (iii) regarding abundance estimation in heterogeneous habitats. We have illustrated how not accounting for the variation in habitat type can result in poor estimation of the density. However, the iREM developed for random effect and for both random effect and habitat can be further refined for efficiency. First of all, the numerical method adopted to approximate the intractable integral requires a large number of quadrature points to obtain a close approximation to the likelihood. Consequently, the method can be computationally intensive as we have experienced in the simulations and analysis of the WWAP data set, and particularly, if there are a large number of random effects (Crouch and Spiegelman, 1990; Rabe-Hesketh et al., 2002). The approximation can also be poor if the responses are conditionally Poisson distributed (Albert and Follmann, 2000) as can be seen, for example, in the case of the water deer species. Moving to simpler approximation methods such as an adaptive Gauss quadrature, which uses a single quadrature point (mode) (Lesaffre and Spiessens, 2001) or a simple matrix transformation (Crouch and Spiegelman, 1990), or a Bayesian framework (King et al., 2016; Reynolds et al., 2009), which may be more straightforward are avenues for further research. Secondly, the results of the more complicated iREM with random effect and habitat model suggested that the model is near parameter redundant, which behaves poorly in terms of the maximum likelihood estimates and relevant standard errors. Therefore, models to address the issue of near parameter redundancy in this type of study is an avenue for further research.

To tackle the problems posed in questions (iv) and (v) we explored a more general model that accounts for the sampling variability in animal speed and the detection zone dimensions (iREM-D). Our approach represents an attractive alternative to Rowcliffe et al. (2008) REM, which requires five independent steps to derive all the necessary components of the analysis (encounter rate, speed, activity level, detection distance and detection angle). Our model provides a description that incorporates the steps, into a single integrated framework, which allows the estimation of not only density but also of the independent parameters required to estimate density. This is, as far as we know, the first comprehensive model, which considers a functional relationship between a camera index and animal density (Jennelle et al., 2002) of unmarked animals, that

accounts for the sampling variability in the independent parameters required for density estimation. However, one feature for further consideration is the influence of activity level on estimated density, which is not currently accounted for by the single integrated model (iREM-D), since animals that are more active are more likely to be captured by camera traps. In Chapter 5 it was shown that the underlying Poisson model in REM produces meaningful estimates of the density but the negative binomial model provides an alternative approach, particularly to account for variation in the encounters. While a ZIP model or a ZINB model are, theoretically, other approaches to account for zero-inflation and additional variation, their application to this type of data is an avenue for further research. It is hoped that the results in this Chapter can inform guidance on minimum and optimal survey effort required to achieve a given power in detecting density difference in time or space, and highlight the importance of accounting for sampling variability in the parameters required for density estimation.

It is unfortunately not only within a given model structure that inferential problems arise. Since the density is a functional relationship with the speed of movement, and given that an average speed is required in the estimation process, this average is likely to be biased towards faster moving species. In connection with this, the size biased models in Chapter 6 were proposed to account for the bias in the speed of faster moving animals. We have illustrated how not accounting for this bias, unless where the sampling variability of speed data is low, can be problematic in terms of induced bias in the average speed, and consequently the density estimator. We proposed models for the probability of encountering animals with given speeds, and showed that this probability is approximately proportional to its speed. We considered the underlying Poisson model in REM, and other alternative models including a negative binomial, a zero-inflated Poisson and zero-inflated negative binomial from which the probability of encounter could be derived. And, we considered nonnegative probability density functions such as a gamma model, a lognormal model and a Weibull model for animal speed. We showed that the approximation method for the probability of encounter for size biased sampling works for all models, but we have only derived the size biased gamma model with the true probability of encounter from a Poisson, which was proven to work well. The derivation of size biased models where the true probability of encounter comes from a negative

binomial or a zero-inflated Poisson or a zero-inflated negative binomial is an area for further research, since these cannot be achieved easily by analytical methods.

The size biased models proposed in Chapter 6 for the speed data were also relevant in the development of models suitable for describing population abundance of unmarked species in Chapter 7. We have shown how induced-bias in the average speed from faster moving animals can bias the density estimator. A key challenge when using integrated population models to describe demographic variables in ecological studies is the independence assumption of the data sets. Generally, this assumption is ignored or violated, and this violation can have some consequence on the parameters (Besbeas et al., 2002). In Chapter 7 we considered two approaches: 1) using dependent sources, and 2) using independent sources for modelling abundance of unmarked species. These approaches were designed to determine whether this assumption, if violated, would have an affect on the density estimator. Despite the strong degree of dependence in the data sources there was only minor consequences on the parameter estimations and precision.

There are some important comments, worth reiterating, regarding the use of REMs developed in this thesis. First and foremost, REMs require randomized placements of camera traps. If traps are placed on trails, for example, or not placed randomly in the surveyed area, this would violate the underpinning assumption of REMs resulting in biased estimates of the density. A classical example of this is the estimates of the density of the mara species from REMs. Secondly, random encounters of animals with camera traps are necessary for density estimation. Thus, REMs do not allow baiting of traps or luring of animals to traps in any way. The data required for REMs must be in the form of numbers of independent contacts between animal (individual or group) and camera. One way to acquire this is to set camera traps to become inactive for 2 minutes after each photograph. This allows for an animal or a large group of animals to leave the camera detection zone, and that same animal (or group) or a different animal (or group) later re-enters to give a second independent contact. From testing and implementing REMs in this thesis we showed the importance of, and recommend the following when using REMs:

1. Modelling the speed data - this resulted in improved precision since the variability

is accounted for.

2. Accounting for the variation in the encounter data - estimated density and its precision improve when this variability is accounted for, e.g., use of NB REM and ZINB REM.

3. Incorporating measurable factors such as habitat type/land cover or random effects in the modelling process - this is necessary in estimating the density as these would explain some of the variation in the density.

4. Accounting for the variability in the detection zone dimensions (whether low or high variability) - this resulted in improved estimates of the density and precision.

5. Accounting for the bias introduced by faster moving animals - this is necessary regardless if the variability in the speed data is low. The coefficient of variation of the estimated average speed for the BCI data is as low as 3% as shown in Chapter 6 but the bias in the density introduced from faster moving animals is substantial as shown in Chapter 7.

All analyses in this thesis were made using maximum likelihood estimations. But many of these modelling suggestions may be most feasible within a Bayesian framework, which can readily incorporate hierarchical models, and fitting of complex models, which include random effects, for example, may be more straightforward. Also, if there are some prior beliefs by ecologists about demographic parameters a Bayesian approach might be more straightforward. Reynolds et al. (2009) for example developed an integrated Bayesian analysis framework that uses valuable prior information about the model parameters. Reynolds et al. (2009) showed that it is feasible to obtain reasonable estimates, including parameters that were otherwise inestimable if the data sources were analysed separately from this comprehensive process model.

Although we have provided some of the associated R code within this thesis, and the appendix of the thesis, the possibility of developing a free, easy to use, statistical package in R, which incorporates general frameworks for abundance estimation of unmarked animals, could encourage the wider application of these methods by producing more accessible tools for users.

Bibliography

Abadi, F., Gimenez, O., Arlettaz, R., and Schaub, M. (2010). An assessment of integrated population models: bias, accuracy, and violation of the assumption of independence. *Ecology*, 91(1):7–14.

Akaike, H. (1992). Information theory and an extension of the maximum likelihood principle. In *Breakthroughs in statistics*, pages 610–624. Springer.

Albert, P. S. and Follmann, D. A. (2000). Modeling repeated count data subject to informative dropout. *Biometrics*, 56(3):667–677.

Arratia, R. and Goldstein, L. (2010). Size bias, sampling, the waiting time paradox, and infinite divisibility: when is the increment independent? *arXiv preprint arXiv:1007.3910*.

Bartmann, R. M., White, G. C., Carpenter, L. H., and Garrott, R. A. (1987). Aerial mark-recapture estimates of confined mule deer in pinyon-juniper woodland. *The Journal of wildlife management*, pages 41–46.

Bedrick, E. J. and Tsai, C.-L. (1994). Model selection for multivariate regression in small samples. *Biometrics*, pages 226–231.

Besbeas, P., Borysiewicz, R. S., and Morgan, B. J. (2009). Completing the ecological jigsaw. In *Modeling demographic processes in marked populations*, pages 513–539. Springer.

Besbeas, P., Freeman, S. N., and Morgan, B. J. (2005). The potential of integrated population modelling. *Australian & New Zealand Journal of Statistics*, 47(1):35–48.

Besbeas, P., Freeman, S. N., Morgan, B. J., and Catchpole, E. A. (2002). Integrating

mark–recapture–recovery and census data to estimate animal abundance and demographic parameters. *Biometrics*, 58(3):540–547.

Besbeas, P., Lebreton, J.-D., and Morgan, B. J. T. (2003). The efficient integration of abundance and demographic data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 52(1):95–102.

Borchers, D., Distiller, G., Foster, R., Harmsen, B., and 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. L., Buckland, S. T., and Zucchini, W. (2002). *Estimating animal abundance: closed populations*, volume 13. Springer Science & Business Media.

Bowden, D. C. and Kufeld, R. C. (1995). Generalized mark-sight population size estimation applied to colorado moose. *The Journal of wildlife management*, pages 840–851.

Bridges, A. S., Vaughan, M. R., and Klenzendorf, S. (2004). Seasonal variation in american black bear ursus americanus activity patterns: quantification via remote photography. *Wildlife Biology*, 10(4):277–284.

Brown, C. E. (1998). Coefficient of variation. In *Applied multivariate statistics in geohydrology and related sciences*, pages 155–157. Springer.

Buckland, S. T. (1992). Fitting density functions with polynomials. *Applied Statistics*, pages 63–76.

Buckland, S. T., Anderson, D. R., Burnham, K. P., Laake, J. L., Borchers, D., and Thomas, L. (2001). Introduction to distance sampling estimating abundance of biological populations.

Buckland, S. T., Burnham, K. P., and Augustin, N. H. (1997). Model selection: an integral part of inference. *Biometrics*, pages 603–618.

Burnham, K. P. and Anderson, D. R. (2004). Multimodel inference understanding aic and bic in model selection. *Sociological methods & research*, 33(2):261–304.

Burton, A. C., Neilson, E., Moreira, D., Ladle, A., Steenweg, R., Fisher, J. T., Bayne, E., and Boutin, S. (2015). Wildlife camera trapping: a review and recommendations for linking surveys to ecological processes. *Journal of Applied Ecology*.

Carbone, C., Christie, S., Conforti, K., Coulson, T., Franklin, N., Ginsberg, J., Griffiths, M., Holden, J., Kawanishi, K., Kinnaird, M., et al. (2001). The use of photographic rates to estimate densities of tigers and other cryptic mammals. *Animal Conservation*, 4(01):75–79.

Carpenter, J. and Bithell, J. (2000). Bootstrap con" dence intervals: when, which, what? a practical guide for medical statisticians.

Catchpole, E. A., Kgosi, P., and Morgan, B. J. (2001). On the near-singularity of models for animal recovery data. *Biometrics*, 57(3):720–726.

Chandler, R. B., Royle, J. A., et al. (2013). Spatially explicit models for inference about density in unmarked or partially marked populations. *The Annals of Applied Statistics*, 7(2):936–954.

Chen, Y. Q. (2010). Semiparametric regression in size-biased sampling. *Biometrics*, 66(1):149–158.

Choquet, R. and Cole, D. J. (2012). A hybrid symbolic-numerical method for determining model structure. *Mathematical Biosciences*, 236(2):117–125.

Cole, D. J. and McCrea, R. S. (2016). Parameter redundancy in discrete state-space and integrated models. *Biometrical Journal*.

Cole, D. J. and Morgan, B. J. (2010). Parameter redundancy with covariates. *Biometrika*, page asq041.

Cole, D. J., Morgan, B. J., McCrea, R. S., Pradel, R., Gimenez, O., and Choquet, R. (2014). Does your species have memory? analyzing capture–recapture data with memory models. *Ecology and evolution*, 4(11):2124–2133.

Cole, D. J., Morgan, B. J., and Ridout, M. S. (2003). Generalized linear mixed models for strawberry inflorescence data. *Statistical Modelling*, 3(4):273–290.

Cole, D. J., Morgan, B. J., and Titterington, D. (2010). Determining the parametric structure of models. *Mathematical biosciences*, 228(1):16–30.

Cox, D. R. (1983). Some remarks on overdispersion. *Biometrika*, 70(1):269–274.

Crouch, E. A. and Spiegelman, D. (1990). The evaluation of integrals of the form $\int +\infty$- ∞ f (t) exp (- t 2) dt: Application to logistic-normal models. *Journal of the American Statistical Association*, 85(410):464–469.

Cutler, T. L. and Swann, D. E. (1999). Using remote photography in wildlife ecology: a review. *Wildlife Society Bulletin*, pages 571–581.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap methods and their application*, volume 1. Cambridge university press.

Dey, S., Dey, T., and Anis, M. (2015). Weighted weibull distribution: Properties and estimation. *Journal of Statistical Theory and Practice*, 9(2):250–265.

Dixon, K. R. and Chapman, J. A. (1980). Harmonic mean measure of animal activity areas. *Ecology*, 61(5):1040–1044.

du Preez, B. D., Loveridge, A. J., and Macdonald, D. W. (2014). To bait or not to bait: a comparison of camera-trapping methods for estimating leopard panthera pardus density. *Biological Conservation*, 176:153–161.

Eberhardt, L. (1968). A preliminary appraisal of line transects. *The Journal of Wildlife Management*, pages 82–88.

Foster, R. J. and Harmsen, B. J. (2012). A critique of density estimation from camera-trap data. *The Journal of Wildlife Management*, 76(2):224–236.

Fournier, D. and Archibald, C. P. (1982). A general theory for analyzing catch at age data. *Canadian Journal of Fisheries and Aquatic Sciences*, 39(8):1195–1207.

Fox, J. (2002). Bootstrapping regression models. *An R and S-PLUS Companion to Applied Regression: A Web Appendix to the Book. Sage, Thousand Oaks, CA. URL http://cran. r-project. org/doc/contrib/Fox-Companion/appendix-bootstrapping. pdf.*

Fuller, T. K., York, E. C., Powell, S. M., Decker, T. A., and DeGraaf, R. M. (2001). An evaluation of territory mapping to estimate fisher density. *Canadian Journal of Zoology*, 79(9):1691–1696.

Gates, C. E., Marshall, W. H., and Olson, D. P. (1968). Line transect method of estimating grouse population densities. *Biometrics*, pages 135–145.

Gschlößl, S. and Czado, C. (2008). Modelling count data with overdispersion and spatial effects. *Statistical papers*, 49(3):531–552.

Hall, P. (2013). *The bootstrap and Edgeworth expansion*. Springer Science & Business Media.

Harmsen, B. J. (2006). *The use of camera traps for estimating abundance and studying the ecology of Jaguars*. PhD thesis, University of Southampton.

Haukoos, J. S. and Lewis, R. J. (2005). Advanced statistics: bootstrapping confidence intervals for statistics with "difficult" distributions. *Academic emergency medicine*, 12(4):360–365.

Hilbe, J. M. (2011). *Negative binomial regression*. Cambridge University Press.

Hutchinson, J. and Waser, P. M. (2007). Use, misuse and extensions of "ideal gas" models of animal encounter. *Biological Reviews*, 82(3):335–359.

Ismail, N. and Zamani, H. (2013). Estimation of claim count data using negative binomial, generalized poisson, zero-inflated negative binomial and zero-inflated generalized poisson regression models. In *Casualty Actuarial Society E-Forum*, volume 41, pages 1–28.

Jennelle, C. S., Runge, M. C., and MacKenzie, D. I. (2002). The use of photographic rates to estimate densities of tigers and other cryptic mammals: a comment on misleading conclusions. *Animal Conservation*, 5(2):119–120.

Karanth, K. and Nichols, J. (2002). Monitoring tigers and their prey. *Centre for Wildlife Studies, Bangalore, India*.

Karanth, K. U. (1995). Estimating tiger panthera tigris populations from camera-trap data using capture—recapture models. *Biological conservation*, 71(3):333–338.

Karanth, K. U., Chundawat, R. S., Nichols, J. D., and Kumar, N. (2004). Estimation of tiger densities in the tropical dry forests of panna, central india, using photographic capture–recapture sampling. *Animal Conservation*, 7(03):285–290.

Karanth, K. U. and Nichols, J. D. (1998). Estimation of tiger densities in india using photographic captures and recaptures. *Ecology*, 79(8):2852–2862.

Kelly, M. J., Noss, A. J., Di Bitetti, M. S., Maffei, L., Arispe, R. L., Paviolo, A., De Angelo, C. D., and Di Blanco, Y. E. (2008). Estimating puma densities from camera trapping across three study sites: Bolivia, argentina, and belize. *Journal of Mammalogy*, 89(2):408–418.

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.

Lam, F., Hung, C., and Perrier, D. (1985). Estimation of variance for harmonic mean half-lives. *Journal of pharmaceutical sciences*, 74(2):229–231.

Lesaffre, E. and Spiessens, B. (2001). On the effect of the number of quadrature points in a logistic random effects model: an example. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 50(3):325–335.

Linkie, M., Dinata, Y., Nugroho, A., and Haidir, I. A. (2007). Estimating occupancy of a data deficient mammalian species living in tropical rainforests: sun bears in the kerinci seblat region, sumatra. *Biological Conservation*, 137(1):20–27.

Lucas, T. C., Moorcroft, E. A., Freeman, R., Rowcliffe, J. M., and Jones, K. E. (2015). A generalised random encounter model for estimating animal density with remote sensor data. *Methods in Ecology and Evolution*, 6(5):500–509.

Maffei, L., Cuéllar, E., and Noss, A. (2004). One thousand jaguars (panthera onca) in bolivia's chaco? camera trapping in the kaa-iya national park. *Journal of Zoology*, 262(3):295–304.

Marques, T., Buckland, S., Borchers, D., Tosh, D., and McDonald, R. (2010). Point transect sampling along linear features. *Biometrics*, 66(4):1247–1255.

Matthews, S. M., Golightly, R. T., and Higley, J. M. (2008). Mark-resight density estimation for american black bears in hoopa, california. *Ursus*, 19(1):13–21.

Maxwell, J. C. (1860). V. illustrations of the dynamical theory of gases.—part i. on the motions and collisions of perfectly elastic spheres. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 19(124):19–32.

Mazzolli, M. (2010). Mosaics of exotic forest plantations and native forests as habitat of pumas. *Environmental Management*, 46(2):237–253.

McCallum, J. (2013). Changing use of camera traps in mammalian field research: habitats, taxa and study types. *Mammal Review*, 43(3):196–206.

McClintock, B. T., White, G. C., Antolin, M. F., and Tripp, D. W. (2009). Estimating abundance using mark–resight when sampling is with replacement or the number of marked individuals is unknown. *Biometrics*, 65(1):237–246.

McCrea, R. S. and Morgan, B. J. (2014). *Analysis of capture-recapture data*. CRC Press.

Mir, K., Ahmed, A., and Reshi, J. (2013). On size baised gamma distribution. *Journal of Modern Mathematics and Statistics*, 7(5):55–57.

Newman, K., Buckland, S., Morgan, B., King, R., Borchers, D., Cole, D., Besbeas, P., Gimenez, O., and Thomas, L. (2014). Integrated population modelling. In *Modelling Population Dynamics*, pages 169–195. Springer.

Noss, A., Cuéllar, R., Barrientos, J., Maffei, L., Cuéllar, E., Arispe, R., Rúmiz, D., and Rivero, K. (2003). A camera trapping and radio telemetry study of lowland tapir (tapirus terrestris) in bolivian dry forests. *Tapir conservation*, 12(1):24–32.

O'Brien, T. G., Kinnaird, M. F., and Wibisono, H. T. (2003). Crouching tigers, hidden prey: Sumatran tiger and prey populations in a tropical forest landscape. *Animal Conservation*, 6(02):131–139.

