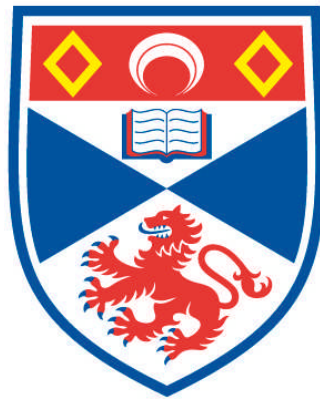# MIXED EFFECT MODELS IN DISTANCE SAMPLING

Cornelia Sabrina Oedekoven

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

2013

MIXED EFFECT MODELS IN DISTANCE SAMPLING

Cornelia Sabrina Oedekoven

Thesis submitted for the degree of
DOCTOR OF PHILOSOPHY
in the Schools of Mathematics and Statistics
UNIVERSITY OF ST ANDREWS
ST ANDREWS
JANUARY 2013

Declarations

I, Cornelia Oedekoven, hereby certify that this thesis, which is approximately 38000 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.
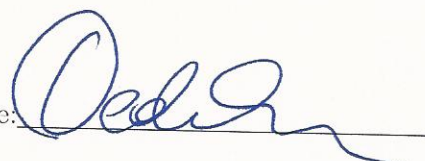
Date: 3/5/2013   Signature of Candidate:

I was admitted as a research student in December 2008 and as a candidate for the degree of Doctor of Philosophy in Statistics in December 2008; the higher study for which this is a record was carried out in the University of St Andrews between 2008 and 2012.

Date: 3/5/2013   Signature of Candidate:

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

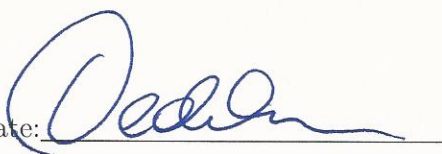Date: 3/5/2013   Signature of Supervisor:

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use unless exempt by award of an embargo as requested below, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. I have obtained any third-party copyright permissions that may be required in order to allow such access and migration, or have requested the appropriate embargo below. The following is an agreed request by candidate and supervisor regarding the electronic publication of this thesis:

Embargo on both printed copy and electronic copy for the same fixed period of one year on the following ground: publication would preclude future publication.

Date: 3/5/2013

Signature of Candidate:

Signature of Supervisor:

iii

Dedication

Diese Arbeit ist meinen Eltern Herta und Peter Oedekoven gewidmet, von denen eine(r) das Talent hatte, Klassenbeste(r) im Statistikkurs an der Universität zu Köln zu sein, und der/die andere die Ausdauer, sich durch den selben Statistikkurs zu kämpfen, um weiterhin gemeinsam Kurse besuchen zu können.

This thesis is dedicated to my parents Herta and Peter Oedekoven, of whom one had the skills to be the best student in their statistics class at University of Cologne and one had the perseverance to struggle through the same statistics class so that they could continue their classes together.

Herta und Peter Oedekoven

Abstract

Recently, much effort has been expended for improving conventional distance sampling methods, e.g. by replacing the design-based approach with a model-based approach where observed counts are related to environmental covariates (Hedley and Buckland, 2004) or by incorporating covariates in the detection function model (Marques and Buckland, 2003).

While these models have generally been limited to include fixed effects, we propose four different methods for analysing distance sampling data using mixed effects models. These include an extension of the two-stage approach (Buckland et al., 2009), where we include site random effects in the second-stage count model to account for correlated counts at the same sites. We also present two integrated approaches which include site random effects in the count model. These approaches combine the analysis stages for the detection and count models and allow simultaneous estimation of all parameters. Furthermore, we develop a detection function model that incorporates random effects.

We also propose a novel Bayesian approach to analysing distance sampling data which

uses a Metropolis-Hastings algorithm for updating model parameters and a reversible jump Markov chain Monte Carlo (RJMCMC) algorithm for assessing model uncertainty. Lastly, we propose using hierarchical centering as a novel technique for improving model mixing and hence facilitating an RJMCMC algorithm for mixed models. We analyse two case studies, both large-scale point transect surveys, where the interest lies in establishing the effects of conservation buffers on agricultural fields. For each case study, we compare the results from one integrated approach to those from the extended two-stage approach. We find that these may differ in parameter estimates for covariates that were both in the detection and the count model and in model probabilities when model uncertainty was included in inference. The performance of the random effects based detection function is assessed via simulation and when heterogeneity in the data is present, one of the new estimators yields improved results compared to conventional distance sampling estimators.