O'Connell, A. F., Nichols, J. D., and Karanth, K. U. (2010). *Camera traps in animal ecology: methods and analyses*. Springer Science & Business Media.

Oliveira-Santos, L. G. R., Tortato, M. A., and Graipel, M. E. (2008). Activity pattern of atlantic forest small arboreal mammals as revealed by camera traps. *Journal of Tropical Ecology*, 24(05):563–567.

Patil, G. P. and Rao, C. R. (1978). Weighted distributions and size-biased sampling with applications to wildlife populations and human families. *Biometrics*, pages 179–189.

Pewsey, A., Neuhäuser, M., and Ruxton, G. D. (2013). *Circular statistics in R*. Oxford University Press.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rabe-Hesketh, S., Skrondal, A., Pickles, A., et al. (2002). Reliable estimation of generalized linear mixed models using adaptive quadrature. *The Stata Journal*, 2(1):1–21.

Ratnaparkhi, M. V. and Naik-Nimbalkar, U. V. (2012). The length-biased lognormal distribution and its application in the analysis of data from oil field exploration studies. *Journal of Modern Applied Statistical Methods*, 11(1):22.

Reynolds, T. J., King, R., Harwood, J., Frederiksen, M., Harris, M. P., and Wanless, S. (2009). Integrated data analysis in the presence of emigration and mark loss. *Journal of agricultural, biological, and environmental statistics*, 14(4):411–431.

Ridout, M. S. and Linkie, M. (2009). Estimating overlap of daily activity patterns from camera trap data. *Journal of Agricultural, Biological, and Environmental Statistics*, 14(3):322–337.

Roberts, N. J. (2011). Investigation into survey techniques of large mammals: surveyor competence and camera-trapping vs. transect-sampling. *Bioscience Horizons*, 4(1):40–49.

Rothman, K. J. (2012). *Epidemiology: an introduction*. Oxford University Press.

Rovero, F. and De Luca, D. W. (2007). Checklist of mammals of the udzungwa mountains of tanzania. *Mammalia*, 71(1/2):47–55.

Rovero, F., Zimmermann, F., Berzi, D., and Meek, P. (2013). " which camera trap type and how many do i need?" a review of camera features and study designs for a

range of wildlife research applications. *Hystrix, the Italian Journal of Mammalogy*, 24(2):148–156.

Rowcliffe, J. M., Cowlishaw, G., and Long, J. (2003). A model of human hunting impacts in multi-prey communities. *Journal of applied ecology*, 40(5):872–889.

Rowcliffe, J. M., Field, J., Turvey, S. T., and Carbone, C. (2008). Estimating animal density using camera traps without the need for individual recognition. *Journal of Applied Ecology*, 45(4):1228–1236.

Rowcliffe, J. M., Jansen, P. A., Kays, R., Kranstauber, B., and Carbone, C. (2016). Wildlife speed cameras: measuring animal travel speed and day range using camera traps. *Remote Sensing in Ecology and Conservation*, 2(2):84–94.

Rowcliffe, M., Carbone, C., Jansen, P. A., Kays, R., and Kranstauber, B. (2011). Quantifying the sensitivity of camera traps: an adapted distance sampling approach. *Methods in Ecology and Evolution*, 2(5):464–476.

Royle, J. A. (2011). Hierarchical spatial capture–recapture models for estimating density from trapping arrays. In *Camera Traps in Animal Ecology*, pages 163–190. Springer.

Royle, J. A., Chandler, R. B., Sollmann, R., and Gardner, B. (2013). *Spatial capture-recapture*. Academic Press.

Royle, J. A., Dawson, D. K., and Bates, S. (2004). Modeling abundance effects in distance sampling. *Ecology*, 85(6):1591–1597.

Royle, J. A., Nichols, J. D., Karanth, K. U., and Gopalaswamy, A. M. (2009). A hierarchical model for estimating density in camera-trap studies. *Journal of Applied Ecology*, 46(1):118–127.

Samejima, H., Ong, R., Lagan, P., and Kitayama, K. (2012). Camera-trapping rates of mammals and birds in a bornean tropical rainforest under sustainable forest management. *Forest Ecology and Management*, 270:248–256.

Sanderson, J. G. and Trolle, M. (2005). Monitoring elusive mammals. *American Scientist*, 93(2):148–155.

Schaub, M., Gimenez, O., Sierro, A., and Arlettaz, R. (2007). Use of integrated modeling to enhance estimates of population dynamics obtained from limited data. *Conservation Biology*, 21(4):945–955.

Silver, S. C., Ostro, L. E., Marsh, L. K., Maffei, L., Noss, A. J., Kelly, M. J., Wallace, R. B., Gómez, H., and Ayala, G. (2004). The use of camera traps for estimating jaguar panthera onca abundance and density using capture/recapture analysis. *Oryx*, 38(02):148–154.

Smith, M., Gardner, R., Gentry, J., Kaufman, D., and O'farrell, M. (1975). Density estimations of small mammal populations. *Small mammals: their productivity and population dynamics. Cambridge University Press, London*, pages 25–54.

Swinnen, K. R., Reijniers, J., Breno, M., and Leirs, H. (2014). A novel method to reduce time investment when processing videos from camera trap studies. *PloS one*, 9(6):e98881.

Thorn, M., Scott, D. M., Green, M., Bateman, P. W., and Cameron, E. Z. (2009). Estimating brown hyaena occupancy using baited camera traps. *South African Journal of Wildlife Research*, 39(1):1–10.

Tobler, M. W., Carrillo-Percastegui, S. E., and Powell, G. (2009). Habitat use, activity patterns and use of mineral licks by five species of ungulate in south-eastern peru. *Journal of Tropical Ecology*, 25(03):261–270.

Trolle, M. and Kéry, M. (2003). Estimation of ocelot density in the pantanal using capture-recapture analysis of camera-trapping data. *Journal of mammalogy*, 84(2):607–614.

Trolle, M. and Kéry, M. (2005). Camera-trap study of ocelot and other secretive mammals in the northern pantanal. *Mammalia mamm*, 69(3-4):409–416.

Trolle, M., Noss, A. J., Cordeiro, J. L. P., and Oliveira, L. F. B. (2008). Brazilian tapir density in the pantanal: A comparison of systematic camera-trapping and line-transect surveys. *Biotropica*, 40(2):211–217.

Trolle, M., Noss, A. J., Lima, E. D. S., and Dalponte, J. C. (2007). Camera-trap studies of maned wolf density in the cerrado and the pantanal of brazil. *Biodiversity and Conservation*, 16(4):1197–1204.

Vaida, F. and Blanchard, S. (2005). Conditional akaike information for mixed-effects models. *Biometrika*, 92(2):351–370.

Van Kempen, G. and Van Vliet, L. (2000). Mean and variance of ratio estimators used in fluorescence ratio imaging. *Cytometry Part A*, 39(4):300–305.

Watts, D. E., Parker, I. D., Lopez, R. R., Silvy, N. J., and Davis, D. S. (2008). Distribution and abundance of endangered florida key deer on outer islands. *The Journal of Wildlife Management*, 72(2):360–366.

Winkelmann, R. (2008). *Econometric analysis of count data*. Springer Science & Business Media.

Yan, J., Chen, Y.-w., Lawrence-Apfel, K., Ortega, I. M., Pozdnyakov, V., Williams, S., and Meyer, T. (2014). A moving–resting process with an embedded brownian motion for animal movements. *Population ecology*, 56(2):401–415.

Yapp, W. (1956). The theory of line transects. *Bird study*, 3(2):93–104.

Zhao, X. (2009). *A Numerical Investigation of the Effects of Loading Conditions on Soil Response*. ProQuest.

Appendix A

A.1 Maximum Likelihood Estimation of the Density in REM

Here we provide an example of the the R-code for the log-likelihood function above used to estimate density.

```
###loglikelihood function to estimate density in REM
###y is the encounter data
densitys <- 468      #starting value for density
r        <- 0.012    #fixed detection distance
theta    <- 0.175    #fixed detection angle
v.bar    <- 0.71     #fixed mean speed from speed data
t        <- 1        #camera trap time period

loglike <- function(par,y){
         Density <- exp(par)
         lambda  <- ((2+theta)/pi)*r*t*v.bar*Density
         -sum(dpois(y, lambda, log=TRUE))}

###minimizing the negative loglikelihood using optim
max.like <- optim(f=loglike,log(densitys),method="Brent",
            lower=1,upper=1000,y=y,hessian=TRUE)
```

A.2 Density split by habitat in REM

This section is linked to Section 2.8.3 where the estimated density for wallaby, water deer, mara and muntjac species are given. Here we give the estimates of the density split by habitat for the wallaby and water deer species. Table A.2.1 gives the results of the wallaby species, while Table A.2.2 gives the results of the water deer species. For the wallaby species the density from the census in three of the four habitats is captured within an approximate 95% confidence interval from the Hessian matrix and a 95% confidence interval using bootstrap, but the bootstrap confidence interval captures the mean density from the census. For the water deer species, the density from the census in only two of the four habitats, and the mean density are captured within a 95% confidence interval.

Table A.2.1: Density from the census of the wallaby species compared with estimated density. The estimated standard error of the density, and estimated standard error on the log-scale are given. We give 95% Confidence Intervals (CI) from the inverse of the negative Hessian matrix and Bootstrap method, using the standard errors on the log-scale are given.

	Habitat	Census Density	Estimate	Bootstrap estimates		Hessian errors		Bootstrap errors		Hessian	Bootstrap
		D (in km^2)	$\hat{D}$	(average) $\hat{D}$	(median) $\hat{D}$	$Se(\hat{D})$	$Se\{\log(\hat{D})\}$	$Se(\hat{D})$	$Se\{\log(\hat{D})\}$	CI	CI
	Downs	1101	834	1118	875	50.87	0.06	766.54	0.56	(740.39, 940.28)	(385.35, 2923.85)
	Institute Paddock	760	1117	1571	1261	81.82	0.07	982.74	0.55	(1027.20, 1348.91)	(519.86, 4219.98)
	Old Farm	803	809	1049	860	62.80	0.08	618.80	0.53	(694.93, 942.06)	(367.99, 2611.61)
	Central Park	96	334	444	356	48.73	0.15	347.86	0.63	(251.01, 444.65)	(109.18, 1289.04)
Total		468	595	791	632	35142.29	41.80	663564.70	478.70	(518.64, 683.03)	(241.79, 2589.91)

Table A.2.2: Density from the census of the water deer species compared with estimated density. The estimated standard error of the density, and estimated standard error on the log-scale are given. We give 95% Confidence Intervals (CI) from the inverse of the negative Hessian matrix and Bootstrap method, using the standard errors on the log-scale are given.

Habitat	Census Density D (in km^2)	Estimate $\hat{D}$	Bootstrap estimates (average) $\hat{D}$	(median) $\hat{D}$	Hessian errors $Se(\hat{D})$	$Se\{\log(\hat{D})\}$	Bootstrap errors $Se(\hat{D})$	$Se\{\log(\hat{D})\}$	Hessian CI	Bootstrap CI
Downs	73	86	101	85	12.66	0.15	65.73	0.48	(64.30, 114.60)	(39.83, 273.53)
Institute Paddock	36	10	15	10	5.93	0.58	61.39	1.05	(3.31, 31.82)	(0.14, 38.52)
Old Farm	577	325	517	331	30.90	0.09	2503.38	0.57	(270.25, 392.06)	(153.24, 1115.44)
Central Park	72	248	304	252	32.57	0.13	224.33	0.49	(191.75, 320.84)	(115.55, 787.13)
Total	119	191	246	193	2649.75	30.45	20596.61	123.23	(139.96, 261.73)	(92.20, 655.61)

Appendix B

B.1 Investigating the fit of the Poisson iREM

In Chapter 3 Section 3.7.1 we test the Poisson iREM under different parameter settings and sample size conditions. In the simulations we investigate the importance of accounting for the variation in animal speed. The results reveal that the Poisson iREM works well in estimating the density and for given sample sizes, although in some cases small, can be used for ecological studies with limited resources. Here we give the parameter estimates from the Poisson iREM, and these are shown in Table B.1.1.

The simulations indicate that precise estimates are obtained for all parameters and there is no difference between the three speed data models used. But large expected animal speeds obtained smaller estimated standard error of the density compared with smaller values of expected speed of movement.

Table B.1.1: Average parameter estimates obtained from fitting a Poisson iREM to encounters simulated from a Poisson REM. The expected speed is set to $\mu = 0.60$ (km/day^{-1}) or $\mu = 4.60$ (km/day^{-1}). The standard errors (in parentheses), Standard deviation (Sd) and Root Mean Square Error (RMSE) are given.

	speed data models					
	gamma		lognormal		Weibull	
	$\hat{D}$	$\hat{\mu}$	$\hat{D}$	$\hat{\mu}$	$\hat{D}$	$\hat{\mu}$
$\mu = 0.60$						
$D = 20$; $n = m = 40$						
estimate	23 (13.66)	0.61 (0.20)	23 (13.27)	0.61 (0.15)	23 (14.52)	0.64 (0.24)
Sd	9.44	0.19	11.85	0.14	11.91	0.25
RMSE	9.79	0.19	12.18	0.14	12.20	0.26
$D = 20$; $n = m = 100$						
estimate	20 (7.94)	0.59 (0.12)	20 (7.37)	0.61 (0.10)	21 (8.36)	0.60 (0.14)
Sd	6.07	0.12	7.15	0.10	7.34	0.14
RMSE	6.09	0.12	7.16	0.10	7.36	0.14
$D = 100$; $n = m = 40$						
estimate	106 (41.69)	0.62 (0.20)	105 (36.06)	0.61 (0.24)	108 (47.40)	0.63 (0.23)
Sd	45.18	0.19	37.16	0.17	50.17	0.25
RMSE	45.53	0.19	37.42	0.17	50.86	0.25
$D = 100$; $n = m = 100$						
estimate	102 (25.46)	0.61 (0.12)	100 (21.68)	0.61 (0.10)	102 (28.26)	0.61 (0.14)
Sd	26.41	0.13	21.17	0.10	26.94	0.14
RMSE	26.47	0.13	21.17	0.10	27.01	0.14
$\mu = 4.60$						
$D = 20$; $n = m = 40$						
estimate	20 (3.84)	4.59 (0.23)	22 (8.25)	4.64 (1.51)	20 (3.82)	4.61 (0.23)
Sd	3.75	0.24	8.14	1.53	3.73	0.24
RMSE	3.74	0.24	8.38	1.53	3.76	0.24
$D = 20$; $n = m = 100$						
estimate	20 (2.38)	4.59 (0.15)	20 (4.90)	4.71 (0.99)	20 (2.38)	4.61 (0.15)
Sd	2.33	0.15	4.90	1.04	2.44	0.15
RMSE	2.33	0.15	4.91	1.05	2.44	0.16
$D = 100$; $n = m = 40$						
estimate	100 (8.23)	4.60 (0.07)	101 (10.68)	4.59 (0.31)	100 (8.23)	4.60 (0.06)
Sd	7.95	0.07	10.28	0.33	7.79	0.06
RMSE	7.79	0.06	10.29	0.33	7.75	0.07
$D = 100$; $n = m = 100$						
estimate	100 (5.21)	4.60 (0.04)	101 (6.78)	4.60 (0.20)	100 (5.21)	4.60 (0.04)
Sd	5.33	0.04	6.88	0.20	5.23	0.05
RMSE	5.35	0.04	6.90	0.20	5.24	0.05

B.1.1 Investigating the importance of accounting for overdispersion in encounter data

In Chapter 3, Section 3.7.2 we investigate the importance of accounting for overdispersion in encounter data. We fit a Poisson iREM to encounters simulated from a NB REM and the results are compared with estimations from fitting a NB iREM to the data. The simulation results reveal that the Poisson iREM can estimate the density well but not accounting for this overdispersion can induce underestimation of the standard error.

Table B.1.2 gives the parameter estimates from the Poisson iREM and Table B.1.3 gives the parameter estimates from a NB iREM.

Table B.1.2: Average parameter estimates obtained from fitting a Poisson iREM to encounters simulated from a NB REM. The expected speed is set to $\mu = 2.60$ (km/day^{-1}) or $\mu = 4.60$ (km/day^{-1}). The standard errors (in parentheses), Standard deviation (Sd) and Root Mean Square Error (RMSE) are given.

	speed data models					
	gamma		lognormal		Weibull	
	$\hat{D}$	$\hat{\mu}$	$\hat{D}$	$\hat{\mu}$	$\hat{D}$	$\hat{\mu}$
$\mu = 2.60$						
$D = 20$; $n = m = 40$						
estimate	21 (5.35)	2.61 (0.13)	21 (5.29)	2.60 (0.05)	21 (5.42)	2.59 (0.13)
Sd	8.31	0.13	8.58	0.06	8.79	0.14
RMSE	8.35	0.13	8.64	0.06	8.86	0.14
$D = 20$; $n = m = 100$						
estimate	20 (3.21)	2.60 (0.08)	21 (3.76)	2.60 (0.25)	20 (3.20)	2.61 (0.0.08)
Sd	5.23	0.09	5.75	0.26	5.25	0.09
RMSE	5.24	0.09	5.79	0.26	5.26	0.09
$D = 100$; $n = m = 40$						
estimate	98 (12.09)	2.60 (0.13)	98 (18.73)	2.63 (0.40)	96 (12.03)	2.60 (0.13)
Sd	31.29	0.13	35.53	0.42	30.34	0.13
RMSE	31.39	0.13	35.57	0.42	30.55	0.13
$D = 100$; $n = m = 100$						
estimate	98 (7.56)	2.60 (0.08)	98 (11.74)	2.63 (0.26)	99 (7.61)	2.59 (0.08)
Sd	22.20	0.09	24.07	0.26	22.72	0.09
RMSE	22.27	0.09	24.16	0.26	22.75	0.09
$\mu = 4.60$						
$D = 20$; $n = m = 40$						
estimate	20 (3.16)	8.62 (0.64)	20 (3.59)	8.71 (0.98)	20 (3.15)	8.64 (0.64)
Sd	6.69	0.68	6.57	1.01	6.66	0.61
RMSE	6.70	0.68	6.57	1.02	6.66	0.61
$D = 20$; $n = m = 100$						
estimate	20 (1.95)	8.57 (0.41)	20 (1.91)	8.56 (0.36)	20 (1.94)	8.62 (0.41)
Sd	4.32	0.42	4.29	0.36	4.33	0.43
RMSE	4.32	0.42	4.30	0.37	4.34	0.43
$D = 100$; $n = m = 40$						
estimate	98 (9.62)	8.60 (0.65)	98 (12.79)	8.64 (0.98)	98 (9.54)	8.60 (0.64)
Sd	33.92	0.67	37.57	0.93	34.24	0.67
RMSE	33.99	0.67	37.60	0.93	34.29	0.67
$D = 100$; $n = m = 100$						
estimate	98 (6.02)	8.59 (0.41)	98 (5.67)	8.62 (0.37)	97 (5.93)	8.63 (0.41)
Sd	18.37	0.42	19.15	0.42	19.33	0.44
RMSE	18.47	0.42	19.24	0.43	19.53	0.44

Table B.1.3: Average parameter estimates obtained from fitting a NB iREM to encounters simulated from a NB REM. For the expected speed, and dispersion parameter we set $\mu = 2.60$ (km/day^{-1}), and $\kappa = 0.25$, respectively. The standard errors, Standard deviation (Sd) and Root Mean Square Error (RMSE) are given.

	gamma			lognormal			Weibull		
	$\hat{D}$	$\hat{\mu}$	$\hat{\kappa}$	$\hat{D}$	$\hat{\mu}$	$\hat{\kappa}$	$\hat{D}$	$\hat{\mu}$	$\hat{\kappa}$
$D = 20$; $n = m = 40$									
estimate	21 (8.51)	2.61 (0.13)	0.38 (0.29)	21 (8.57)	2.60 (0.05)	0.38 (0.28)	21 (8.65)	2.59 (0.13)	0.38 (0.28)
Sd	8.05	0.13	0.50	8.39	0.06	0.43	8.55	0.14	0.47
RMSE	8.10	0.13	0.52	8.46	0.06	0.45	8.64	0.14	0.48
$D = 20$; $n = m = 100$									
estimate	21 (5.25)	2.60 (0.08)	0.28 (0.11)	21 (5.66)	2.60 (0.25)	0.29 (0.13)	21 (5.32)	2.61 (0.08)	0.28 (0.09)
Sd	5.13	0.08	0.10	5.70	0.25	0.11	5.30	0.09	0.09
RMSE	5.16	0.09	0.11	5.77	0.25	0.11	5.36	0.09	0.10
$D = 100$; $n = m = 40$									
estimate	99 (33.12)	2.60 (0.13)	0.30 (0.10)	99 (36.36)	2.66 (0.41)	0.30 (0.10)	99 (32.88)	2.60 (0.13)	0.30 (0.10)
Sd	31.58	0.13	0.11	34.86	0.43	0.11	30.61	0.13	0.11
RMSE	31.59	0.13	0.12	34.87	0.44	0.12	30.64	0.13	0.12
$D = 100$; $n = m = 100$									
estimate	100 (21.54)	2.60 (0.08)	0.26 (0.06)	100 (23.45)	2.63 (0.26)	0.26 (0.06)	101 (21.76)	2.59 (0.08)	0.27 (0.06)
Sd	22.26	0.09	0.06	23.81	0.26	0.05	22.17	0.09	0.05
RMSE	22.26	0.09	0.06	23.81	0.27	0.05	22.21	0.09	0.06

Table B.1.4: Average parameter estimates obtained from fitting a NB iREM to encounters simulated from a NB REM. For the expected speed, and dispersion parameter we set $\mu = 8.60$ (km/day^{-1}), and $\kappa = 0.25$, respectively. The standard errors, Standard deviation (Sd) and Root Mean Square Error (RMSE) are given.

	gamma			lognormal			Weibull		
	$\hat{D}$	$\hat{\mu}$	$\hat{\kappa}$	$\hat{D}$	$\hat{\mu}$	$\hat{\kappa}$	$\hat{D}$	$\hat{\mu}$	$\hat{\kappa}$
$D = 20$; $n = m = 40$									
estimate	21 (7.30)	8.66 (0.64)	0.29 (0.12)	21 (7.55)	8.69 (0.97)	0.29 (0.12)	21 (6.70)	8.65 (0.64)	0.29 (0.12)
Sd	6.46	0.69	0.13	6.52	0.99	0.11	6.30	0.63	0.11
RMSE	6.51	0.69	0.13	6.58	1.00	0.12	6.33	0.63	0.12
$D = 20$; $n = m = 100$									
estimate	20 (4.52)	8.56 (0.40)	0.27 (0.07)	20 (4.47)	8.59 (0.37)	0.28 (0.07)	20 (4.47)	8.60 (0.404)	0.27 (0.07)
Sd	4.17	0.43	0.06	4.15	0.38	0.06	4.22	0.40	0.06
RMSE	4.19	0.44	0.06	4.17	0.38	0.06	4.23	0.40	0.06
$D = 100$; $n = m = 40$									
estimate	100 (33.36)	8.60 (0.64)	0.27 (0.30)	102 (35.13)	8.76 (0.99)	0.27 (0.07)	101 (33.51)	8.57 (0.63)	0.27 (0.07)
Sd	33.65	0.62	0.07	36.00	0.99	0.07	32.56	0.64	0.07
RMSE	33.65	0.62	0.07	36.03	1.00	0.07	32.56	0.64	0.07
$D = 100$; $n = m = 100$									
estimate	100 (20.99)	8.61 (0.41)	0.26 (0.04)	100 (20.78)	8.62 (0.37)	0.26 (0.04)	99 (20.85)	8.59 (0.40)	0.26 (0.04)
Sd	19.98	0.41	0.04	18.52	0.42	0.04	18.52	0.38	0.04
RMSE	19.99	0.41	0.04	18.52	0.42	0.04	18.53	0.38	0.04

B.1.2 Investigating the importance of accounting for zero-inflation in the encounter data

We have shown in Chapter 3, Section 3.7.3 that disregarding zero-inflation in the encounter data can induce a negative bias and an underestimation of the standard error of the density.

Table B.1.5 gives the parameter estimates from fitting a Poisson iREM to encounters simulated from a ZIP REM, and Table B.1.6 gives the parameter estimates from a ZIP iREM where encounters are simulated from the ZIP REM. The simulation results reveal that the models obtained precise estimates of the expected speed but the Poisson iREM underestimates the density, and the true density is not captured within a 95% confidence interval, while the ZIP iREM obtained precise estimates of the density.

Table B.1.5: Average parameter estimates obtained from fitting a Poisson iREM to encounters simulated from a ZIP REM. The expected animal speed is set to $\mu = 4.60$ (km/day^{-1}). The standard errors (in parentheses), Standard deviation (Sd) and Root Mean Square Error (RMSE) are given.

	speed data models					
	gamma		lognormal		Weibull	
	$\hat{D}$	$\hat{\mu}$	$\hat{D}$	$\hat{\mu}$	$\hat{D}$	$\hat{\mu}$
$D = 20$; $n = m = 40$						
estimate	14 (3.41)	4.60 (0.47)	14 (3.26)	4.61 (0.36)	14 (3.39)	4.62 (0.46)
Sd	3.35	0.43	3.15	0.38	3.50	0.48
RMSE	6.92	0.43	6.92	0.38	7.02	0.49
$D = 20$; $n = m = 100$						
estimate	14 (2.13)	4.62 (0.30)	14 (1.97)	4.60 (0.14)	14 (2.14)	4.59 (0.29)
Sd	2.06	0.30	1.90	0.13	2.05	0.28
RMSE	6.35	0.30	6.29	0.13	6.22	0.29
$D = 100$; $n = m = 40$						
estimate	70 (9.89)	4.60 (0.47)	69 (8.72)	4.61 (0.36)	70 (9.76)	4.62 (0.46)
Sd	11.78	0.43	10.76	0.38	12.29	0.48
RMSE	32.30	0.43	32.36	0.38	32.65	0.49
$D = 100$;$n = m = 100$						
estimate	70 (6.25)	4.62 (0.30)	70 (4.80)	4.60 (0.14)	70 (4.80)	4.57 (0.29)
Sd	7.20	0.30	5.95	0.13	7.19	0.28
RMSE	30.64	0.30	30.36	0.13	30.01	0.29

Table B.1.6: Simulation results from fitting a ZIP iREM to encounters simulated from a ZIP REM. The expected animal speed is set to $\mu = 4.60$ (km/day^{-1}), and the probability of the zero-response category, $\rho = 0.30$. The standard errors, Standard deviation (Sd) and Root Mean Square Error (RMSE) are given.

	gamma			lognormal			Weibull		
	$\hat{D}$	$\hat{\mu}$	$\hat{\rho}$	$\hat{D}$	$\hat{\mu}$	$\hat{\rho}$	$\hat{D}$	$\hat{\mu}$	$\hat{\rho}$
$\mu = 4.60;\ D = 20;\ n = m = 40$									
estimate	20 (7.07)	4.60 (0.47)	0.25 (0.27)	20 (6.78)	4.61 (0.36)	0.25 (0.23)	20 (7.06)	4.62 (0.46)	0.25 (0.27)
Sd	6.64	0.43	0.19	6.53	0.38	0.19	6.98	0.48	0.19
RMSE	6.65	0.43	0.20	6.54	0.38	0.20	6.98	0.49	0.20
$\mu = 0.90;\ D = 20;\ n = m = 100$									
estimate	20 (4.80)	4.62 (0.30)	0.28 (0.23)	20 (4.56)	4.60 (0.14)	0.28 (0.19)	20 (4.83)	4.57 (0.29)	0.28 (0.23)
Sd	4.50	0.30	0.13	4.33	0.13	0.13	4.39	0.28	0.13
RMSE	4.50	0.30	0.14	4.33	0.13	0.13	4.39	0.29	0.13
$\mu = 0.90;\ D = 100;\ n = m = 40$									
estimate	100 (14.56)	4.60 (0.47)	0.30 (0.11)	100 (12.91)	4.61 (0.36)	0.30 (0.11)	100 (14.36)	4.62 (0.46)	0.30 (0.11)
Sd	13.73	0.43	0.08	12.48	0.38	0.08	14.73	0.48	0.08
RMSE	13.74	0.43	0.08	12.48	0.38	0.08	14.74	0.49	0.08
$\mu = 0.90;\ D = 100;\ n = m = 100$									
estimate	99 (9.05)	4.62 (0.30)	0.29 (0.07)	99 (7.06)	4.60 (0.14)	0.29 (0.07)	100 (9.09)	4.57 (0.29)	0.29 (0.07)
Sd	8.68	0.30	0.05	6.88	0.13	0.05	8.55	0.28	0.05
RMSE	8.70	0.30	0.05	6.90	0.13	0.05	8.56	0.29	0.05

B.1.3 Investigating the importance of accounting for overdispersion and zero-inflation in encounter data

In this section we give the parameter estimates from fitting a Poisson iREM to encounters simulated from a ZINB REM, which is discussed in Chapter 3, Section 3.7.4. We investigate the importance of accounting for overdispersion and zero-inflation in the encounter data.

The simulation results from the Poisson iREM are compared with the results from a ZINB iREM in Table B.1.7, while Table B.1.8 compares estimations from a NB iREM and a ZINB iREM. The simulations reveal that both the Poisson iREM and the NB iREM obtained similar estimates of the density to the ZINB iREM for small density values but the Poisson iREM underestimates the standard error. Larger density values resulted in an underestimation of the density and the standard error from both the Poisson iREM and NB iREM but the underestimation is worse from the Poisson iREM. In all cases the true values are captured within an approximate 95% confidence interval. The parameter estimates from a Poisson iREM are given in Table B.1.9, Table B.1.10 gives the parameter estimates from a NB iREM, and the parameter estimates from a ZINB REM are given in Table B.1.11.

.

Table B.1.7: Average density estimates from fitting a Poisson iREM and a ZINB iREM to the data compared with average estimates from Rowcliffe et al. (2008) REM (standard errors in parentheses). We set $\mu = 6.60$ (km/day^{-1}); and for the dispersion parameter, we set $\kappa = 0.10$, and for the probability of zero-response category, $\rho = 0.10$. The Standard deviation (Sd) and Root Mean Square Error (RMSE) are given.

	gamma model				**lognormal model**				**Weibull model**			
	Poisson		ZINB		Poisson		ZINB		Poisson		ZINB	
	iREM $\hat{D}$	REM $\hat{D}$	iREM $\hat{D}$	REM $\hat{D}$	iREM $\hat{D}$	REM $\hat{D}$	iREM $\hat{D}$	REM $\hat{D}$	iREM $\hat{D}$	REM $\hat{D}$	iREM $\hat{D}$	REM $\hat{D}$
Poisson model												
$D = 20$; $n = m = 40$												
estimate	18 (3.44)	18 (3.29)	19 (11.13)	20 (9.98)	18 (3.48)	18 (3.28)	19 (10.64)	20 (9.98)	18 (3.42)	18 (3.23)	20 (10.64)	20 (9.97)
Sd	10.28	10.40	9.61	11.57	10.44	10.27	9.17	11.42	10.11	10.27	10.29	11.41
RMSE	10.52	10.66	9.64	11.58	10.64	10.54	9.19	11.43	10.36	10.54	10.30	11.43
$D = 20$; $n = m = 100$												
estimate	18 (1.99)	18 (1.89)	20 (6.97)	20 (6.29)	18 (1.94)	18 (1.88)	20 (6.96)	20 (6.27)	18 (1.98)	18 (1.89)	20 (6.90)	20 (6.28)
Sd	5.72	5.82	5.71	6.46	5.67	5.72	5.49	6.36	5.69	5.78	5.51	6.42
RMSE	6.02	6.19	5.71	6.47	5.97	6.12	5.49	6.36	6.06	6.16	5.51	6.42
$D = 100$; $n = m = 40$												
estimate	90 (8.68)	90 (7.14)	100 (49.15)	100 (48.09)	89 (8.79)	89 (7.10)	100 (48.83)	99 (47.82)	90 (8.67)	90 (7.13)	101 (48.77)	99 (48.03)
Sd	46.48	47.09	46.77	52.28	46.28	46.21	45.84	51.33	46.75	46.39	46.30	51.47
RMSE	47.63	48.19	46.06	52.28	47.49	47.45	42.06	51.34	47.86	47.53	46.31	51.47
$D = 100$; $n = m = 100$												
estimate	88 (5.16)	89 (4.21)	100 (33.52)	98 (29.92)	89 (4.68)	88 (4.21)	101 (34.21)	98 (29.93)	89 (5.18)	89 (4.22)	101 (33.10)	99 (30.00)
Sd	30.01	30.78	31.16	34.22	30.72	30.89	30.96	34.34	31.72	31.59	31.58	35.13
RMSE	32.25	32.89	31.16	34.26	32.76	32.99	30.57	34.38	33.74	33.57	31.59	35.16

Table B.1.8: Average density estimates from fitting a NB iREM and a ZINB iREM to the data compared with average estimates from Rowcliffe et al. (2008) REM (standard errors in parentheses). We set $\mu = 6.60$ (km/day^{-1}); and for the dispersion parameter, we set $\kappa = 0.10$, and for the probability of zero-response category, $\rho = 0.10$. The Standard deviation (Sd) and Root Mean Square Error (RMSE) are given.

	gamma model				**lognormal model**				**Weibull model**			
	NB		ZINB		NB		ZINB		NB		ZINB	
	iREM $\hat{D}$	REM $\hat{D}$	iREM $\hat{D}$	REM $\hat{D}$	iREM $\hat{D}$	REM $\hat{D}$	iREM $\hat{D}$	REM $\hat{D}$	iREM $\hat{D}$	REM $\hat{D}$	iREM $\hat{D}$	REM $\hat{D}$
Poisson model												
$D = 20$; $n = m = 40$												
estimate	19 (11.00)	18 (9.49)	19 (11.13)	20 (9.98)	19 (11.05)	18 (9.49)	19 (10.64)	20 (9.98)	19 (11.43)	18 (9.48)	20 (10.64)	20 (9.97)
Sd	9.63	10.40	9.61	11.57	9.44	10.27	9.17	11.42	10.74	10.27	10.29	11.41
RMSE	9.70	10.66	9.64	11.58	9.49	10.54	9.19	11.43	10.76	10.54	10.30	11.43
$D = 20$; $n = m = 100$												
estimate	19 (6.89)	18 (5.97)	20 (6.97)	20 (6.29)	19 (6.98)	18 (5.95)	20 (6.96)	20 (6.27)	19 (6.99)	18 (5.96)	20 (6.90)	20 (6.28)
Sd	5.85	6.19	5.71	6.46	5.68	6.12	5.49	6.36	5.65	6.16	5.51	6.42
RMSE	5.91	6.19	5.71	6.47	5.71	6.12	5.49	6.36	5.72	6.16	5.51	6.42
$D = 100$; $n = m = 40$												
estimate	96 (46.06)	90 (45.49)	100 (49.15)	100 (48.09)	97 (47.50)	89 (45.24)	100 (48.83)	99 (47.82)	97 (40.35)	90 (45.43)	101 (48.77)	99 (48.03)
Sd	45.89	47.09	46.77	52.28	42.31	46.22	45.84	51.33	48.00	46.39	46.30	51.47
RMSE	46.10	48.19	46.06	52.28	42.44	47.45	42.06	51.34	48.10	47.53	46.31	51.47
$D = 100$; $n = m = 100$												
estimate	94 (31.48)	88 (28.28)	100 (33.52)	98 (29.92)	95 (33.21)	88 (28.29)	101 (34.21)	98 (29.93)	95 (30.66)	89 (28.36)	101 (33.10)	99 (30.00)
Sd	31.21	30.78	31.16	34.22	30.06	30.89	30.96	34.34	31.22	31.59	31.58	35.13
RMSE	31.84	32.89	31.16	34.26	30.59	32.99	34.38	32.99	31.67	33.57	31.59	35.16

Table B.1.9: Average parameter estimates obtained from fitting a Poisson iREM to encounters simulated from a ZINB REM. The expected speed is set to $\mu = 6.60$ (km/day^{-1}). The standard errors (in parentheses), Standard deviation (Sd) and Root Mean Square Error (RMSE) are given.

	speed data models					
	gamma		lognormal		Weibull	
	$\hat{D}$	$\hat{\mu}$	$\hat{D}$	$\hat{\mu}$	$\hat{D}$	$\hat{\mu}$
$D = 20$; $n = m = 40$						
estimate	18 (3.44)	6.60 (0.35)	18 (3.48)	6.59 (0.37)	18 (3.42)	6.58 (0.35)
Sd	10.28	0.40	10.44	0.37	10.11	0.33
RMSE	10.52	0.40	10.64	0.37	10.36	0.33
$D = 20$; $n = m = 100$						
estimate	18 (1.99)	6.60 (0.22)	18 (1.94)	6.60 (0.15)	18 (1.98)	6.60 (0.22)
Sd	5.72	0.23	5.67	0.15	5.69	0.22
RMSE	6.02	0.23	5.97	0.15	6.06	0.22
$D = 100$; $n = m = 40$						
estimate	90 (8.69)	6.60 (0.35)	89 (8.79)	6.64 (0.37)	90 (8.67)	6.58 (0.35)
Sd	46.48	0.42	46.28	0.42	46.75	0.34
RMSE	47.63	0.43	47.49	0.42	47.86	0.34
$D = 100$;$n = m = 100$						
estimate	88 (5.15)	6.60 (0.22)	89 (4.68)	6.60 (0.15)	89 (5.18)	6.61 (0.22)
Sd	30.01	0.24	30.72	0.19	31.72	0.25
RMSE	32.25	0.24	32.76	0.19	33.74	0.25

Table B.1.10: Average parameter estimates obtained from fitting a NB iREM to encounters simulated from a ZINB REM. For expected speed and the dispersion parameter we set $\mu = 6.60$ (km/day^{-1}) and $\kappa = 0.10$, respectively. The standard errors, Standard deviation (Sd) and Root Mean Square Error (RMSE) are given.

	gamma			lognormal			Weibull		
	$\hat{D}$	$\hat{\mu}$	$\hat{\kappa}$	$\hat{D}$	$\hat{\mu}$	$\hat{\kappa}$	$\hat{D}$	$\hat{\mu}$	$\hat{\kappa}$
$\mu = 0.90,\ D = 20;\ n = m = 40$									
estimate	19 (11.00)	6.61 (0.35)	0.12 (0.06)	19 (11.05)	6.60 (0.38)	0.12 (0.06)	19 (11.43)	6.59 (0.35)	0.12 (0.06)
Sd	9.63	0.40	0.06	9.44	0.39	0.06	10.74	0.34	0.06
RMSE	9.70	0.40	0.06	9.49	0.38	0.07	10.76	0.34	0.06
$D = 20;\ n = m = 100$									
estimate	19 (6.89)	6.59 (0.22)	0.10 (0.03)	19 (6.98)	6.61 (0.15)	0.10 (0.03)	19 (6.99)	6.61 (0.22)	0.10 (0.03)
Sd	5.85	0.22	0.03	5.68	0.14	0.03	5.65	0.21	0.03
RMSE	5.91	0.22	0.03	5.71	0.14	0.03	5.72	0.21	0.03
$D = 100;\ n = m = 40$									
estimate	96 (46.06)	6.59 (0.35)	0.10 (0.03)	97 (47.50)	6.63 (0.37)	0.10 (0.03)	97 (40.35)	6.59 (0.35)	0.10 (0.03)
Sd	45.89	0.39	0.03	42.31	0.39	0.03	48.00	0.33	0.04
RMSE	46.10	0.39	0.03	42.44	0.39	0.03	48.10	0.33	0.04
$D = 100;\ n = m = 100$									
estimate	94 (31.48)	6.61 (0.22)	0.09 (0.02)	95 (33.21)	6.60 (0.15)	0.09 (0.02)	95 (30.66)	6.59 (0.22)	0.09 (0.02)
Sd	31.21	0.23	0.02	30.06	0.16	0.02	31.22	0.22	0.02
RMSE	31.84	0.23	0.02	30.59	0.16	0.02	31.67	0.22	0.02