Acknowledgements

Lindesay Scott-Hayward, Angelika Studeny, Laura Marshall and Joanne Potts for their camaraderie and friendship.

Data acknowledgements

Institutional funding acknowledgements

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1   Conventional distance sampling

Distance sampling is a tool for assessing wildlife abundance that is commonly used when the interest lies in evaluating how many individuals (or clusters of individuals) of the species of interest occur in a defined study area (e.g. Buckland et al., 2000; Cañadas and Hammond, 2006; Marques et al., 2007). Although the methods are also applicable to plants (Buckland et al., 2007), we will generally speak of animals in the following.

Distance sampling comprises a suite of methods, e.g. line transect sampling (e.g. Burnham et al., 1980), point transect sampling (e.g. Buckland, 2006), cue counting (e.g. Borchers et al., 2009) or trapping point transects (e.g. Buckland et al., 2006; Potts et al., 2012). We focus on line or point transects in the following. Traditionally, each of these methods requires that samplers such as lines or points are placed within the study area according to some sampling design and that an observer makes detections of the species of interest along or at these samplers.

These methods share an underlying concept which recognises that some of the animals within the search area are not detected and that the proportion of those that were missed can be estimated by collecting additional information. This additional information usually consists of the distance to the detection, i.e. perpendicular distance from the line for line transects or radial distance from the point for point transects (Buckland et al., 2001).

These distances are used to estimate a detection function which models the decay in detection probabilities with increasing distance from the sampler. This detection function may then be used to estimate the average detection probability within the search area, which is used to scale up the number of observed detections to an estimate of the number of individuals in the search area (or number of clusters in the case that detections are made of clusters of individuals). The latter is converted into an estimate of abundance in the study area using a design-based approach where the number of individuals in the search area is divided by the proportion of the study area that was searched. Estimators of abundance or density of the study species are summarised in Buckland et al. (2001). These conventional distance sampling (CDS) methods rely on several assumptions (Buckland et al., 2001; Burnham et al., 1980):

I. All animals on the line or point are detected with certainty. If some animals on the line are missed, due to perception or availability bias, resulting abundance estimates are likely to be negatively biased.

II. Samplers are located according to a survey with an element of randomisation. Meeting this assumption has two main consequences.

A: It insures that the area that was searched is a good representation of the whole study area, allowing use of a design-based approach for scaling up from

an encounter rate estimate in the search area to an encounter rate estimate in the study area.

B: Placement of samplers in the study area is independent from the distribution of animals. Estimators for the average detection probability in the search area incorporate a function that describes the expected distribution of animals in the search area with respect to increasing distance from the line or point. Using CDS methods, we assume that this distribution is on average uniform for lines and linearly increasing for points, which requires that this assumption is met.

III. Distances are measured without error. Measurement errors may lead to biased abundance estimates by over- or underestimating the average detection probability.

IV. Observation process is a snapshot and animals are detected at their initial location. For line and point transects, bias in the estimate of average detection probability may arise due to animal movement whether the movement is random or responsive to the observer.

V. Detections are independent.

In addition to these assumptions, reliable estimation of abundance using distance sampling methods requires that:

i. The detection function is sufficiently flexible to capture the decay in detection probability well and allow unbiased estimation of the average detection probability - often referred to as the pooling robustness property. For CDS methods, a flexible fit of the detection function is obtained by choosing a best fitting model

from a suite of contending models including different key functions, possibly in combination with adjustment terms (Buckland et al., 2001).

ii. The detection function has a 'shoulder', i.e. that animals out to some distance from the line are detected with certainty - often referred to as the shape criterion.

iii. The number of samplers is large and that the samplers constitute a good representation of the study area. This allows reliable scaling up from number of animals in the search area to abundance in the study area. In addition, this ensures that the distribution of animals with respect to the samplers is on average as described under assumption II.

iv. Samplers are independent from each other.

1.2 Recent developments in distance sampling

Over the past decade or so, a lot of effort has been invested into developing distance sampling methods that allow one or more of these assumptions to be relaxed. With reference to the above lists, these include:

I. *Mark-recapture distance sampling*

Borchers et al. (1998) developed mark-recapture distance sampling methods (MRDS) for line transect surveys where detection on the line is not certain. These authors combined mark-recapture and distance sampling methods where two independent observers simultaneously conduct the transect survey and set up mark-recapture trials for each other. This allows the number of animals missed on the line to be estimated. Studies following this work incorporated

heterogeneity in the detection models and explored different levels of independence between the two observers (e.g. Borchers et al., 2006; Buckland et al., 2010). Laake et al. (2011) developed MRDS methods for point transects.