Table B.1.11: Simulation results from fitting a ZINB iREM to encounters simulated from a ZINB REM for $\rho = 0.10$, $\mu = 6.60$ (km/day^{-1}), and $\kappa = 0.10$. The standard errors, Standard deviation (Sd) and Root Mean Square Error (RMSE) are given.

	gamma				lognormal				Weibull			
	$\hat{D}$	$\hat{\mu}$	$\hat{\kappa}$	$\hat{\rho}$	$\hat{D}$	$\hat{\mu}$	$\hat{\kappa}$	$\hat{\rho}$	$\hat{D}$	$\hat{\mu}$	$\hat{\kappa}$	$\hat{\rho}$
$D = 20$; $n = m = 40$												
estimate	19 (11.13)	6.60 (0.35)	0.12 (0.07)	0.11 (0.10)	19 (10.97)	6.59 (0.37)	0.12 (0.07)	0.11 (0.04)	20 (10.64)	6.58 (0.07)	0.13 (0.07)	0.11 (0.06)
Sd	9.61	0.39	0.06	0.05	9.17	0.38	0.05	0.05	10.29	0.34	0.07	0.05
RMSE	9.64	0.39	0.06	0.05	9.19	0.38	0.06	0.05	10.30	0.34	0.08	0.07
$D = 20$; $n = m = 100$												
estimate	20 (6.97)	6.60 (0.22)	0.11 (0.03)	0.10 (0.03)	20 (6.96)	6.60 (0.15)	0.11 (0.03)	0.11 (0.01)	20 (6.91)	6.60 (0.22)	0.11 (0.03)	0.10 (0.03)
Sd	5.71	0.23	0.03	0.03	5.49	0.14	0.03	0.03	5.51	0.21	0.03	0.03
RMSE	5.71	0.23	0.03	0.03	5.49	0.14	0.03	0.03	5.51	0.21	0.03	0.03
$D = 100$; $n = m = 40$												
estimate	100 (49.15)	6.60 (0.35)	0.11 (0.04)	0.12 (0.02)	100 (48.83)	6.63 (0.37)	0.11 (0.04)	0.12 (0.02)	101 (48.77)	6.59 (0.35)	0.11 (0.04)	0.12 (0.02)
Sd	46.77	0.40	0.04	0.05	42.04	0.40	0.04	0.05	46.30	0.32	0.04	0.05
RMSE	46.06	0.40	0.04	0.06	42.04	0.41	0.04	0.05	46.31	0.32	0.04	0.05
$D = 100$; $n = m = 100$												
estimate	100 (33.52)	6.60 (0.22)	0.10 (0.02)	0.12 (0.01)	101 (32.21)	6.60 (0.15)	0.11 (0.02)	0.12 (0.01)	101 (31.10)	6.59 (0.22)	0.11 (0.02)	0.12 (0.01)
Sd	31.16	0.22	0.02	0.03	30.96	0.15	0.02	0.03	31.58	0.22	0.02	0.03
RMSE	31.16	0.22	0.02	0.04	30.57	0.16	0.02	0.04	31.59	0.22	0.02	0.03

Appendix C

This section of the appendices is linked to Chapter 4 in the main thesis. Here we provide additional results from those given in the Chapter. The appendix begins, in Section C.1 with some simulation results from iREM with habitat models as discussed in Section 4.2. It continues in Section C.2 to give some analyses of the application of iREM with habitat models to WWAP data as discussed in Section 4.3. We give an example of the R codes for iREM with random effect in Section C.3. Finally, we provide some analyses of WWAP data set using an iREM with random effect model in Section C.4, and as described in Section 4.5.

C.1 Simulation results from iREM with habitat models

This section of the appendices is linked to Section 4.2. The results are given for iREM with habitat where animal speed is assumed to follow a gamma, lognormal or Weibull model. Section C.1.1 gives the results from fitting a Poisson iREM with habitat and Section C.1.2 investigates the importance of accounting for variation in the encounter data. This is done by fitting a NB iREM with habitat and a Poisson iREM with habitat to encounter data simulated from a NB REM. It continues in Section C.1.3 to investigate the importance of accounting for zero-inflation in the encounter data. To do this we fit a ZIP iREM with habitat and a Poisson iREM with habitat to encounter data simulated from a ZIP REM. The importance of accounting for both zero-inflation and variation in encounter is investigated in Section C.1.4 by fitting a Poisson iREM with habitat, a NB iREM with habitat and a ZINB iREM with habitat to encounter data simulated

from a ZINB REM.

C.1.1 Simulation results from a Poisson REM with habitat

The results from a gamma model and a Weibull model are given here in Table C.1.1, and C.1.2, respectively. The results from the lognormal model are given in Table 4.2.2 in Chapter 4, Section 4.2.1. Broadly the patterns are similar for each speed data model but the lognormal model (Table 4.2.2, Section 4.2.1) generally obtained smaller standard errors and lower RMSEs compared with the gamma model and Weibull model. The simulations show that the regression estimators have some bias when sample sizes are small, but this bias is minimal when sample sizes increase.

Table C.1.1: Average parameter estimates where a Poisson iREM with habitat is fitted to encounters simulated from a Poisson REM and a gamma model is fitted to data simulated from the relevant fitted model (standard error in parentheses). The sample sizes are n trap days, and m animal speed data. The Standard deviation (Sd) and Root Mean Square Error (RMSE) are also given.

	Parameter estimates				
	Habitat 1 $\hat{\beta}_1$	Habitat 2 $\hat{\beta}_2$	Habitat 3 $\hat{\beta}_3$	Habitat 4 $\hat{\beta}_4$	$\hat{\mu}_x$
$\beta_1 = 4.62$, $\beta_2 = 0.31$, $\beta_3 = 3.13$, $\beta_4 = 0.30$, and $\mu_x = 0.71$;					
$n = m = 40$					
Estimate	4.60 (0.24)	0.32 (0.27)	3.16 (0.21)	0.33 (0.27)	0.71 (0.09)
Sd	0.24	0.27	0.21	0.26	0.09
RMSE	0.24	0.27	0.21	0.26	0.09
$n = m = 100$					
Estimate	4.61 (0.15)	0.32 (0.17)	3.14 (0.13)	0.29 (0.17)	0.71 (0.05)
Sd	0.16	0.17	0.13	0.17	0.05
RMSE	0.16	0.17	0.13	0.17	0.05
$\beta_1 = 3.56$, $\beta_2 = 1.61$, $\beta_3 = 0.18$, $\beta_4 = 0.40$, and $\mu_x = 0.71$					
$n = m = 40$					
Estimate	3.51 (0.38)	1.65 (0.40)	0.17 (0.50)	0.40 (0.47)	0.70 (0.08)
Sd	0.35	0.35	0.49	0.40	0.07
RMSE	0.35	0.35	0.49	0.40	0.07
$n = m = 100$					
Estimate	3.55 (0.23)	1.61 (0.24)	0.15 (0.30)	0.40 (0.29)	0.71 (0.05)
Sd	0.23	0.23	0.31	0.29	0.05
RMSE	0.23	0.23	0.31	0.29	0.05

Table C.1.2: Average parameter estimates where a Poisson iREM with habitat is fitted to encounters simulated from a Poisson REM and a Weibull model is fitted to data simulated from the relevant fitted model (standard error in parentheses). The sample sizes are n trap days, and m animal speed data. The Standard deviation (Sd) and Root Mean Square Error (RMSE) are also given.

	Parameter estimates				
	Habitat 1 $\hat{\beta}_1$	Habitat 2 $\hat{\beta}_2$	Habitat 3 $\hat{\beta}_3$	Habitat 4 $\hat{\beta}_4$	$\hat{\mu}_x$
$\beta_1 = 4.62$, $\beta_2 = 0.31$, $\beta_3 = 3.13$, $\beta_4 = 0.30$; $\mu_x = 0.71$					
$n = m = 40$					
Estimate	4.59 (0.24)	0.32 (0.27)	3.16 (0.21)	0.34 (0.27)	0.71 (0.08)
Sd	0.23	0.27	0.21	0.26	0.09
RMSE	0.23	0.27	0.21	0.26	0.09
$n = m = 100$					
Estimate	4.61 (0.15)	0.32 (0.17)	3.14 (0.13)	0.29 (0.17)	0.71 (0.05)
Sd	0.15	0.17	0.13	0.17	0.05
RMSE	0.15	0.17	0.13	0.17	0.05
$\beta_1 = 3.56$, $\beta_2 = 1.61$, $\beta_3 = 0.18$, $\beta_4 = 0.40$, and $\mu_x = 0.71$					
$n = m = 40$					
Estimate	3.49 (0.38)	1.65 (0.40)	0.17 (0.50)	0.40 (0.47)	0.72 (0.09)
Sd	0.34	0.34	0.49	0.40	0.08
RMSE	0.35	0.34	0.49	0.40	0.08
$n = m = 100$					
Estimate	3.55 (0.23)	1.61 (0.24)	0.15 (0.30)	0.40 (0.29)	0.72 (0.07)
Sd	0.23	0.23	0.31	0.29	0.05
RMSE	0.23	0.23	0.31	0.29	0.05

C.1.2 Investigating the importance of accounting for variation in encounter data

The results from the gamma model (Table C.1.3) and Weibull model (Table C.1.4) are given in this section. In Chapter 4, Section 4.2.2 we give the results from the lognormal model. The simulation results show minimal differences between the three speed data models. However, the simulations reveal that ignoring the variation in encounter data can induce an underestimation of the standard error of the density.

Table C.1.3: Average parameter estimates from fitting a Poisson iREM with habitat and a NB iREM with habitat to encounters simulated from a NB REM (standard error in parentheses). Animal speed is assumed to follow a gamma model. The sample sizes are n trap days, and m animal speed data. The Standard deviation (Sd) and Root Mean Square Error (RMSE) are also given.

	Parameter estimates					
	Habitat 1	Habitat 2	Habitat 3	Habitat 4		
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\mu}_x$	$\hat{\kappa}$
$\beta_1 = 4.62$, $\beta_2 = 0.31$, $\beta_3 = 3.13$, $\beta_4 = 0.30$; $\mu_x = 0.71$; $\kappa = 0.56$						
$n = m = 40$						
Poisson	4.61 (0.24)	0.31 (0.28)	3.13 (0.21)	0.30 (0.28)	0.70 (0.08)	-
Sd	0.33	0.42	0.38	0.39	0.08	-
RMSE	0.33	0.42	0.38	0.39	0.08	-
NB	4.60 (0.32)	0.33 (0.40)	3.14 (0.36)	0.32 (0.40)	0.70 (0.12)	0.60 (0.11)
Sd	0.29	0.38	0.33	0.35	0.08	0.12
RMSE	0.29	0.38	0.33	0.35	0.08	0.12
$n = m = 100$						
Poisson	4.61 (0.16)	0.32 (0.17)	3.14 (0.13)	0.30 (0.17)	0.71 (0.05)	-
Sd	0.16	0.17	0.13	0.17	0.05	-
RMSE	0.16	0.17	0.13	0.17	0.05	-
NB	4.61 (0.15)	0.32 (0.17)	3.14 (0.13)	0.30 (0.17)	0.71 (0.07)	0.57 (0.07)
Sd	0.16	0.17	0.13	0.17	0.07	0.07
RMSE	0.16	0.17	0.13	0.17	0.05	0.07

Table C.1.4: Average parameter estimates from fitting a Poisson iREM with habitat and a NB iREM with habitat to encounters simulated from a NB REM (standard error in parentheses). Animal speed is assumed to follow a Weibull model. The sample sizes are n trap days, and m animal speed data. The Standard deviation (Sd) and Root Mean Square Error (RMSE) are also given.

	Parameter estimates					
	Habitat 1	Habitat 2	Habitat 3	Habitat 4		
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\mu}_x$	$\hat{\kappa}$
$\beta_1 = 4.62$, $\beta_2 = 0.31$, $\beta_3 = 3.13$, $\beta_4 = 0.30$; $\mu_x = 0.71$; $\kappa = 0.56$						
$n = m = 40$						
Poisson	4.59 (0.24)	0.31 (0.28)	3.13 (0.22)	0.31 (0.28)	0.71 (0.09)	-
Sd	0.33	0.41	0.37	0.39	0.09	-
RMSE	0.33	0.41	0.37	0.39	0.09	-
NB	4.59 (0.32)	0.31 (0.41)	3.13 (0.37)	0.31 (0.41)	0.71 (0.12)	0.59 (0.12)
Sd	0.34	0.42	0.38	0.40	0.09	0.12
RMSE	0.35	0.42	0.38	0.40	0.09	0.12
$n = m = 100$						
Poisson	4.61 (0.15)	0.31 (0.17)	3.16 (0.13)	0.31 (0.17)	0.70 (0.05)	-
Sd	0.17	0.23	0.21	0.25	0.05	
RMSE	0.17	0.23	0.21	0.25	0.05	
NB	4.61 (0.20)	0.31 (0.26)	3.16 (0.23)	0.31 (0.26)	0.70 (0.07)	0.57 (0.07)
Sd	0.18	0.24	0.21	0.26	0.05	0.07
RMSE	0.18	0.24	0.22	0.26	0.05	0.07

C.1.3 Investigating the importance of accounting for zero-inflation in encounter data

In this section we give the results from fitting a gamma model (Table C.1.5) and a Weibull model (Table C.1.6) to data simulated from the relevant fitted models. The results from the lognormal model are given in Table 4.2.4, Section 4.2.3 in Chapter 4. The Simulation results (Table C.1.5 and Table C.1.6) show that not accounting for zero-inflation in encounter data can introduce bias and an underestimation of the standard error of the density.

Table C.1.5: Average parameter estimates from fitting a Poisson iREM with habitat and a ZIP iREM with habitat to encounters simulated from a ZIP REM (standard error in parentheses). The parameters are set to $\beta_1 = 4.62$, $\beta_2 = 0.31$, $\beta_3 = 3.13$, $\beta_4 = 0.30$; $\mu_x = 0.71$; $\rho = 0.35$. Animal speed is assumed to follow a gamma model. The sample sizes are n trap days, and m animal speed data. The Standard deviation (Sd) and Root Mean Square Error (RMSE) are also given.

	Parameter estimates					
	Habitat 1	Habitat 2	Habitat 3	Habitat 4		
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\mu}_x$	$\hat{\rho}$
$\beta_1 = 4.62$, $\beta_2 = 0.31$, $\beta_3 = 3.13$, $\beta_4 = 0.30$; $\mu_x = 0.71$; $\rho = 0.35$						
$n = m = 40$						
Poisson	4.13 (0.29)	0.35 (0.35)	3.18 (0.27)	0.33 (0.35)	0.71 (0.09)	-
Sd	0.31	0.40	0.32	0.41	0.08	-
RMSE	0.58	0.40	0.33	0.41	0.08	-
ZIP	4.57 (0.32)	0.33 (0.38)	3.18 (0.31)	0.32 (0.38)	0.70 (0.07)	0.35 (0.10)
Sd	0.34	0.41	0.33	0.41	0.07	0.07
RMSE	0.34	0.41	0.33	0.41	0.07	0.07
$n = m = 100$						
Poisson	4.16 (0.18)	0.33 (0.21)	3.14 (0.17)	0.30 (0.22)	0.71 (0.05)	-
Sd	0.18	0.22	0.19	0.24	0.05	-
RMSE	0.49	0.23	0.19	0.24	0.05	-
ZIP	4.60 (0.20)	0.31 (0.24)	3.14 (0.19)	0.28 (0.24)	0.71 (0.05)	0.35 (0.06)
Sd	0.18	0.23	0.18	0.24	0.04	0.04
RMSE	0.18	0.23	0.18	0.24	0.04	0.04

Table C.1.6: Average parameter estimates from fitting a Poisson iREM with habitat and a ZIP iREM with habitat to encounters simulated from a ZIP REM (standard error in parentheses). Animal speed is assumed to follow a Weibull model. The parmamters are set to $\beta_1 = 4.62$, $\beta_2 = 0.31$, $\beta_3 = 3.13$, $\beta_4 = 0.30$; $\mu_x = 0.71$; and $\rho = 0.35$. The sample sizes are n trap days, and m animal speed data. The Standard deviation (Sd) and Root Mean Square Error (RMSE) are also given.

	Parameter estimates					
	Habitat 1	Habitat 2	Habitat 3	Habitat 4		
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\mu}_x$	$\hat{\rho}$
$\beta_1 = 4.62$, $\beta_2 = 0.31$, $\beta_3 = 3.13$, $\beta_4 = 0.30$; $\mu_x = 0.71$; $\rho = 0.35$						
$n = m = 40$						
Poisson	4.15 (0.29)	0.34 (0.35)	3.18 (0.27)	0.33 (0.35)	0.70 (0.08)	-
Sd	0.31	0.39	0.32	0.40	0.08	-
RMSE	0.56	0.39	0.32	0.40	0.08	-
ZIP	4.59 (0.32)	0.33 (0.38)	3.18 (0.31)	0.33 (0.38)	0.70 (0.11)	0.35 (0.10)
Sd	0.34	0.41	0.33	0.41	0.07	0.07
RMSE	0.35	0.41	0.33	0.41	0.07	0.07
$n = m = 100$						
Poisson	4.16 (0.18)	0.33 (0.21)	3.15 (0.17)	0.31 (0.22)	0.71 (0.05)	-
Sd	0.18	0.23	0.19	0.24	0.05	
RMSE	0.18	0.23	0.19	0.24	0.05	
ZIP	4.61 (0.20)	0.31 (0.24)	3.14 (0.19)	0.28 (0.24)	0.71 (0.05)	0.35 (0.06)
Sd	0.19	0.23	0.17	0.23	0.05	0.04
RMSE	0.19	0.23	0.17	0.23	0.05	0.04

C.1.4 Investigating the importance of accounting for zero-inflation and variation in encounter data

In this section we account for zero-inflation and variation in the encounter data when habitat is incorporated into iREM.

The results from fitting a gamma model to animal speed data are given in Table C.1.7 for small sample sizes, and Table C.1.8 for large sample sizes. Table C.1.9 and Table C.1.10 give the results from fitting a lognormal model to animal speed for small and large sample sizes, respectively. While Table C.1.11 and Table C.1.12 give the results from fitting a Weibull model to the speed data for small and large sample sizes, respectively. The simulation results reveal that not accounting for zero-inflation and/or variation in the data can introduce bias and an underestimation of the standard error. The speed data models obtained similar estimates of the parameters.

Table C.1.7: Average parameter estimates from fitting a Poisson iREM with habitat, a NB iREM with habitat and a ZINB iREM with habitat to encounters simulated from a ZINB REM (standard error in parentheses). The parameters are $\beta_1 = 4.62$, $\beta_2 = 0.31$, $\beta_3 = 3.13$, $\beta_4 = 0.30$; $\mu_x = 0.71$; $\kappa = 0.56$; $\rho = 0.35$. Animal speed is assumed to follow a gamma model. The sample sizes are $n = 40$ trap days, and $m = 40$ animal speed data. The Standard deviation (Sd) and Root Mean Square Error (RMSE) are also given.