II. *Modelling non-independent distribution of animals with respect to samplers*

Marques et al. (2010) and Marques et al., in press, developed estimators that replace the assumed distribution of animals with respect to increasing distance from the samplers for CDS methods with a model of the estimated distribution of animals with respect to the linear feature from which the survey was conducted.

III. *Models for measurement errors in distance sampling*

Borchers et al. (2010) developed estimators for a detection function for distance data with systematic and with stochastic measurement errors. Marques (2004) proposed estimators for density in the case of multiplicative errors in distance measurements.

IV. *Dealing with animal movement*

Fewster et al. (2008) applied MRDS methods to double observer line transect data to show that animal movement may constitute a problem for species of high mobility. DiTraglia (2007) proposed adjusted line transect estimators which incorporate movement models. Buckland (2006) showed that for some songbirds, point transects using the snapshot method may produce results with less bias. Spear et al. (1992) and Spear and Ainley (1997a,b) proposed methods for correcting abundance estimates for directionally flying seabirds obtained from strip transects by taking into account the birds' flight speed and direction in relation to the survey ship.

V. *Covariate models for observed counts at the samplers*

Hedley and Buckland (2004) replaced the design-based approach from CDS with a model-based approach using spatial models that relate animal density to spatial and/or habitat covariates. These models may then be used to make predictions on animal densities throughout the study area, including those parts that were not surveyed. These methods do not require that the survey followed a random design.

The two-stage approach (Buckland et al., 2009) may be used for those studies where the interest lies in the relationship between animal densities and the covariates, e.g. for designed experiments where a treatment was applied to part of the study area.

i. *Increasing the flexibility of detection functions*

Marques and Buckland (2003, 2004) increased the flexibility of detection functions by modelling heterogeneity in detection probabilities between detected objects. Their approach incorporates covariates affecting detection probabilities in the scale parameter of the half-normal or hazard-rate detection function.

Miller and Thomas, unpublished manuscript, proposed using mixture models. These are composed of two or more detection functions which are scaled using a corresponding mixing proportion.

1.3 Developments of distance sampling methods proposed in this thesis

While this list of developments in distance sampling methodology is far from exhaustive, it demonstrates the need to supplement or replace some of the methods within CDS. However, in most cases (with the exception of Potts (2011) and Yuan et al., unpublished manuscript) these approaches do not make use of random effects in their models. The main objective of this thesis is to develop estimators, likelihood formulations and algorithms for incorporating random effects, for which we generally assume normality with a zero-mean and unknown standard deviation, into models applied to distance sampling data. In particular, we address two main areas of incorporating random effects: the covariate model for counts or densities on the plot (chapters 2 to 5) and the detection function model (chapter 6).

In chapter 2 we begin by describing an extended version of the two-stage approach (Buckland et al., 2009) where we include random effects in the count model to accommodate correlated measurements, e.g. due to closeness of samplers in space or repeat sampling at the same line or point. Hence, we present methods that do not rely on assumption V. (the assumption of random placement of samplers) or on item iv. (independence of samplers) from the above lists. In addition, we incorporate models for heterogeneity in detection probabilities using MCDS, addressing item i. (flexible detection functions). Each of these items is further addressed in chapters 3 and 4. However, like the original approach described by Buckland et al. (2009), the extended two-stage approach from chapter 2 has the disadvantage that the second-stage density model conditions on the first-stage detection model and hence, uncertainty from the

latter does not propagate into the density model. We address this issue in chapters 3 and 4 by proposing integrated likelihoods that combine the likelihood components of the first and second stage into one.

For the integrated likelihood approach presented in chapter 3, counts at the sampler are divided into counts n_i by distance interval $i = 1, ..., I$. The Poisson model for n_i comprises two components: a mixed effect log-linear Poisson model for the expected number of animals within the search area of the sampler N which is adjusted for imperfect detection using the estimated proportion of N that was detected within the ith interval. The latter is estimated using the unconditional likelihood of observed distances (Royle et al., 2004). This approach is applicable to interval distance data or exact distance data. For the latter, however, the exact distance measurements need to be converted into interval data.

In chapter 4, we propose integrated likelihood formulations that are applicable to both exact and interval distance data. Here, counts are adjusted for imperfect detection within the search area by incorporating the effective area into the mixed effect log-linear Poisson model as an offset. This approach uses the conditional probability density function of observed distances (Buckland et al., 2001).

Recognising that these integrated likelihoods may be difficult to maximise in some cases, we present a novel Bayesian approach to distance sampling in chapter 4 which uses the integrated likelihood formulations presented in the same chapter. This approach uses a random walk single-update Metropolis-Hasting algorithm (Hastings, 1970; Metropolis et al., 1953) to update model parameters. Model uncertainty may be assessed using an RJMCMC algorithm (Green, 1995).

In chapter 5, we present a novel technique for dealing with some of the model mixing

difficulties one may encounter using hierarchical models for an RJMCMC algorithm. The difficulties we refer to may arise when the random effects coefficients absorb the effect of one or more of the fixed effect covariates and prevent the acceptance of these covariates into the model as the effects are already accounted for. We use hierarchical centering to reparameterise the model: the generally assumed zero-mean of the random effect is replaced with a model incorporating the intercept and one or more covariates from the Poisson model. Now, the random effects coefficients are supposed to absorb the effects of the covariates included in the centering, given that they have an effect, and models with these covariates are favoured over those without.

In chapter 6, we address item i. from the above list (flexibility in detection functions) and present a new detection function model that models heterogeneity in detection probabilities between different detections by including random effects in the scale parameter of the half-normal key function. Two estimators for abundance and associated variance are described and assessed via simulation in comparison to CDS methods.

For each of the chapters, we analyse case studies or simulated data and contrast results from competing methods. In chapter 7, we conclude with a general discussion on these topics.

Chapter 2

Fitting random effects models to distance sampling data using a two-stage approach

2.1 Introduction

Traditionally, inference on abundance from distance sampling data relies on a model-based component (the estimation of the detection function to account for imperfect detection) and a design-based component (estimation of the encounter rate in the study area based on encounter rate estimates along the transect lines or points, Buckland et al., 2001). The design-based component assumes that transect lines or points are randomly distributed within the study area. There is currently much interest in replacing the design-based component by a modelling approach, for which random line location is not assumed, and which allows animal density to be related to spatial covariates such as habitat (Burt et al., 2003; Buckland et al., 2004; Hedley and

Buckland, 2004; Royle et al., 2004; Kéry et al., 2005). Commonly, the abundance is modelled as a function of covariates using a generalized linear model (GLM) or generalized additive model (GAM) but may also be modelled as a spatial point process (Johnson et al., 2010).

Increasingly, large-scale experimental studies are needed to assess the effects of some intervention on numbers of species of conservation interest. The intervention might be a change in agricultural or forestry practice that may have unintended consequences on population abundance, or it might be the introduction of a management practice that is intended to increase population abundance. Buckland et al. (2009) describe a two-stage model-based approach for analysing distance sampling count data from such studies. In the first stage, a detection function model is fitted to the distance data, from which an offset is estimated to account for imperfect detection within the surveyed strip or circle. In the second stage, this offset is incorporated in a count model using a log-link and a Poisson error structure in a GLM. The problem arising then is that an assumption has to be made that the estimate of the detection function in the first stage represents the true detection function. However, non-parametric bootstrapping may be used to quantify precision of parameter estimates, allowing uncertainty from fitting the detection function to propagate into the second stage.

Buckland et al. (2009) recommended that when the study consists of a large number of sites, these should be included as a random effect. This has the advantage that inference is not limited to those sites included in the study (McCulloch and Searle, 2001). Only one site parameter is then required (as opposed to the $j - 1$ parameters for j sites if treated as fixed), and the approach accommodates positive correlation in counts from a single site.

We adopt the suggestion of Buckland et al. (2009), and include site random effects into the two-stage approach using a generalized linear mixed model (GLMM) for the counts. In contrast with a GLM, the likelihood of a GLMM includes a random effect component (McCulloch and Searle, 2001). Although other distributions have been suggested for random effects (e.g. Komárek and Lesaffre, 2008), most commonly a normal distribution is assumed. In this chapter, we present an extended version of the two-stage approach of Buckland et al. (2001) which includes random effects in the second stage count model for which we assume normality with a zero-mean and unknown standard deviation σ_b. Our approach is presented for line and point transect data and applicable to either exact or interval distance data. In the following section, we present the likelihoods for comparison to the following chapters where these formulations are modified. We then analyse two case studies, point transects of indigo buntings (*Passerina cyanea* L.) and point transects of northern bobwhite (*Colinus virginianus* L.) coveys. Results from these analyses are presented here and are compared to results from analyses in chapters 3 and 4 in the results sections of those chapters.