	Parameter estimates						
	Habitat 1	Habitat 2	Habitat 3	Habitat 4			
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\mu}_x$	$\hat{\kappa}$	ρ
$n = m = 40$							
Poisson	4.13 (0.14)	0.30 (0.19)	3.12 (0.11)	0.33 (0.19)	0.71 (0.11)	-	-
Sd	0.37	0.48	0.41	0.46	0.08	-	-
RMSE	0.61	0.48	0.41	0.57	0.08	-	-
NB	4.33 (0.39)	0.58 (0.54)	3.28 (0.50)	0.60 (0.54)	0.64 (0.11)	0.41 (0.22)	-
Sd	0.29	0.18	0.16	0.18	0.06	0.19	-
RMSE	0.41	0.33	0.22	0.35	0.10	0.39	-
ZINB	4.61 (0.41)	0.30 (0.49)	3.12 (0.44)	0.32 (0.49)	0.71 (0.11)	1.00 (0.38)	0.35 (0.40)
Sd	0.39	0.49	0.41	0.47	0.08	0.44	0.10
RMSE	0.38	0.49	0.41	0.47	0.08	0.50	0.11

Table C.1.8: Average parameter estimates from fitting a Poisson iREM with habitat, a NB iREM with habitat and a ZINB iREM with habitat to encounters simulated from a ZINB REM (standard error in parentheses). The parameters are $\beta_1 = 4.62$, $\beta_2 = 0.31$, $\beta_3 = 3.13$, $\beta_4 = 0.30$; $\mu_x = 0.71$; $\kappa = 0.56$; $\rho = 0.35$. Animal speed is assumed to follow a gamma model. The sample sizes are $n = 100$ trap days, and $m = 100$ animal speed data. The Standard deviation (Sd) and Root Mean Square Error (RMSE) are also given.

	Parameter estimates						
	Habitat 1	Habitat 2	Habitat 3	Habitat 4			
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\mu}_x$	$\hat{\kappa}$	ρ
$n = m = 100$							
Poisson	4.16 (0.08)	0.37 (0.11)	3.15 (0.06)	0.34 (0.11)	0.71 (0.07)	-	-
Sd	0.24	0.32	0.27	0.32	0.06	-	-
RMSE	0.53	0.33	0.27	0.32	0.06	-	-
NB	4.41 (0.25)	0.61 (0.34)	3.31 (0.32)	0.60 (0.34)	0.63 (0.08)	0.38 (0.14)	-
Sd	0.24	0.14	0.14	0.14	0.05	0.14	-
RMSE	0.32	0.32	0.22	0.33	0.09	0.40	-
ZINB	4.62 (0.26)	0.36 (0.31)	3.15 (0.28)	0.34 (0.31)	0.71 (0.07)	0.86 (0.25)	0.37 (0.28)
Sd	0.27	0.31	0.28	0.32	0.06	0.21	0.07
RMSE	0.27	0.31	0.28	0.33	0.06	0.24	0.07

Table C.1.9: Average parameter estimates from fitting a Poisson iREM with habitat, a NB iREM with habitat and a ZINB iREM with habitat to encounters simulated from a ZINB REM (standard error in parentheses). The parameters are $\beta_1 = 4.62$, $\beta_2 = 0.31$, $\beta_3 = 3.13$, $\beta_4 = 0.30$; $\mu_x = 0.71$; $\kappa = 0.56$; $\rho = 0.35$. Animal speed is assumed to follow a lognormal model. The sample sizes are $n = 40$ trap days, and $m = 40$ animal speed data. The Standard deviation (Sd) and Root Mean Square Error (RMSE) are also given.

	Parameter estimates						
	Habitat 1	Habitat 2	Habitat 3	Habitat 4			
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\mu}_x$	$\hat{\kappa}$	ρ
$n = m = 40$							
Poisson	4.11 (0.12)	0.31 (0.19)	3.14 (0.11)	0.34 (0.19)	0.71 (0.09)	-	-
Sd	0.36	0.48	0.40	0.46	0.07	-	-
RMSE	0.62	0.48	0.40	0.46	0.07	-	-
NB	4.33 (0.39)	0.57 (0.53)	3.32 (0.50)	0.60 (0.52)	0.68 (0.11)	0.43 (0.22)	-
Sd	0.31	0.19	0.17	0.20	0.07	0.19	-
RMSE	0.43	0.33	0.26	0.36	0.07	0.37	-
ZINB	4.59 (0.38)	0.31 (0.48)	3.14 (0.42)	0.33 (0.48)	0.71 (0.08)	0.99 (0.38)	0.37 (0.39)
Sd	0.38	0.50	0.40	0.47	0.07	0.43	0.11
RMSE	0.39	0.50	0.40	0.47	0.07	0.49	0.11

Table C.1.10: Average parameter estimates from fitting a Poisson iREM with habitat, a NB iREM with habitat and a ZINB iREM with habitat to encounters simulated from a ZINB REM (standard error in parentheses). The parameters are $\beta_1 = 4.62$, $\beta_2 = 0.31$, $\beta_3 = 3.13$, $\beta_4 = 0.30$; $\mu_x = 0.71$; $\kappa = 0.56$; $\rho = 0.35$. Animal speed is assumed to follow a lognormal model. The sample sizes are $n = 100$ trap days, and $m = 100$ animal speed data. The Standard deviation (Sd) and Root Mean Square Error (RMSE) are also given.

	Parameter estimates						
	Habitat 1	Habitat 2	Habitat 3	Habitat 4			
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\mu}_x$	$\hat{\kappa}$	ρ
$n = m = 100$							
Poisson	4.15 (0.07)	0.36 (0.11)	3.16 (0.06)	0.34 (0.11)	0.71 (0.06)	-	-
Sd	0.23	0.32	0.26	0.31	0.04	-	-
RMSE	0.53	0.33	0.26	0.32	0.04	-	-
NB	4.42 (0.25)	0.59 (0.33)	3.33 (0.32)	0.57 (0.33)	0.66 (0.04)	0.40 (0.14)	-
Sd	0.23	0.15	0.13	0.15	0.06	0.13	-
RMSE	0.30	0.31	0.24	0.31	0.07	0.38	-
ZINB	4.62 (0.25)	0.35 (0.31)	3.15 (0.28)	0.32 (0.31)	0.71 (0.06)	0.86 (0.25)	0.37 (0.28)
Sd	0.25	0.32	0.29	0.32	0.04	0.21	0.06
RMSE	0.25	0.32	0.29	0.32	0.04	0.23	0.07

Table C.1.11: Average parameter estimates from fitting a Poisson iREM with habitat, a NB iREM with habitat and a ZINB iREM with habitat to encounters simulated from a ZINB REM (standard error in parentheses). The parameters are $\beta_1 = 4.62$, $\beta_2 = 0.31$, $\beta_3 = 3.13$, $\beta_4 = 0.30$; $\mu_x = 0.71$; $\kappa = 0.56$; $\rho = 0.35$. Animal speed is assumed to follow a Weibull model. The sample sizes are $n = 40$ trap days, and $m = 40$ animal speed data. The Standard deviation (Sd) and Root Mean Square Error (RMSE) are also given.

	Parameter estimates						
	Habitat 1	Habitat 2	Habitat 3	Habitat 4			
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\mu}_x$	$\hat{\kappa}$	ρ
$n = m = 40$							
Poisson	4.13 (0.14)	0.30 (0.19)	3.12 (0.11)	0.33 (0.19)	0.71 (0.11)	-	-
Sd	0.39	0.49	0.41	0.45	0.08	-	-
RMSE	0.61	0.49	0.41	0.45	0.08	-	-
NB	4.27 (0.37)	0.60 (0.51)	3.34 (0.49)	0.62 (0.51)	0.69 (0.11)	0.44 (0.22)	-
Sd	0.34	0.21	0.16	0.20	0.08	0.20	-
RMSE	0.48	0.36	0.26	0.38	0.09	0.37	-
ZINB	4.60 (0.41)	0.30 (0.50)	3.12 (0.44)	0.32 (0.49)	0.71 (0.11)	0.98 (0.38)	0.37 (0.40)
Sd	0.39	0.48	0.41	0.47	0.08	0.42	0.10
RMSE	0.39	0.48	0.41	0.47	0.08	0.48	0.11

Table C.1.12: Average parameter estimates from fitting a Poisson iREM with habitat, a NB iREM with habitat and a ZINB iREM with habitat to encounters simulated from a ZINB REM (standard error in parentheses). The parameters are $\beta_1 = 4.62$, $\beta_2 = 0.31$, $\beta_3 = 3.13$, $\beta_4 = 0.30$; $\mu_x = 0.71$; $\kappa = 0.56$; $\rho = 0.35$. Animal speed is assumed to follow a Weibull model. The sample sizes are $n = 100$ trap days, and $m = 100$ animal speed data. The Standard deviation (Sd) and Root Mean Square Error (RMSE) are also given.

	Parameter estimates						
	Habitat 1	Habitat 2	Habitat 3	Habitat 4			
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\mu}_x$	$\hat{\kappa}$	ρ
$n = m = 100$							
Poisson	4.15 (0.08)	0.35 (0.11)	3.15 (0.06)	0.34 (0.11)	0.71 (0.07)	-	-
Sd	0.23	0.31	0.26	0.31	0.05	-	-
RMSE	0.52	0.32	0.26	0.31	0.05	-	-
NB	4.37 (0.26)	0.59 (0.31)	3.34 (0.30)	0.60 (0.31)	0.69 (0.07)	0.41 (0.14)	-
Sd	0.26	0.16	0.15	0.16	0.07	0.16	-
RMSE	0.37	0.32	0.26	0.34	0.08	0.37	-
ZINB	4.61 (0.26)	0.36 (0.31)	3.15 (0.28)	0.34 (0.31)	0.71 (0.07)	0.86 (0.25)	0.37 (0.28)
Sd	0.27	0.32	0.28	0.32	0.05	0.21	0.06
RMSE	0.27	0.32	0.28	0.32	0.05	0.24	0.07

C.1.5 Simulation results for animals moving in groups

In this section we give results for animals moving in pairs or family groups within habitats. We give the results from fitting a lognormal model (Table C.1.13) and a Weibull model (Table C.1.13) here, and the results from fitting a gamma model to animal speed data are given in Table 4.2.5 in Section 4.2.5 in the main thesis. The results show some bias in the regression estimators for animals with low expected speeds. But for larger expected speeds precise estimates are obtained for large sample sizes (Table 4.2.6, Section 4.2.5).

Table C.1.13: Average parameter estimates from fitting a Poisson iREM to encounter data simulated from a Poisson REM, and from fitting a ZTP model to group data and a lognormal model to animal speed data simulated from the relevant fitted models. The standard errors are in parentheses. The parameters are set to $\beta_1 = 4.62$, $\beta_2 = 0.31$, $\beta_3 = 3.13$, $\beta_4 = 0.30$; $\mu_x = 0.71$; and $g = 2.50$ or $g = 5.50$. The sample sizes are n trap days, m animal speed data, and group size, $S = 10$ or $S = 50$. The Standard deviation (Sd) and Root Mean Square Error (RMSE) are also given.

	Estimated Parameters						
	Habitat 1	Habitat 2	Habitat 3	Habitat 4			
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\mu}_x$	$\hat{\phi}$	$\hat{g}$
$n = m = 40$; $S = 10$							
Estimate	4.63 (0.18)	0.40 (0.22)	3.21 (0.12)	0.36 (0.23)	0.70 (0.09)	2.50 (0.22)	2.73 (0.59)
Sd	0.08	0.19	0.09	0.19	0.05	0.23	0.20
RMSE	0.08	0.21	0.12	0.20	0.05	0.35	0.30
$n = m = 100$; $S = 10$							
Estimate	4.62 (0.11)	0.37 (0.15)	3.19 (0.09)	0.36 (0.15)	0.70 (0.06)	2.38 (0.17)	2.62 (0.44)
Sd	0.05	0.16	0.06	0.16	0.04	0.12	0.10
RMSE	0.05	0.17	0.09	0.17	0.04	0.19	0.16
$n = m = 40$; $S = 50$							
Estimate	4.68 (0.14)	0.36 (0.23)	3.25 (0.12)	0.33 (0.23)	0.70 (0.09)	2.83 (0.09)	3.01 (0.27)
Sd	0.11	0.21	0.10	0.21	0.06	0.23	0.20
RMSE	0.12	0.22	0.16	0.21	0.06	0.64	0.54
$n = m = 100$; $S = 50$							
Estimate	4.63 (0.10)	0.38 (0.14)	3.23 (0.08)	0.34 (0.14)	0.70 (0.05)	2.57 (0.09)	2.78 (0.26)
Sd	0.06	0.14	0.07	0.17	0.04	0.16	0.14
RMSE	0.06	0.16	0.12	0.17	0.04	0.37	0.31

Table C.1.14: Average parameter estimates from fitting a Poisson iREM to encounter data simulated from a Poisson REM, and from fitting a ZTP model to group data and a Weibull model to animal speed data simulated from the relevant fitted models. The standard errors are in parentheses. The parameters are set to $\beta_1 = 4.62$, $\beta_2 = 0.31$, $\beta_3 = 3.13$, $\beta_4 = 0.30$; $\mu_x = 0.71$; and $g = 2.50$ or $g = 5.50$. The sample sizes are n trap days, m animal speed data, and group size, $S = 10$ or $S = 50$. The Standard deviation (Sd) and Root Mean Square Error (RMSE) are also given.

	Estimated Parameters						
	Habitat 1	Habitat 2	Habitat 3	Habitat 4			
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\mu}_x$	$\hat{\phi}$	$\hat{g}$
$n = m = 40$; $S = 10$							
Estimate	4.62 (0.18)	0.40 (0.23)	3.21 (0.13)	0.33 (0.23)	0.71 (0.12)	2.50 (0.21)	2.72 (0.58)
Sd	0.09	0.20	0.09	0.22	0.07	0.24	0.21
RMSE	0.09	0.22	0.12	0.22	0.07	0.36	0.30
$n = m = 100$; $S = 10$							
Estimate	4.61 (0.12)	0.38 (0.15)	3.20 (0.09)	0.34 (0.15)	0.70 (0.07)	2.37 (0.17)	2.62 (0.44)
Sd	0.06	0.17	0.06	0.16	0.04	0.14	0.12
RMSE	0.06	0.18	0.09	0.16	0.04	0.20	0.17
$n = m = 40$; $S = 50$							
Estimate	4.66 (0.16)	0.36 (0.23)	3.25 (0.12)	0.31 (0.23)	0.72 (0.12)	2.82 (0.10)	3.00 (0.23)
Sd	0.11	0.22	0.11	0.22	0.07	0.23	0.20
RMSE	0.12	0.23	0.16	0.22	0.07	0.63	0.54
$n = m = 100$; $S = 50$							
Estimate	4.62 (0.11)	0.36 (0.14)	3.24 (0.08)	0.33 (0.15)	0.70 (0.08)	2.60 (0.09)	2.81 (0.26)
Sd	0.06	0.18	0.07	0.18	0.05	0.17	0.14
RMSE	0.06	0.18	0.13	0.18	0.05	0.40	0.34

C.1.6 Performance of iREM with habitat

The results from fitting iREM with habitat and iREM for data simulated from a Poisson iREM with habitat is given in this section. We set $\mu_x = 0.71$; $\beta_1 = 3.56$, $\beta_2 = 1.61$, $\beta_3 = 0.18$, $\beta_4 = 0.40$; $D_1 = 35.16$, $D_2 = 175.91$, $D_3 = 42.10$, $D_4 = 52.46$; $D_T = 68.30$. The results suggests that the model is working well under these conditions. The results of using larger regression coefficients, and hence, higher density values are given in Table C.1.15 in Section 4.2.6.

Table C.1.15: Average parameter estimates from a Poisson iREM with habitat fitted to encounters simulated from a Poisson REM (standard error in parentheses). Animal speed data is assumed to follow a lognormal model. The expected animal speed, regression coefficients, and hence density within habitats and across habitats are set to $\mu_x = 0.71$; $\beta_1 = 3.56$, $\beta_2 = 1.61$, $\beta_3 = 0.18$, $\beta_4 = 0.40$; $D_1 = 35.16$, $D_2 = 175.91$, $D_3 = 42.10$, $D_4 = 52.46$; $D_T = 68.30$, respectively. The sample sizes are n trap days, and m animal speed data. The Standard deviation (Sd) and Root Mean Square Error (RMSE) are also given.

	iREM with habitat						iREM	
	Habitat 1 $\hat{\beta}_1$	Habitat 2 $\hat{\beta}_2$	Habitat 3 $\hat{\beta}_3$	Habitat 4 $\hat{\beta}_4$	$\hat{\mu}_x$	$\hat{D_T}$	$\hat{\mu}_x$	$\hat{D}$
Estimates of β_p								
$n = m = 40$								
Estimate	3.51 (0.37)	1.65 (0.40)	0.17 (0.50)	0.41 (0.47)	-	-	-	-
Sd	0.37	0.40	0.23	0.47	-	-	-	-
RMSE	0.38	0.41	0.23	0.47	-	-	-	-
$n = m = 100$								
Estimate	3.54 (0.23)	1.62 (0.24)	0.18 (0.30)	0.40 (0.29)	-	-	-	-
Sd	0.24	0.24	0.31	0.28	-	-	-	-
RMSE	0.24	0.24	0.31	0.28	-	-	-	-
Estimates of D_p and μ								
$n = m = 40$								
Estimate	33.52 (12.53)	173.84 (133.85)	39.83 (34.69)	50.66 (42.70)	0.71 (0.09)	66.47 (11.43)	0.71 (0.06)	77.29 (11.46)
Sd	13.15	34.58	12.97	15.13	0.06	11.38	0.07	11.45
RMSE	13.17	34.60	12.97	15.13	0.06	11.52	0.06	14.56
$n = m = 100$								
Estimate	34.54 (7.96)	175.27 (83.19)	41.36 (22.06)	51.77 (26.86)	0.71 (0.06)	67.64 (7.23)	0.71 (0.06)	76.59 (7.18)
Sd	7.97	20.13	9.05	9.31	0.04	7.18	0.07	7.61
RMSE	7.97	20.13	9.05	9.31	0.04	7.21	0.05	11.25

C.2 Application of iREM with habitat to WWAP data set

This section is linked to Section 4.3, in Chapter 4. Estimates of the density within and across habitats are given in Section C.2.1 for the wallaby species, while Section C.2.2 gives the estimates for the water deer species. It then goes on to give the results for the muntjac species in Section C.2.3 and the mara species in Section C.2.4.

C.2.1 Estimated parameters for the wallaby species

The results from a gamma model are given in Table 4.3.1, in Section 4.3.1. Here we give the results from fitting a lognormal model (Table C.2.1) and a Weibull model (Table C.2.2). Estimates of the density from the alternative speed data models show minimal difference, which is confirmed by the ΔAIC values in Table 4.3.2, Section 4.3.1. The results also show that incorporating habitat in iREM is highly relevant as estimates of the density improved. There was support for models allowing for variation in the encounter data and iREM with habitat model over iREM.

Table C.2.1: Estimations from iREM with habitat compared with estimations from iREM and the density from the census of the wallaby species (standard error in parentheses). Animal speed is assumed to follow a lognormal model.

	Habitat				**Total density**	
	Central Park	Downs	Institute Paddock	Old Farm	**iREM with habitat**	**iREM**
Census density ((D), km^2)	96	1101	760	803	468	468
Poisson	260 (142.41)	648 (458.18)	915 (649.58)	629 (448.80)	463 (245.88)	636 (333.94)
NB	260 (151.28)	648 (561.67)	915 (815.77)	629 (557.44)	463 (249.24)	635 (340.94)
ZIP	381 (251.67)	753 (647.92)	1077 (929.32)	713 (617.60)	582 (370.60)	739 (392.60)
ZINB	257 (152.17)	638 (557.91)	905 (814.49)	618 (552.68)	457 (250.16)	633 (340.57)

Table C.2.2: Estimations from iREM with habitat compared with estimations from iREM and the density from the census of the wallaby species (standard error in parentheses). Animal speed is assumed to follow a Weibull model.