2.2 The two-stage approach

Consider a wildlife study carried out at a number of sites, at each of which point or line transects are placed according to some design. Each site is surveyed at least once following a distance sampling protocol (Buckland et al., 2001). For line transects the observer travels down the line and records the perpendicular distances to the line for each detection of the species of interest. For point transects, the observer remains at

the point for a fixed amount of time and records the distances from the point to the detections. Distances can be recorded either exactly or in intervals. We assume that animals on the line or point are certain to be detected.

If all animals within the search radius were detectable with certainty, then counts at the line or point could be modelled via a log-link using a GLMM with a Poisson or negative binomial error structure. Including site as a random effect allows counts from the same site to covary. For the two-stage approach, we consider the total count n_{jpr} at visit r to line or point p of site j to be a Poisson random variable with $E(n_{jpr}) = \lambda_{jpr}$ which can be modelled by a linear predictor via a log-link function using a GLMM:

$$\lambda_{jpr} = \exp\left(\beta_0 + b_j + \sum_{k=1}^{K} x_{kjpr}\beta_k\right). \tag{2.1}$$

Here β_0 represents the fixed effect intercept, b_j the random effect coefficient for site j with $b_j \sim N(0, \sigma_b^2)$, x_{kjpr} the value of the kth fixed effect covariate measured during the respective visits to that line (point), and β_k the associated coefficients.

In this formulation (eqn (2.1)) we assume perfect detection on the plot. As this is generally not the case, we need a formulation to allow for detectability decreasing with distance from the line or point. Hence, in the first stage, a probability density function $f(y)$ is fitted to the observed detection distances where y represents the distances from the line or point to the observed detections given that the animal is in the strip of half-width w centered on the line (lines) or in the circle of radius w around the point (points) (Buckland et al., 2001). It describes the probability that an animal was in interval $(y, y + dy)$ given that it was detected within distance w of the line or point, where $\int_0^w f(y)dy = 1$ and w is the truncation distance:

$$f(y) = \frac{\pi(y)g(y)}{\int\limits_0^w \pi(y)g(y)dy}. \tag{2.2}$$

The function $\pi(y)$ describes the expected distribution of animals (whether detected or not) with distance from the line or point. When lines or points are randomly positioned, $\pi(y) = 1/w$ for line transects and $\pi(y) = 2y/w^2$ for point transects where w is the truncation distance as before.

The detection function $g(y)$ may be modelled using a key function and adjustment terms (Buckland et al., 2001). However, for simplicity, we omit adjustment terms from the equations presented here and will revert to this topic in chapter 6. Commonly used key functions include the half-normal $g(y) = \exp\left(-y^2/2\sigma^2\right)$ and the hazard-rate $g(y) = 1 - exp\left(-(y/\sigma)^{-\tau}\right)$.

The parameters of the detection function (denoted by $\boldsymbol{\theta}$ in the following) are the scale parameter σ and, additionally for the hazard-rate model, the shape parameter τ. If distances were measured exactly, the parameter estimates are found by maximizing the following likelihood, which is conditional on the number of detections n (Buckland et al., 2004, p. 16):

$$L_y\left(\boldsymbol{\theta}\right) = \prod_{e=1}^{n} f\left(y_e\right) \tag{2.3}$$

where y_e refers to the eth detection.

When detections are made in distance intervals, let f_i be the probability that a detected animal is in interval i. The ith interval is delineated by the cutpoints c_{i-1} and c_i where $c_0 = 0$ unless the data are left-truncated (Buckland et al., 2001), and the outer cutpoint of the outermost interval is $c_I = w$. The f_is can be obtained by

integrating $f(y)$ between the cutpoints of the intervals where:

$$f_i = \frac{\int\limits_{c_{i-1}}^{c_i} f\left(y\right) dy}{\int\limits_{0}^{w} f\left(y\right) dy}. \tag{2.4}$$

Then, parameters within $\boldsymbol{\theta}$ can be estimated by maximising the multinomial likelihood (Buckland et al., 2004, ch. 2):

$$L_{yG}\left(\boldsymbol{\theta}\right) = \left(\frac{n!}{\prod\limits_{i=1}^{I} n_i!}\right) \prod\limits_{i=1}^{I} f_i^{n_i} \tag{2.5}$$

where f_i is the probability that the detected animal falls in interval i, n the total number of detections and n_i the number of detections in the ith interval with I being the outermost interval. Note that in this formulation detections from all sites are assumed to arise from a single detection function. See below for modelling heterogeneity.