	Habitat				**Total density**	
	Central Park	Downs	Institute Paddock	Old Farm	**iREM with habitat**	**iREM**
Census density ((D), km^2)	96	1101	760	803	468	468
Poisson	270 (113.64)	673 (390.26)	950 (553.88)	653 (383.11)	480 (191.70)	659 (261.72)
NB	243 (140.78)	804 (693.78)	1092 (970.10)	680 (597.60)	514 (275.48)	659 (267.76)
ZIP	367 (149.10)	715 (432.17)	1000 (607.64)	708 (431.52)	555 (204.46)	768 (305.42)
ZINB	299 (139.05)	772 (577.96)	1018 (786.22)	707 (542.91)	532 (217.75)	688 (286.72)

C.2.2 Estimated density for the water deer species

In this Section we give the results for the water species where animal speed is assumed to follow a lognormal model or a gamma model. The results from the Weibull model (Table 4.3.3) are given in Section 4.3.2, in the Chapter 4.

Estimates of density from the gamma model and Weibull model are similar with inflated density in Central park and an underestimation of the density in Institute Paddock. These results might suggest that the water deer frequented Central Park during the survey period. The lognormal model performed poorly compared with the gamma model and Weibull model, particularly when allowing for variation in the encounter data. However, this may be a reflection of the small speed data set for which a lognormal model may not be appropriate to fit. There is support for incorporating habitat in iREM, and for accounting for variation in the encounter data as shown by the ΔAIC values in Table 4.3.4, Section 4.3.2.

Table C.2.3: Estimations from iREM with habitat compared with estimations from iREM and the density from the census of the water deer species (standard error in parentheses). Animal speed is assumed to follow a lognormal model

	Habitat				**Total density**	
	Central Park	Downs	Institute Paddock	Old Farm	**iREM with habitat**	**iREM**
Census density ((D), km^2)	72	73	36	577	119	119
Poisson	143 (92.19)	49 (40.91)	6 (7.38)	177 (142.88)	109 (69.63)	59 (47.60)
NB	136 (96.86)	42 (42.69)	5 (6.64)	164 (164.49)	102 (71.19)	60 (48.13)
ZIP	162 (123.50)	54 (53.69)	6 (8.02)	187 (173.67)	122 (92.24)	113 (91.00)
ZINB	97 (80.84)	34 (38.17)	4 (5.94)	122 (137.57)	74 (60.78)	75 (63.60)

Table C.2.4: Density estimates from iREM with habitat compared with density estimates from iREM and the density from the census of the water deer species (standard error in parentheses). Animal speed is assumed to follow a gamma model

	Habitat				**Total density**	
	Central Park	Downs	Institute Paddock	Old Farm	**iREM with habitat**	**iREM**
Census density ((D), km^2)	72	73	36	577	119	119
Poisson	149 (62.34)	51 (31.74)	6 (6.23)	187 (108.59)	114 (46.66)	91 (36.76)
NB	148 (68.44)	52 (39.40)	6 (6.83)	185 (140.64)	113 (49.37)	91 (38.32)
ZIP	202 (85.13)	74 (48.18)	9 (8.77)	247 (146.22)	155 (63.71)	174 (70.35))
ZINB	149 (68.55)	52 (39.32)	7 (6.81)	141 (108.98)	114 (49.40)	115 (55.95)

C.2.3 Estimated density of the muntjac species

Table 4.3.5, in Section 4.3.3, compares estimated density from iREM with habitat model with an iREM model of the muntjac species where animal speed is assumed to follow a Weibull distribution. Here, in Tables C.2.5 and C.2.6, we give the estimates of the density of the muntjac species where animal speed is assumed to follow a lognormal distribution or a gamma distribution, respectively.

Like the wallaby and water deer species, estimates of the density of the muntjac species improved when habitat is included in the model. For all iREM with habitat models estimated density is non-zero in Institute Paddock. Note that the census for the muntjac species is zero in Institute Paddock. These results suggest that a flexible model that incorporate habitat-specific covariates is the best model to estimate densities for species where the data set is small. There is support for models allowing for variation in encounter data within habitats as confirmed by the ΔAIC values (Table 4.3.6, Section 4.3.3)

Table C.2.5: Estimated density from iREM with habitat is compared with estimated density from an iREM and the density from the census of the muntjac species (standard error in parentheses). Animal speed is assumed to follow a lognormal model

	Habitat				**Total density**	
	Central Park	Downs	Institute Paddock	Old Farm	**iREM with habitat**	**iREM**
Census density ((D), km^2)	18	6	0	22	13	13
Poisson	24 (7.79)	6 (3.75)	1 (0.95)	4 (3.18)	15 (4.74)	7 (2.21)
NB	21 (9.18)	5 (4.17)	1 (0.95)	4 (3.51)	13 (5.63)	7 (2.43)
ZIP	34 (12.05)	9 (6.78)	1 (1.53)	7 (5.40)	22 (7.49)	22 (7.55)
ZINB	25 (10.61)	7 (5.82)	1 (1.14)	5 (4.97)	16 (6.47)	15 (10.68)

Table C.2.6: Density estimates from iREM with habitat compared with density estimates from iREM and the density from the census of the muntjac species (standard error in parentheses). Animal speed is assumed to follow a gamma model

	Habitat				**Total density**	
	Central Park	Downs	Institute Paddock	Old Farm	**iREM with habitat**	**iREM**
Census density ((D), km^2)	18	6	0	22	13	13
Poisson	25 (7.05)	6 (3.63)	1 (0.95)	4 (3.11)	15 (4.25)	8 (1.94)
NB	24 (8.52)	6 (4.55)	1 (1.00)	5 (3.93)	15 (5.01)	8 (2.19)
ZIP	34 (12.05)	9 (6.78)	1 (1.52)	7 (5.39)	22 (6.86)	23 (6.92)
ZINB	36 (11.12)	10 (6.63)	1 (1.53)	7 (5.31)	23 (6.86)	16 (10.65)

C.2.4 Estimated density of the mara species

The results from the lognormal model (Table C.2.7) and the Weibull model (Table C.2.8) are given here. The results from the gamma model (Table 4.3.7) are given in Section 4.3.4.

For all models, estimated density improved when habitat is incorporated, even in the habitat where there are no census records, an iREM with habitat obtained non-zero estimates of the density within the habitat. However, there was support for the underlying Poisson REM without habitat with the gamma model as the best model that explains the speed data of the mara species.

Table C.2.7: Estimated density from an iREM with habitat model compared with estimated density from an iREM and the density from the census of the mara species (standard error in parentheses). Animal speed is assumed to follow a lognormal model.

	Habitat				**Total density**	
	Central Park	Downs	Institute Paddock	Old Farm	**iREM with habitat**	**iREM**
Census density ((D), km^2)	108	30	7	0	68	68
Poisson	5 (5.29)	2 (4.23)	1 (3.08)	2 (4.66)	4 (3.60)	3 (2.32)
NB	5 (5.29)	2 (4.23)	1 (3.08)	2 (4.68)	4 (3.60)	3 (2.32)
ZIP	5 (5.29)	2 (4.23)	1 (3.08)	2 (4.66)	4 (3.60)	3 (2.52)
ZINB	5 (5.31)	2 (4.23)	1 (3.06)	2 (4.68)	4 (3.62)	3 (2.36)

C.2.5 Analysis of data for species with observed zero speed

This section is linked to Section 4.3.5. Here we give the results for the species with observed zero speed of movement. The results of the wallaby species are given in Table C.2.9. The results of the water deer species are given in Table C.2.10, while the results of the mara species are given in Table C.2.11.

Table C.2.8: Estimated density from an iREM with habitat compared with estimated density from an iREM and the density from the census of the mara species (standard error in parentheses). Animal speed is assumed to follow a Weibull model.

	Habitat				**Total density**	
	Central Park	Downs	Institute Paddock	Old Farm	**iREM with habitat**	**iREM**
Census density ((D), km^2)	108	30	7	0	68	68
Poisson	7 (5.31)	3 (4.95)	2 (3.69)	3 (5.51)	5 (3.37)	4 (2.02)
NB	7 (5.31)	3 (4.95)	2 (3.67)	3 (5.50)	5 (3.37)	4 (2.12)
ZIP	8 (7.90)	4 (6.75)	2 (4.84)	4 (6.84)	6 (5.36)	4 (2.02)
ZINB	8 (6.23)	3 (5.47)	2 (4.07)	4 (6.08)	6 (4.07)	4 (2.02)

Table C.2.9: Estimated density of the wallaby species where animal speed follows a ZAGA model (standard error in parentheses). Estimated density from iREM with habitat is compared with estimates of the density from iREM , and the density from the census.

	Habitat				**Total density**	
	Central Park	Downs	Institute Paddock	Old Farm	**iREM with habitat**	**iREM**
Census density (D, animals km^2)	96	1101	760	803	468	468
Poisson	334 (143.12)	834 (489.35)	1177 (694.46)	809 (480.31)	595 (241.88)	816 (330.23)
NB	76 (27.53)	808 (535.47)	795 (528.89)	734 (496.99)	391 (103.26)	816 (337.66)
ZIP	322 (426.12)	647 (986.16)	948 (1447.04)	625 (956.59)	501 (657.38)	952 (385.18)
ZINB	341 (189.19)	744 (630.43)	1123 (982.28)	779 (677.09)	570 (287.32)	790 (330.00)

Table C.2.10: Estimated density of the water deer species where animal speed follows a ZAGA model (standard error in parentheses). Estimated density from iREM with habitat is compared with estimated density from iREM, and the density from the census.

	Habitat				**Total density**	
	Central Park	Downs	Institute Paddock	Old Farm	**iREM with habitat**	**iREM**
Census density (D, animals km^2)	72	73	36	577	119	119
Poisson	70 (56.70)	102 (118.93)	71 (87.61)	530 (603.90)	124 (93.97)	152 (72.73)
NB	252 (133.56)	85 (71.17)	10 (12.26)	308 (254.78)	192 (97.13)	152 (74.87)
ZIP	361 (183.57)	138 (101.31)	17 (18.95)	445 (302.68)	279 (139.34)	310 (145.41)
ZINB	248 (130.94)	86 (71.30)	10 (12.04)	311 (256.58)	190 (95.80)	192 (105.66)

Table C.2.11: Estimated density of the mara species where animal speed follows a ZAGA model (standard error in parentheses). Estimated desnity from iREM with habitat is compared with estimated density from iREM, and the density from the census.

	Habitat				**Total density**	
	Central Park	Downs	Institute Paddock	Old Farm	**iREM with habitat**	**iREM**
Census density (D, animals km^2)	108	30	7	0	68	68
Poisson	11 (8.01)	4 (6.97)	3 (5.11)	5 (7.81)	8 (5.11)	5 (3.06)
NB	10 (7.81)	5 (7.13)	3 (5.18)	5 (7.63)	8 (5.02)	5 (3.06)
ZIP	11 (8.68)	5 (7.67)	3 (5.80)	5 (8.53)	8 (5.65)	5 (3.62)
ZINB	10 (7.94)	5 (7.99)	3 (6.02)	5 (8.91)	7 (5.22)	5 (3.13)

C.3 Example R codes for iREM with random effect

Example

This example is a survey design with c = 3 camera traps, which were set up for $n = 10$ camera trap days. The parameters required to computer the encounter rate, λ_{ij}, are $r = 0.012$ (km) and $\theta = 0.175$ (radians); true density, $D = 100$; camera trap time period, $t = 1$ (day), and mean animal speed, $\bar{v} = 0.60$ km/day^{-1}.

```
###generating encounters, Y, on each camera from a Poisson model
###with a random effect of the camera location drawn from a
###normal distribution with mean 0 and variance sigmab^2.
Camera   <- 3            #number of camera traps
n        <- 10           #number of camera trap days (same for all cameras)
Density  <- 100          #starting value for density
sigmab   <- 0.10         #starting value for random effect
theta    <- 0.175        #fixed detection angle
r        <- 0.012        #fixed detection distance
t        <- 1            #camera trap time period
v.bar    <- 0.60         #mean animal speed
###creating vectors to store information
 Y       <- c()          #stores encounters from
                         #all cameras as a vector
 b       <- numeric(k)   #stores the random effects
 Dens    <- numeric(k)   #stores the density values
 lambda  <- numeric(k)   #stores the encounter rates


  for(k in 1:Camera){
    b[k]      <- rnorm(1,0,sigmab)
    Dens[k]   <- exp(log(Density) + b[k])
    lambda[k] <- ((2 + theta)/pi)*r*t*v.bar.Dens[k]
    Ydata     <- rpois(n,lambda[k])
    Y         <-c(Y,Ydata)}
```

Example continued

If the number of camera trap days varies for each trap, we show how this can be accommodated in the model in the R code below.

```
###generating encounters, Y, on each camera from a Poisson model
###with a random effect of the camera location drawn from a
###normal distribution with mean 0 and variance sigmab^2.
Camera   <- 3            #number of camera traps
Density  <- 100          #starting value for density
sigmab   <- 0.10         #starting value for random effect
theta    <- 0.175        #fixed detection angle
r        <- 0.012        #fixed detection distance
t        <- 1            #camera trap time period
v.bar    <- 0.60         #mean animal speed

###If the number of days each camera was out varies, we can accommodate
###this also in the model. First we can generate the number of days each
###camera was out from some discrete distribution, say a zero truncated
###Poisson model N is a vector of camera days, where 2 represents the
### minimum number of days and 8 represents the mean camera trap days

N = rktp(Camera, 2 ,8)

    for(k in seq_along(N)){
    b[k]    = rnorm(1,0,sigmaE)
    Dens[k] = exp(log(Density) + b[k])
    lambda[k]= ((2+theta1)/pi)*r*t*velocity*Dens[k]
    Y=rpois(N[k],lam)
    Ydata = c(Ydata,Y)}
```

Example continued

This example consists of sample size of $m = 10$ with speeds (in km/day^{-1}) following a lognormal distribution with parameters $\epsilon = -1.09$ and, the standard error from the normal distribution, $\sigma = 1.08$.

```
###generating animal speed data from a lognormal model
m       <- 10          #number of animal speed data
epsilon <- -1.09       #starting value for mean of the logarithm
sigma   <-  1.08       #starting value for variance from
                       #the normal distribution
X <- rlnorm(m, epsilon, sigma)
```

C.4 Application of iREM with random effect to WWAP data set

In this section of the appendix we give results from iREM with random effect models fitted to real data at WWAP for the four species of interest. Section C.4.1 gives estimates of the density of the wallaby species while Section C.4.2 gives estimates of the density of the water deer species. It then goes on, in Section C.4.3, to give estimates of the density of the muntjac species and finally, Section C.4.4 gives estimates of the density of the mara species.

C.4.1 Analysis of the wallaby species

In this section we give the results from fitting a lognormal model (Table C.4.1) and a Weibull model (Table C.4.2) to animal speed data. The results from fitting a gamma model (Table 4.5.1) are given in Section 4.5.1. The results show that the gamma model and lognormal model obtained stable estimates of the density with the lognormal obtaining smaller differences in estimated density and the density from the census. The Weibull model performed poorly obtaining large differences between estimated density and density from the census when camera random effect is incorporated in the model. The results also show that there is support for models accounting for variation in the encounter data but an iREM with habitat has been selected as the best model that explains the wallaby data set (Table 4.5.2, Section 4.5.1).

Table C.4.1: Estimated parameters for the wallaby species from iREM with random effect models where animal speed data is assumed to follow a lognormal distribution (standard error in parentheses).

	Count data models			
	Poisson	NB	ZIP	ZINB
Census density D = 468				
$\hat{D}$	483 (260.48)	597 (325.36)	711 (378.54)	594 (324.82)
$\hat{\mu}_x$	0.91 (0.48)	0.91 (0.48)	0.91 (0.48)	0.91 (0.48)
$\hat{\nu}$	1.16 (0.29)	1.16 (0.29)	1.16 (0.29)	1.16 (0.29)
$\hat{\sigma}_b$	0.54 (0.11)	0.30 (0.18)	0.53 (0.08)	0.31 (0.18)
$\hat{k}$	-	1.03 (0.17)	-	0.97 (0.16)
$\hat{\rho}$	-	-	0.12 (0.03)	0.0002 (0.001)
$(-\ell)$	549.93	389.28	516.91	389.28
AIC	1109.86	790.56	1049.82	792.56
ΔAIC	319.30	0	258.64	2

Table C.4.2: Estimated parameters for the wallaby species from iREM with random effect models where animal speed data is assumed to follow a Weibull distribution (standard error in parentheses).

	Count data models			
	Poisson	NB	ZIP	ZINB
Census density D = 468				
$\hat{D}$	691 (276.58)	620 (258.36)	738 (295.60)	610.50 (264.80)
$\hat{\mu}_x$	0.87 (0.35)	0.87 (0.35)	0.87 (0.35)	0.90 (0.38)
$\hat{\nu}$	0.89 (0.23)	0.89 (0.23)	0.89 (0.23)	0.86 (0.23)
$\hat{\sigma}_b$	0.60 (0.11)	0.30 (0.18)	0.53 (0.08)	0.26 (0.22)
$\hat{k}$	-	1.03 (0.17))	-	1.01 (0.17)
$\hat{\rho}$	-	-	0.12 (0.03)	0.005 (0.01)
$(-\ell)$	551.44	389.79	517.42	389.71
AIC	1112.88	791.58	1058.84	793.42
ΔAIC	321.30	0	267.26	1.84

C.4.2 Analysis of the water deer data set

In this section we give the results of the water deer species. The results from fitting a gamma model to animal speed data are given in Table 4.5.3, Section 4.5.2, while Table C.4.3 gives the results fitting a lognormal model to animal speed data and Table C.4.4 gives the results from fitting a Weibull model to animal speed data. The results show that the lognormal model generally obtained better estimates of the density compared with the other speed data models. The results also show that accounting for the variation from the random location of camera traps in estimating density is important when interpreting the results. The ΔAIC values suggest that an iREM with random effect model best explains the water deer data set.

Table C.4.3: Estimated parameters for the water deer species from iREM with random effect models where animal speed data is assumed to follow a lognormal distribution (standard error in parentheses).

	Count data models			
	Poisson	NB	ZIP	ZINB
Census density D = 119				
$\hat{D}$	122 (97.53)	117 (94.10)	46 (37.73)	121 (97.33)
$\hat{\mu}_x$	2.99 (2.38)	2.55 (1.89)	3.01 (2.41)	2.98 (2.38)
$\hat{\nu}$	1.39 (0.40)	1.47 (0.40)	1.39 (0.40)	1.39 (0.40)
$\hat{\sigma}_b$	1.76 (0.18)	1.93 (0.28)	1.54 (0.31)	1.80 (0.25)
$\hat{k}$	-	2.21 (1.04)	-	0.20 (0.13)
$\hat{\rho}$	-	-	0.14 (0.07)	0.09 (0.08)
$(-\ell)$	232.04	225.46	221.81	223.91
AIC	474.08	462.92	455.62	461.82
ΔAIC	18.46	7.30	0	6.20

Table C.4.4: Estimated parameters for the water deer species from iREM with random effect models where animal speed data is assumed to follow a Weibull distribution (standard error in parentheses).

	Count data models			
	Poisson	NB	ZIP	ZINB
Census density D = 119				
$\hat{D}$	188 (68.11)	144 (48.28)	71 (28.56)	186 (70.32)
$\hat{\mu}_x$	1.93 (0.68)	2.37 (0.68)	1.94 (0.69)	1.94 (0.69)
$\hat{\nu}$	1.15 (0.41)	1.65 (0.51)	1.15 (0.41)	1.15 (0.41)
$\hat{\sigma}_b$	1.77 (0.18)	1.51 (0.30)	1.54 (0.31)	1.80 (0.25)
$\hat{k}$	-	2.07 (0.81)	-	0.20 (0.13)
$\hat{\rho}$	-	-	0.14 (0.07)	0.09 (0.08)
$(-\ell)$	230.73	225.59	220.50	222.60
AIC	471.46	463.18	453.00	459.20
ΔAIC	18.46	10.18	0	6.20

C.4.3 Analysis of the muntjac species data set

In this section we give estimates of the density of the muntjac species within and across habitats where animal speed is assumed to follow a lognormal model (Table C.4.5) or a Weibull model (Table C.4.6). The results from fitting a gamma model to animal speed are given in Table 4.5.5, in Section 4.3.3. The results show that the difference in parameter estimates from the alternative speed data models is minimal. A Poisson iREM does well in estimating the density compared with the other iREM but the ΔAIC values suggest that a NB iREM best explains the muntjac species data. Also, accounting for the variation the from the random location of camera traps has proven to be relevant in this case.