$f(y)$ can be used to estimate the effective area ν which is defined as the area beyond which as many animals were seen as were missed within (Buckland et al., 2001). For line transects, the effective strip half-width $\mu = \int_0^w g(x)dx = 1/f(0)$ and the effective area $\nu = 2l_{jpr}\mu$, where l_{jpr} is the length of the line surveyed at the respective visit to the line (this changes to $\nu_{jpr} = 2l_{jpr}\mu$ in case the lengths of the individual lines differ). Similarly for point transects, the effective area at a point is $\nu = 2\pi \int\limits_{0}^{w} yg(y)dy = 2\pi/h(0)$, where $h(0)$ is the slope of $f(y)$ evaluated at distance 0. Consequently, the observed count n_{jpr} divided by an estimate of the effective area at the line or point ν is a valid estimator of density D_{jpr} at the line or point. Hence, given a log-link function, the log effective area gives the appropriate offset to add to

the count model from eqn (2.1), giving:

$$\lambda_{jpr} = E\left(n_{jpr}\right) = \exp\left(\beta_0 + b_j + \sum_{k=1}^{K} x_{kjpr}\beta_k + \ln\left(\nu\right)\right). \tag{2.6}$$

Note that the offset is an estimate, whereas offsets are treated as known constants. Hence, uncertainty about estimating the detection function parameters does not naturally propagate into the count model. We address this issue in section 2.2.3.

Using the formulation for the expected counts including the offset estimated from the first-stage detection model from eqn (2.6), the likelihood for the second-stage count model is given by (modified from McCulloch and Searle, 2001):

$$L_{y,n}(\boldsymbol{\beta}, \sigma_b|\boldsymbol{\theta}) = \prod_{j=1}^{J} \int_{-\infty}^{\infty} \left[\prod_{p=1}^{P_j} \prod_{r=1}^{R_j} \frac{(\lambda_{jpr})^{n_{jpr}} \exp\left(-\lambda_{jpr}\right)}{n_{jpr}!}\right] \frac{1}{\sqrt{2\pi\sigma_b^2}} \exp\left(-\frac{b_j^2}{2\sigma_b^2}\right) db_j \tag{2.7}$$

which is conditional on the parameter estimates for $\boldsymbol{\theta}$ from the first stage. Parameter vector $\boldsymbol{\beta}$ combines the coefficients for the fixed effect covariates and intercept from eqn (2.6). J equals the total number of sites and P_j and R_j refer to the total number of lines (or points) at the jth site and total number of visits to the jth site, respectively. P_j and R_j may vary between different sites. The integral in eqn (2.7) denotes that we integrate out the random effects for which normality is assumed. Inside the integral we have two main components: the product of the Poisson likelihoods for all observed counts at the jth site (inside the square brackets) and the normal density for the random effects coefficient b_j. The random effects are integrated out by integrating the product of these components over all possible values for b_j, i.e. from negative infinity to positive infinity.

However, mixed-effect Poisson models of this form including an offset can be fitted using the *glmer* function of the *lme4* package (Bates, 2009b) in R. This function uses the adaptive Gauss-Hermite approximation to evaluate the integral in calculating the marginalized log-likelihood (Bates, 2009a). The number of quadrature points can be manually chosen with the argument $nAGQ$. If the default is used, where $nAGQ$ equals one, the approximation corresponds to Laplace (e.g. MacKay, 2003, ch. 27). Lesaffre and Spiessens (2001), however, recommend using 10 quadrature points. Larger values may increase the accuracy in the evaluation at the cost of computing time (Rabe-Hesketh et al., 2002). To determine how many quadrature points to choose, a model can be fitted with varying values for $nAGQ$ while using the same model of covariates. For a range of values, the approximated marginalized likelihood may stabilize. Out of this range, it is recommended to choose a small value for $nAGQ$ and use the same value for all models.