Table C.4.5: Estimated parameters for the muntjac species from iREM with random effect models where animal speed data is assumed to follow a lognormal distribution (standard error in parentheses).

	Count data models			
	Poisson	NB	ZIP	ZINB
Census density D = 13				
$\hat{D}$	14 (6.23)	4 (1.88)	4 (2.64)	3 (2.48)
$\hat{\mu}_x$	8.27 (2.24)	8.52 (2.25)	8.50 (2.24)	8.52 (2.25)
$\hat{\nu}$	0.78 (0.18)	0.74 (0.17)	0.74 (0.17)	0.74 (0.17)
$\hat{\sigma}_b$	1.81 (0.55)	1.33 (0.37)	1.34 (0.49)	1.33 (0.37)
$\hat{k}$	-	1.70 (1.38)	-	0.39 (0.76)
$\hat{\rho}$	-	-	0.27 (0.21)	0.11 (0.41)
$(-\ell)$	134.22	126.81	127.17	126.78
AIC	278.44	265.62	266.34	267.56
ΔAIC	12.82	0	0.72	1.94

Table C.4.6: Estimated parameters for the muntac species from iREM with random effect models where animal speed data is assumed to follow a Weibull distribution (standard error in parentheses).

	Count data models			
	Poisson	NB	ZIP	ZINB
Census density D = 13				
$\hat{D}$	15 (4.27)	4 (1.86)	6 (2.63)	3 (3.68)
$\hat{\mu}_x$	8.32 (1.76)	8.32 (1.76)	8.32 (1.76)	8.26 (1.75)
$\hat{\nu}$	1.53 (0.36)	1.53 (0.36)	1.53 (0.36)	1.52 (0.36)
$\hat{\sigma}_b$	2.17 (0.39)	1.33 (0.36)	1.86 (0.82)	1.33 (0.36)
$\hat{k}$	-	1.70 (1.38)	-	0.43 (1.66)
$\hat{\rho}$	-	-	0.24 (0.21)	0.08 (0.95)
$(-\ell)$	133.64	126.81	127.42	126.78
AIC	278.44	265.62	266.34	267.56
ΔAIC	12.82	0	0.72	1.94

C.4.4 Analysis of the mara species data set

Estimates of the density of the mara species from fitting a gamma model are given in Table 4.5.7, in Section 4.5.4 while the results from fitting a lognormal model (Table C.4.7) and a Weibull model (Table C.4.8) are given here. Estimated density from the three speed data models show minimal differences but the density is underestimated in all cases since the camera placement strategies adopted in the area where the mara frequented (Central Park) were not randomized, which is critical in REM. There is support for a Poisson iREM as shown by the ΔAIC values in Table 4.5.8. Note that there were a small number of encounter records of the mara species and therefore it may not be possible to make precise inference about the population parameters; see for instance the huge estimate of the parameter κ, which suggests that the NB iREM may be non-identifiable. Also the relationship between a Poisson iREM and a NB iREM exists through the variance, and larger values of κ indicate that the variance of a NB iREM

would be approximately equal to the mean, hence, a NB iREM tends to a Poisson iREM.

Table C.4.7: Estimated parameters for the mara species from iREM with random effect models where animal speed data is assumed to follow a lognormal distribution (standard error in parentheses).

	Count data models			
	Poisson	NB	ZIP	ZINB
Census density D = 68				
$\hat{D}$	1 (1.98)	1 (1.46)	1 (1.46)	1 (1.44)
$\hat{\mu}_x$	5.03 (4.00)	5.03 (4.00)	5.02 (4.00)	5.09 (4.09)
$\hat{\nu}$	1.46 (0.39)	1.46 (0.39)	1.46 (0.39)	1.47 (0.39)
$\hat{\sigma}_b$	1.34 (0.62)	1.34 (0.62)	1.34 (0.62)	1.34 (0.62)
$\hat{k}$	-	9.66e+09 (1.69e+10)	-	0.001 (0.03)
$\hat{\rho}$	-	-	1e-06 (1.6e-14)	1e-04 (0.01)
$(-\ell)$	48.74	48.74	48.74	48.74
AIC	107.84	109.48	109.48	111.48
ΔAIC	0	2	2	4

Table C.4.8: Estimated parameters for the mara species from iREM with random effect models where animal speed data is assumed to follow a Weibull distribution (standard error in parentheses).

	Count data models			
	Poisson	NB	ZIP	ZINB
Census density D = 68				
$\hat{D}$	2 (1.55)	2 (1.58)	2 (1.62)	2 (1.58)
$\hat{\mu}_x$	3.68 (1.59)	3.68 (1.59)	3.69 (1.61)	3.67 (1.58)
$\hat{\nu}$	0.88 (0.28)	0.88 (0.28)	0.88 (0.28)	0.88 (0.28)
$\hat{\sigma}_b$	1.34 (0.62)	1.34 (0.62)	1.34 (0.62)	1.34 (0.62)
$\hat{k}$	-	7.19e+06 (3.52e+08)	-	4e-04 (0.02)
$\hat{\rho}$	-	-	2e-04 (0.01)	4e-04 (0.02)
$(-\ell)$	48.31	48.31	48.31	48.31
AIC	106.62	108.62	108.62	110.62
ΔAIC	0	2	2	4

Appendix D

The results discussed in Chapter 6, and additional results from Chapter 6 are given in this section of the appendix.

Section D.1 compares estimations of expected animal speed from parametric likelihood-based methods with estimations from nonparametric based methods such as the harmonic mean formula and the standard mean formula. Encounters are simulated from a Poisson REM to test the approximation of the probability of encounter. This section gives the Tables with the results for the gamma model and Weibull model discussed in Section 6.5.1.3 in the main thesis.

Section D.2 gives the results from fitting a lognormal model and a Weibull model model to animal speed data where encounters are drawn from a NB REM. The findings from these models and from fitting a gamma model to animal speed are discussed in Section 6.5.2 in the main thesis.

In Section D.3 we give the results from fitting a lognormal model and a Weibull model to animal speed data for encounters simulated from a ZIP REM. The findings from these models and from fitting a gamma model to animal speed can be found in Section 6.5.3 in the main thesis.

Section D.4 gives the results from fitting a lognormal model and a Weibull model to animal speed data where encounters are drawn from a ZINB REM. The findings from

these models and from fitting a gamma model to animal speed are discussed in Section 6.5.4 in the main thesis. Finally, Section D.5 gives estimates from fitting size biased models to real data at BCI, Panama.

D.1 Comparing estimations of expected speed from parametric models and nonparametric methods for encounters simulated from a Poisosn REM

This section gives the results from the gamma models and Weibull models discussed in Section 6.5.1.3 in the main thesis. In Section 6.5.1.3, we discussed the findings from fitting a size biased lognormal model and a lognormal model to animal speed data. Estimations from these models are also compared with estimations from the harmonic mean formula and the standard mean formula. The results show that under certain conditions (small difference between ν^2 and ϵ) estimations from the size biased parametric likelihood-based method are approximately equal to estimations from the nonparametric harmonic mean method. But when this difference increases it would be more appropriate to use a size biased parametric likelihood-based method to estimate average speed. Table D.1.1 gives the results from fitting a size biased gamma model and a standard gamma model to animal speed data. We also compare the estimated harmonic mean and the estimated standard (arithmetic) mean. Estimates obtained from the size biased gamma model are better than the estimates from the standard model. However, the harmonic mean formula underestimates the expected animal speed, and more so when expected speed is low. Hence, we would recommend a parametric likelihood based approach to estimate animal speeds accounting for the bias is the speed of faster moving animals, if the data is assumed to follow a gamma distribution. Table D.1.2 gives the results from fitting a size biased Weibull model and a standard Weibull model to animal speed data. These results are compared with estimates from the harmonic mean formula and the standard mean formula. Like the gamma model, the size biased Weibull model performs better than the standard Weibull model in obtaining relatively accurate and stable estimations of expected animal speed. But the harmonic mean method underestimates the expected speed, and therefore, we would not recommend the use of the harmonic mean method to estimate expected animal speed in practice if the speed data follows a Weibull model.

Table D.1.1: Estimates of average expected mean speed μ_x (in ms^{-1}) for 100 simulation runs from a size biased gamma distribution compared with estimates from a gamma distribution, and nonparametric methods. The Standard deviation (Sd) and Root Mean Square Error (RMSES) are also given.

				Parametric Methods						**Nonparametric Methods**					
Population size	True value			size biased gamma			gamma			Harmonic			Standard		
M	μ_x	ν	α	$\hat{\mu}_x$	Sd	RMSE	$\hat{\mu}_x$	Sd	RMSE	$\hat{\mu}_x$	Sd	RMSE	$\hat{\mu}_x$	Sd	RMSE
$t = 1$															
40	0.45	45	20.25	0.46 (0.02)	0.02	0.03	0.48 (0.02)	0.03	0.04	0.43	0.02	0.03	0.451	0.02	0.02
160	0.45	45	20.25	0.45 (0.01)	0.01	0.01	0.47 (0.01)	0.01	0.02	0.43	0.007	0.02	0.450	0.008	0.008
280	0.45	45	20.25	0.45 (0.01)	0.01	0.01	0.47 (0.01)	0.01	0.03	0.43	0.006	0.02	0.452	0.006	0.007
40	0.45	3	1.35	0.47 (0.14)	0.14	0.14	0.77 (0.12)	0.20	0.37	0.16	0.06	0.30	0.45	0.06	0.06
160	0.45	3	1.35	0.44 (0.002)	0.07	0.08	0.77 (0.06)	0.10	0.33	0.14	0.04	0.31	0.45	0.03	0.03
280	0.45	3	1.35	0.45 (0.05)	0.06	0.06	0.77 (0.05)	0.07	0.33	0.13	0.03	0.32	0.45	0.02	0.02
40	0.15	15	2.25	0.17 (0.04)	0.04	0.05	0.22 (0.04)	0.05	0.09	0.09	0.02	0.06	0.15	0.001	0.02
160	0.15	15	2.25	0.16 (0.03)	0.03	0.03	0.22 (0.03)	0.03	0.08	0.08	0.01	0.07	0.15	0.01	0.01
280	0.15	15	2.25	0.15 (0.02)	0.02	0.02	0.21 (0.02)	0.02	0.07	0.08	0.01	0.07	0.15	0.01	0.01
40	0.15	2.5	0.375	0.23 (0.23)	0.18	0.20	0.52 (0.13)	0.22	0.43	0.0007	0.001	0.15	0.15	0.04	0.04
160	0.15	2.5	0.375	0.17 (0.12)	0.12	0.12	0.56 (0.10)	0.15	0.43	0.0001	0.0004	0.15	0.15	0.01	0.02
280	0.15	2.5	0.375	0.17 (0.09)	0.09	0.09	0.55 (0.08)	0.11	0.42	3.6e-05	0.0001	0.15	0.15	0.01	0.01

Table D.1.2: Estimates of average expected mean speed μ_x (in ms^{-1}) for 100 simulation runs from a size biased Weibull distribution compared with estimates from a Weibull distribution, and nonparametric methods. The Standard deviation (Sd) and Root Mean Square Error (RMSES) are also given.

				Parametric Methods						**Nonparametric Methods**					
Population size	True value			size biased Weibull			Weibull			Harmonic			Standard		
M	$\hat{\mu}_x$	ν	β	$\hat{\mu}_x$	Sd	RMSE	$\hat{\mu}_x$	Sd	RMSE	$\hat{\mu}_x$	Sd	RMSE	$\hat{\mu}_x$	Sd	RMSE
$t = 1$															
40	0.45	5.12	0.49	0.45 (0.03)	0.03	0.03	0.47 (0.02)	0.03	0.03	0.424	0.02	0.03	0.449	0.01	0.01
160	0.45	5.12	0.49	0.45 (0.01)	0.02	0.02	0.47 (0.01)	0.01	0.03	0.422	0.01	0.03	0.450	0.008	0.008
280	0.45	5.12	0.49	0.45 (0.01)	0.01	0.01	0.47 (0.009)	0.01	0.03	0.423	0.007	0.03	0.450	0.006	0.006
40	0.45	1.18	0.48	0.47 (0.12)	0.12	0.12	0.74 (0.10)	0.15	0.32	0.13	0.05	0.33	0.45	0.05	0.05
160	0.45	1.18	0.48	0.46 (0.06)	0.07	0.07	0.77 (0.06)	0.08	0.33	0.12	0.04	0.34	0.45	0.03	0.03
280	0.45	1.18	0.48	0.46 (0.05)	0.06	0.06	0.78 (0.04)	0.07	0.33	0.11	0.04	0.34	0.45	0.001	0.02
40	0.15	1.55	0.17	0.17 (0.04)	0.04	0.04	0.21 (0.03)	0.04	0.08	0.09	0.02	0.08	0.15	0.01	0.02
160	0.15	1.55	0.17	0.15 (0.03)	0.03	0.03	0.22 (0.03)	0.03	0.07	0.07	0.02	0.08	0.15	0.01	0.01
280	0.15	1.55	0.17	0.15 (0.02)	0.02	0.02	0.22 (0.02)	0.02	0.07	0.07	0.01	0.08	0.15	0.005	0.005
40	0.15	0.587	0.10	0.22 (0.12)	0.13	0.15	0.51 (0.13)	0.24	0.43	0.004	0.004	0.15	0.143	0.04	0.04
160	0.15	0.587	0.10	0.17 (0.07)	0.07	0.08	0.58 (0.11)	0.21	0.48	0.001	0.001	0.15	0.15	0.02	0.02
280	0.15	0.587	0.10	0.17 (0.06)	0.07	0.08	0.62 (0.09)	0.20	0.52	0.0007	0.001	0.15 0.15	0.01	0.02	

D.2 Simulation results for encounters drawn from a negative binomial REM

This section is linked to Section 6.5.2 in the main thesis where we fit a gamma model, a lognormal model and a Weibull to animal speed data with probability of encounters coming from a NB REM. In this section we give the Tables with the results from fitting a lognormal model and a Weibull model to animal speed data. The results of the actual sample sizes, M^* for the first three simulation runs used in the simulation process are also given in this section.

Table D.2.1 gives the sample sizes from a negative binomial REM for the first 3 simulation runs used in the estimation process. As population size, M increases, the actual number of speeds recorded by the camera traps increase. Also, M^* increases for larger values of expected animal speed, μ_x and camera trap time, t.

Table 6.5.10, given in Section 6.5.2 in the main thesis, gives the results from fitting a gamma model to animal speed data. The results suggest that the size biased gamma model performs better than the standard gamma model. The results also suggest that the approximation works well but as expected animal speed increases, expected encounter rate increases and the size bias approximation does slightly worse. However, the slight increase in RMSE is minimal for larger population sizes, M.

Table D.2.2 gives the results from fitting a lognormal model to the speed data. The approximation works well for small values of expected animal speed, μ_x. The RMSE increases slightly for larger values of μ_x but this slight increase is minimal for larger population sizes. Also, estimations from a size biased lognormal model and a lognormal model are approximately equal, suggesting that the variance from the normal model is small relative to the mean of the logarithm.

Table D.2.3 gives the results from fitting a size biased Weibull model and a standard Weibull model to animal speed data. The results show that the size biased Weibull model performs better than the standard Weibull model, and that the approximation works well.

Table D.2.1: Actual speed data sample sizes M^*, for the first 3 simulation runs used in the estimation process where $\mu_x = 0.15$ and $\mu_x = 0.45$. These are given for $t = 1$, $t = 5$ and speed data population sizes of $M = 40$, $M = 160$ and $M = 280$.

$M = 40$			$M = 160$			$M = 280$		
gamma	lognormal	Weibull	gamma	lognormal	Weibull	gamma	lognormal	Weibull
$t = 1$								
$\mu_x = 0.15$								
7	8	7	28	25	26	48	31	36
10	9	10	19	15	20	30	34	54
7	7	7	24	17	20	33	38	38
$\mu_x = 0.45$								
17	22	24	49	83	79	137	82	103
9	21	10	59	62	74	118	102	129
9	13	35	63	59	68	126	106	83
$t = 5$								
$\mu_x = 0.15$								
29	39	29	102	122	136	262	173	191
17	28	21	113	101	97	194	186	219
17	22	50	99	104	91	154	177	169
$\mu_x = 0.45$								
77	100	86	246	354	383	746	579	446
70	75	82.	305	301	340	479	562	618
78	95	81	419	298	382	515	500	412

Table D.2.2: Average estimates of expected speed μ_x (in ms^{-1}) from a size biased lognormal distribution compared with estimates from the lognormal distribution.

Population size	True value		size biased lognormal			lognormal		
M	μ_x	ν	$\hat{\mu}_x$	Sd	RMSE	$\hat{\mu}_x$	Sd	RMSE
$t = 1$								
40	0.15	15	0.15 (0.001)	0.001	0.001	0.15 (0.001)	0.001	0.001
160	0.15	15	0.15 (0.0002)	0.0003	0.0003	0.15 (0.0002)	0.0003	0.0003
280	0.15	15	0.15 (0.0001)	0.0002	0.0002	0.15 (0.0001)	0.0002	0.0002
40	0.45	3	0.45 (0.01)	0.01	0.01	0.45 (0.01)	0.01	0.01
160	0.45	3	0.45 (0.002)	0.003	0.003	0.45 (0.002)	0.003	0.003
280	0.45	3	0.45 (0.001)	0.001	0.001	0.45 (0.001)	0.001	0.001
$t = 5$								
40	0.15	15	0.15 (0.0004)	0.001	0.001	0.15 (0.0004)	0.001	0.001
160	0.15	15	0.15 (0.0001)	0.0002	0.0002	0.15 (0.0001)	0.0002	0.0002
280	0.15	15	0.15 (0.0001)	0.0001	0.0001	0.15 (0.0001)	0.0001	0.0001
40	0.45	3	0.45 (0.003)	0.01	0.01	0.45 (0.003)	0.01	0.01
160	0.45	3	0.45 (0.001)	0.002	0.002	0.45 (0.001)	0.002	0.002
280	0.45	3	0.45 (0.0004)	0.001	0.001	0.45 (0.0004)	0.001	0.001

Table D.2.3: Average estimates of expected animal speed μ_x (in ms^{-1}) from a size biased Weibull distribution compared with estimates from a Weibull distribution.

Population size	True value		size biased Weibull			Weibull		
M	$\hat{\mu}_x$	ν	$\hat{\mu}_x$	Sd	RMSE	$\hat{\mu}_x$	Sd	RMSE
$t = 1$								
40	0.15	15	0.17 (0.04)	0.06	0.07	0.22 (0.03)	0.06	0.09
160	0.15	15	0.15 (0.03)	0.03	0.03	0.21 (0.02)	0.03	0.07
280	0.15	15	0.15 (0.02)	0.02	0.02	0.21 (0.02)	0.02	0.06
40	0.45	3	0.51 (0.12)	0.19	0.20	0.75 (0.10)	0.20	0.36
160	0.45	3	0.47 (0.06)	0.09	0.09	0.77 (0.06)	0.11	0.34
280	0.45	3	0.46 (0.04)	0.04	0.05	0.77 (0.04)	0.06	0.33
$t = 5$								
40	0.15	15	0.16 (0.02)	0.04	0.04	0.21 (0.02)	0.04	0.07
160	0.15	15	0.15 (0.01)	0.02	0.02	0.21 (0.01)	0.02	0.07
280	0.15	15	0.15 (0.01)	0.02	0.02	0.21 (0.01)	0.02	0.07
40	0.45	3	0.48 (0.05)	0.13	0.14	0.74 (0.05)	0.18	0.34
40	0.45	3	0.46 (0.03)	0.06	0.06	0.76 (0.02)	0.09	0.33
280	0.45	3	0.45 (0.02)	0.06	0.05	0.77 (0.02)	0.08	0.33

D.3 Simulation results for encounters drawn from a ZIP REM

The results given in this section are linked to the results and discussion given in Section 6.5.3 in the main thesis. The actual sample sizes, M^* used in the simulation process for the first three simulation runs are given in Table D.3.1. As population size, M, camera trap time, t and expected animal speed, μ_x increase, M^* increases.

The results from fitting a gamma model to animal speed data are given in Table 6.5.11 in the main thesis. The findings suggest that the approximation of the size bias works, and better estimations are obtained from a size biased gamma model compared with a standard gamma model.

The Tables with the results from fitting a lognormal model and a Weibull model to

animal speed data are given in this section.

Table D.3.2 gives the results from fitting a size biased lognormal model and a lognormal model to animal speed data. Accurate estimations of expected animal speed are obtained from the size biased lognormal model, which suggests that the approximation of the size bias works well. Also, estimations from both models are approximately equal, and the RMSEs are much smaller compared with the RMSEs from the size biased gamma model and the size biased Weibull model.

Table D.3.3 gives the results from fitting a size biased Weibull model and a Weibull model to animal speed data. The performance of the size biased Weibull model is better than a Weibull model, with relatively accurate and stable estimations provided by the size biased model. The results also suggest that the approximation of the size bias works well.

Table D.3.1: Actual speed data sample sizes M^* for the first 3 simulation runs used in the estimation process where $\mu_x = 0.15$ and $\mu_x = 0.45$. These are given for $t = 1$, $t = 5$ and speed data population sizes of $M = 40$, $M = 160$ and $M = 280$.

$M = 40$			$M = 160$			$M = 280$		
gamma	lognormal	Weibull	gamma	lognormal	Weibull	gamma	lognormal	Weibull
$t = 1$								
$\mu_x = 0.15$								
8	9	7	9	9	13	15	11	27
9	7	7	9	14	12	27	20	19
8	7	7	9	10	14	19	13	20
$\mu_x = 0.45$								
13	15	11	30	23	29	59	49	70
7	17	7	28	34	34	57	61	48
7	7	10	25	44	28	51	44	40
$t = 5$								
$\mu_x = 0.15$								
17	21	13	48	46	52	86	75	121
15	16	20	61	59	66	99	99	87
21	12	8	49	64	65	93	78	80
$\mu_x = 0.45$								
25	42	47	134	144	162	314	261	323
42	46	44	152	161	175	293	284	256
50	39	27	162	173	159	247	274	219

Table D.3.2: Average estimates of expected animal speed μ_x (in ms^{-1}) from a size biased lognormal distribution compared with estimates from a lognormal distribution.

Population size	True value		size biased lognormal			lognormal		
M	μ_x	ν	$\hat{\mu}_x$	Sd	RMSE	$\hat{\mu}_x$	Sd	RMSE
$t = 1$								
40	0.15	15	0.15 (0.001)	0.001	0.001	0.15 (0.001)	0.001	0.001
160	0.15	15	0.15 (0.0003)	0.0003	0.0003	0.15 (0.0003)	0.0003	0.0003
280	0.15	15	0.15 (0.0001)	0.0002	0.0002	0.15 (0.0001)	0.0002	0.0002
40	0.45	3	0.45 (0.01)	0.01	0.01	0.45 (0.01)	0.01	0.01
160	0.45	3	0.45 (0.002)	0.002	0.002	0.45 (0.002)	0.003	0.002
280	0.45	3	0.45 (0.001)	0.001	0.001	0.45 (0.001)	0.001	0.001
$t = 5$								
40	0.15	15	0.15 (0.0005)	0.001	0.001	0.15 (0.0005)	0.001	0.001
160	0.15	15	0.15 (0.0001)	0.0002	0.0002	0.15 (0.0001)	0.0002	0.0002
280	0.15	15	0.15 (0.0001)	0.0001	0.0001	0.15 (0.0001)	0.0001	0.0001
40	0.45	3	0.45 (0.003)	0.01	0.01	0.45 (0.003)	0.01	0.01
160	0.45	3	0.45 (0.001)	0.001	0.001	0.45 (0.001)	0.001	0.001
280	0.45	3	0.45 (0.0005)	0.001	0.001	0.45 (0.0005)	0.001	0.001

Table D.3.3: Average estimates of expected animal speed μ_x (in ms^{-1}) from a size biased Weibull distribution compared with estimates from a Weibull distribution.

Population size	True value		size biased Weibull			Weibull		
M	μ_x	ν	$\hat{\mu}_x$	Sd	RMSE	$\hat{\mu}_x$	Sd	RMSE
$t = 1$								
40	0.15	15	0.16 (0.04)	0.05	0.0	0.21 (0.03)	0.04	0.08
160	0.15	15	0.16 (0.03)	0.03	0.03	0.21 (0.02)	0.03	0.07
280	0.15	15	0.15 (0.02)	0.02	0.02	0.21 (0.02)	0.02	0.07
40	0.45	3	0.50 (0.12)	0.16	0.17	0.77 (0.11)	0.18	0.37
160	0.45	3	0.46 (0.06)	0.08	0.08	0.78 (0.06)	0.10	0.35
280	0.45	3	0.46 (0.05)	0.06	0.06	0.77 (0.05)	0.06	0.33
$t = 5$								
40	0.15	15	0.15 (0.02)	0.03	0.03	0.21 (0.02)	0.03	0.07
160	0.15	15	0.15 (0.01)	0.02	0.02	0.21 (0.01)	0.02	0.07
280	0.15	15	0.15 (0.01)	0.01	0.01	0.21 (0.01)	0.01	0.06
40	0.45	3	0.47 (0.05)	0.09	0.09	0.75 (0.05)	0.14	0.33
40	0.45	3	0.46 (0.03)	0.05	0.05	0.78 (0.03)	0.08	0.34
280	0.45	3	0.46 (0.02)	0.04	0.04	0.77 (0.02)	0.05	0.33

D.4 Simulation results for encounters drawn from a ZINB REM

This section is linked to Section 6.5.4 in Chapter 6. The Tables with the results from fitting a size biased and standard lognormal models and Weibull models to animal speed data are given in this section. We also give the actual sample sizes for the first 3 simulation runs used in the simulation.

Table D.4.1, shows the actual sample sizes used in the simulations for the first 3 simulation runs. The results show that as population size, M increases, the actual animal speeds, M^* recorded by the camera traps increases. The results show that the size bias approximation works well, and in this case estimations from the size biased model and the standard model are approximately equal. The RMSEs are small but as μ_x increases, and hence, λ, the size bias approximation does slightly worse since the bias increases

slightly. However, this slight increase in the bias is minimal for larger population sizes

The estimations from a size biased Weibull model, given in Table D.4.2, are better than the estimations from a Weibull model. The bias is smaller for the size biased model, which shows that the approximation of size bias works.

Table D.4.1: Actual speed data sample sizes M^* for the first 3 simulation runs used in the estimation process where $\mu_x = 0.15$ and $\mu_x = 0.45$. These are given for $t = 1$, $t = 5$ and speed data population sizes of $M = 40$, $M = 160$ and $M = 280$.

$M = 40$			$M = 160$			$M = 280$		
gamma	lognormal	Weibull	gamma	lognormal	Weibull	gamma	lognormal	Weibull
$t = 1$								
$\mu_x = 0.15$								
8	9	8	9	10	13	15	9	28
9	7	7	8	14	12	29	20	19
9	8	7	9	9	14	18	12	21
$\mu_x = 0.45$								
14	16	9	30	19	30	55	53	72
7	17	8	28	35	37	58	64	47
9	10	7	27	45	32	51	48	38
$t = 5$								
$\mu_x = 0.15$								
15	22	15	47	42	51	79	73	122
16	16	21	57	60	67	108	101	85
24	14	8	47	64	72	96	75	83
$\mu_x = 0.45$								
19	44	53	144	129	159	308	251	331
39	45	51	161	161	189	294	292	243
53	39	26	159	185	165	249	266	199

Table D.4.2: Average estimates of expected animal speed μ_x (in ms^{-1}) from a size biased lognormal distribution compared with estimates from a lognormal distribution.

Population size	True value		size biased lognormal			lognormal		
M	μ_x	ν	$\hat{\mu}_x$	Sd	RMSE	$\hat{\mu}_x$	Sd	RMSE
$t = 1$								
40	0.15	15	0.15 (0.001)	0.001	0.001	0.15 (0.001)	0.001	0.001
160	0.15	15	0.15 (0.0003)	0.0003	0.0003	0.15 (0.0003)	0.0003	0.0003
280	0.15	15	0.15 (0.0002)	0.0002	0.0002	0.15 (0.0002)	0.0002	0.0002
40	0.45	3	0.45 (0.01)	0.01	0.01	0.45 (0.01)	0.01	0.01
160	0.45	3	0.45 (0.002)	0.002	0.002	0.45 (0.002)	0.002	0.002
280	0.45	3	0.45 (0.001)	0.001	0.001	0.45 (0.001)	0.001	0.002
$t = 5$								
40	0.15	15	0.15 (0.0005)	0.001	0.001	0.15 (0.001)	0.001	0.001
160	0.15	15	0.15 (0.0001)	0.0002	0.0002	0.15 (0.0001)	0.0002	0.0002
280	0.15	15	0.15 (0.0001)	0.0001	0.0001	0.15 (0.0001)	0.0001	0.0001
40	0.45	3	0.45 (0.003)	0.01	0.01	0.45 (0.003)	0.01	0.01
160	0.45	3	0.45 (0.001)	0.0001	0.002	0.45 (0.001)	0.002	0.002
280	0.45	3	0.45 (0.0005)	0.001	0.001	0.45 (0.0005)	0.001	0.001

Table D.4.3: Average estimates of expected animal speed μ_x (in ms^{-1}) from a size biased Weibull distribution compared with estimates from a Weibull distribution.

Population size	True value		size biased Weibull			Weibull		
M	$\hat{\mu}_x$	ν	$\hat{\mu}_x$	Sd	RMSE	$\hat{\mu}$	Sd	RMSE
$t = 1$								
40	0.15	15	0.17 (0.04)	0.05	0.05	0.21 (0.03)	0.05	0.08
160	0.15	15	0.16 (0.03)	0.03	0.03	0.22 (0.02)	0.03	0.07
280	0.15	15	0.15 (0.02)	0.02	0.02	0.21 (0.02)	0.02	0.07
40	0.45	3	0.51 (0.12)	0.16	0.17	0.76 (0.11)	0.20	0.36
160	0.45	3	0.47 (0.06)	0.09	0.09	0.78 (0.06)	0.11	0.35
280	0.45	3	0.46 (0.05)	0.06	0.06	0.78 (0.05)	0.07	0.33
$t = 5$								
40	0.15	15	0.16 (0.02)	0.03	0.03	0.21 (0.02)	0.04	0.07
160	0.15	15	0.15 (0.01)	0.02	0.02	0.21 (0.01)	0.02	0.07
280	0.15	15	0.15 (0.01)	0.01	0.01	0.21 (0.01)	0.02	0.07
40	0.45	3	0.47 (0.05)	0.13	0.10	0.73 (0.05)	0.15	0.32
40	0.45	3	0.46 (0.03)	0.06	0.06	0.78 (0.03)	0.09	0.34
280	0.45	3	0.46 (0.02)	0.05	0.05	0.78 (0.02)	0.07	0.33

D.5 Analysis of real data at BCI using size biased sampling

This section is linked to Section 6.6. Here we give the results from fitting a size biased gamma model and a gamma model to speed at BCI in Table D.5.1. In Table D.5.2 we give the results from fitting a size biased lognormal model and a lognormal model to speed data at BCI. The results show big differences between the estimated mean speeds from size biased models and standard models. For a gamma model, when the shape parameter is small (variance is large), the difference between estimates from size biased models and standard models will be large, and a size bias model will perform better. For a lognormal model, when the variance from the normal distribution is large relative to the mean of the logarithm, then the difference between the means of the two models will be large (variance will be large also).

Table D.5.1: Estimates (in ms^{-1}) from a size biased gamma distibution compared with estimates from a gamma distribution (standard errors in parentheses). The negative loglikelihood ($-\ell$) is also given.

Species	**Nonparametric Method**		**Parametric Method**					
	Harmonic mean	Standard mean	size biased gamma			gamma		
	$\hat{\mu}_x$	$\hat{\mu}_x$	$\hat{\mu}_x$	$\hat{\nu}$	$(-\ell)$	$\hat{\mu}_x$	$\hat{\nu}$	$(-\ell)$
ocelot	0.27 (0.03)	0.40 (0.04)	0.27 (0.02)	8.03 (1.22)	-18.37139	0.40 (0.02)	8.03 (1.22)	-18.37139
coati	0.15 (0.01)	0.32 (0.03)	0.12 (0.02)	5.08 (0.69)	-25.01367	0.32 (0.02)	5.08 (0.69)	-25.01367
rat	0.11 (0.01)	0.22 (0.02)	0.09 (0.02)	7.38 (0.97)	-74.13077	0.22 (0.02)	7.38 (0.97)	-74.13077
peccary	0.15 (0.01)	0.30 (0.02)	0.12 (0.02)	5.60 (0.52)	-74.0031	0.30 (0.01)	5.60 (0.52)	-74.0031
brocket	0.15 (0.01)	0.27 (0.02)	0.13 (0.02)	7.13 (0.79)	-74.16115	0.27 (0.01)	7.13 (0.79)	-74.16115
paca	0.17 (0.01)	0.26 (0.02)	0.15 (0.01)	9.18 (0.97)	-99.60046	0.26 (0.01)	9.18 (0.97)	-99.90046
agouti	0.14 (0.004)	0.25 (0.01)	0.11 (0.01)	6.84 (0.33)	-429.0687	0.25 (0.01)	6.84 (0.33)	-429.0687
squirrel	0.12 (0.02)	0.25 (0.03)	0.08 (0.03)	5.93 (1.12)	-28.7527	0.25 (0.03)	5.93 (1.12)	-28.7527
mouse	0.09 (0.01)	0.17 (0.03)	0.08 (0.02)	11.22 (2.56)	-38.21456	0.17 (0.02)	11.22 (2.56)	-38.21456

Table D.5.2: Estimates (in ms^{-1}) from a size biased lognormal distibution compared with estimates from a lognormal distribution (standard errors in parentheses). The negative loglikelihood ($-\ell$) is also given.

Species	**Nonparametric Method**		**Parametric Method**					
	Harmonic mean	Standard mean	size biased lognormal			lognormal		
	$\hat{\mu}_x$	$\hat{\mu}_x$	$\hat{\mu}_x$	$\hat{\nu}$	$(-\ell)$	$\hat{\mu}_x$	$\hat{\nu}$	$(-\ell)$
ocelot	0.27 (0.03)	0.40 (0.04)	0.28 (0.02)	0.61 (0.04)	-14.86784	0.41 (0.03)	0.61 (0.04)	-14.86794
coati	0.15 (0.01)	0.32 (0.03)	0.16 (0.01)	0.86 (0.05)	-25.41895	0.33 (0.03)	0.87 (0.05)	-25.41895
rat	0.11 (0.01)	0.22 (0.02)	0.11 (0.01)	0.82 (0.05)	-80.41684	0.22 (0.02)	0.82 (0.05)	-80.41684
peccary	0.15 (0.01)	0.30 (0.02)	0.15 (0.01)	0.82 (0.04)	-85.52358	0.30 (0.02)	0.82 (0.04)	-85.52358
brocket	0.15 (0.01)	0.27 (0.02)	0.15 (0.01)	0.76 (0.04)	-78.14788	0.28 (0.02)	0.76 (0.04)	-78.14788
paca	0.17 (0.01)	0.26 (0.02)	0.17 (0.01)	0.65 (0.03)	-111.6609	0.26 (0.01)	0.65 (0.03)	-111.6609
agouti	0.14 (0.004)	0.25 (0.01)	0.13 (0.004)	0.81 (0.02)	-461.8157	0.26 (0.01)	0.81 (0.02)	-461.8157
squirrel	0.12 (0.02)	0.25 (0.03)	0.12 (0.01)	0.86 (0.07)	-33.22363	0.25 (0.03)	0.86 (0.07)	-33.22363
mouse	0.09 (0.01)	0.17 (0.03)	0.09 (0.01)	0.75 (0.08)	-40.68071	0.17 (0.02)	0.75 (0.08)	-40.68071

Appendix E

In this section of the appendix, we give additional results from Chapter 7, in the main thesis. One important assumption of REM is that the encounter data and animal speed data are independent. We investigate, using simulations, the effect of violating the independence assumption on the density estimator. This section is linked to Section 7.2.1 in Chapter 7. In this section we give some results of testing the independence assumption using the simulation algorithm in Section 7.2, in the main thesis. This is done by simulating dependent data and independent data and fitting size biased models to these data. Table E.0.3 gives the results for small value of the density ($D = 10$) and expected animal speed $\mu = 0.150$ (ms^{-1}) with low variability, $\text{Var}(x_l) = 0.05$, or high variability, $\text{Var}(x_l) = 0.31$. The results show that violating the independence assumption has minor consequences on the parameter estimates and precision. Note that correcting bias in speed of faster moving animals accounts for the dependence of these animals being more likely to encounter the trap. As expected, REM-SB and iREM-Sb gave similar estimates of the density but REM-Sb gave smaller estimates of the standard error, particularly when the variability in the speed is high.

Table E.0.3: Simulation results from using all the speed data with a large density D; small and large number of camera traps c; and a fixed value of expected speed $\mu_x = 0.150$ (ms^{-1}) with low and high variability in animal speed. The average parameter estimates (standard errors are in parentheses), the Standard deviation (Sd) and Root Mean Square Error (RMSE) are given.

	Dependent data			**Independent data**			
	REM-SB	iREM-SB		REM-SB	iREM-SB		Camera trap days
	$\hat{D}$	$\hat{D}$	$\hat{\mu}_x$	$\hat{D}$	$\hat{D}$	$\hat{\mu}_x$	$\bar{n}$
			Low variability in animal speed (= 0.05)				
$D = 10;\ c = 10; \mu_x = 0.156; n = 75.12$							
estimate	10.05 (3.49)	10.05 (3.80)	0.152 (0.01)	10.10 (3.51)	10.13 (3.79)	0.152 (0.01)	75.93
Sd	3.87	3.87	0.01	3.84	3.87	0.001	1.89
RMSE	3.49	3.80	0.01	3.51	3.80	0.01	2.05
$D = 10;\ c = 100; \mu_x = 0.156; n = 759.62$							
estimate	10.14 (1.12)	10.14 (1.16)	0.152 (0.003)	10.18 (1.12)	10.18 (1.16)	0.152 (0.003)	758.74
Sd	1.95	1.95	0.01	1.99	1.96	0.001	1.65
RMSE	1.13	1.16	0.01	1.14	1.17	0.004	1.87
			High variability in animal speed (= 0.31)				
$D = 10;\ c = 10; \mu_x = 0.156; n = 76.78$							
estimate	11.94 (4.07)	11.82 (5.10)	0.150 (0.02)	11.83 (3.98)	11.69 (4.98)	0.154 (0.03)	75.84
Sd	4.11	5.99	0.03	4.78	4.39	0.04	1.73
RMSE	4.51	5.42	0.03	4.38	5.26	0.03	1.98
$D = 10;\ c = 100; \mu_x = 0.156; n = 759.59$							
estimate	10.77 (1.23)	10.76 (1.48)	0.151 (0.01)	10.61 (1.21)	10.72 (1.50)	0.152 (0.01)	758.98
Sd	1.17	1.19	0.02	1.29	1.47	0.01	1.64
RMSE	1.45	1.66	0.03	1.35	1.66	0.02	1.75