2.2.1   Heterogeneity in Detection Probabilities

When there is no heterogeneity in the detection probabilities, it is sufficient to include detections from all sites in one detection function and estimate one common effective area. However, detection probabilities may vary between different lines or points or even between different detections. There are two main strategies within distance sampling to account for heterogeneity in detection probabilities (Buckland et al., 2001, ch. 3.7). One strategy is post-stratification where the observed distances are divided into different strata based on one of the available covariates. A best fitting detection function is found independently for each stratum and stratum-specific estimates of

the effective area included in the count model.

A generally more parsimonious approach is multiple covariate distance sampling (MCDS) (Marques and Buckland, 2003, 2004; Marques et al., 2007). Here, the scale parameter is modelled as a function of covariates and the conditional density of the observed distances given the associated covariates $\boldsymbol{z}$ becomes $f(y|\boldsymbol{z})$. This allows us to model detection probability not only as a function of increasing distance from the point or line but also with respect to covariates affecting detection conditions and detectabilities of animals. We thus have (Buckland et al., 2004, p. 33):

$$f(y|\boldsymbol{z}) = \frac{\pi(y)g(y, \boldsymbol{z})}{\int_0^w \pi(y)g(y, \boldsymbol{z})dy}. \tag{2.8}$$

The conditional likelihood is thus

$$L_y(\boldsymbol{\theta}) = \prod_{e=1}^{n} f(y_e|\boldsymbol{z}_e). \tag{2.9}$$

Using the same key functions as above, the scale parameter of the detection function is now modelled as the exponential of a linear function of these covariates:

$$\sigma(\boldsymbol{z}) = \delta_0 \times \exp\left(\sum_{q=1}^{Q} z_q \delta_q\right), \tag{2.10}$$

where δ_0 and the δ_q represent the intercept and the coefficients for the Q covariates. In turn, the effective area can now be expressed for each visit r to line (point) p of site j using covariates $\boldsymbol{z}$: for line transects $\nu_{jpr} = 2l_{jpr}/F_{n_{jpr}}$, where $F_{n_{jpr}} = \left[\sum_{e=1}^{n_{jpr}} f(0|\boldsymbol{z}_e)\right]/n_{jpr}$. For point transects, $\nu_{jpr} = 2\pi/H_{n_{jpr}}$, where $H_{n_{jpr}} = \left[\sum_{e=1}^{n_{jpr}} h(0|\boldsymbol{z}_e)\right]/n_{jpr}$.

AIC can be used to compare models from the different strategies. When using stratification, the sum of AIC values from all different strata can be compared to the AIC value from the MCDS model as long as both analyses are based on exactly the same data.

2.2.2 Model Selection

For the first-stage detection function, a best fitting model may be found by comparing AIC values using Distance software (e.g. Distance 6, Thomas et al., 2010; Newson et al., 2010). However, an automatic model selection for the detection function based on AIC values can be set up in R, e.g. by using calls to the MCDS engine of the Distance software or by using functions from the *mrds* package. Similarly for the second-stage count model, a best fitting model may be found using AIC values.

2.2.3 Estimating the Precision

The precision of parameter estimates can be estimated using a non-parametric bootstrap routine (Buckland et al., 2009). For each bootstrap iteration sites are resampled with replacement until the original number of sites is obtained. Each time a site is picked, all visits to that site are included to avoid the assumption of independence between visits to the same site. Models for the detection function and for the counts are fitted to the bootstrapped data. Here two main strategies can be followed. To obtain precision estimates conditional on the best fitting models (for detection and counts) for the original data, the same models selected for the original data are refitted to the bootstrapped data for all iterations. Precision of the estimates is estimated using

bootstrap standard errors and 95% percentile confidence intervals. To incorporate model selection uncertainty into inference, the best fitting models for both stages are found independently within each bootstrap iteration (Buckland et al., 1997). This can be done by fitting the same set of models that were fitted to the original data to the bootstrapped data and applying the same model selection routine during each iteration. Model probabilities are given by the proportion of times the respective models were selected.

2.3 Case study 1: point transect surveys of indigo buntings

2.3.1 The data

The National CP-33 Monitoring Program coordinated by the Mississippi State University, Department of Wildlife, Fisheries, and Aquaculture was set up to monitor beneficial effects of herbaceous buffers around agricultural fields on bird densities in several Southeastern and Midwestern states (Evans et al., 2013). To set up a monitoring scheme, a minimum of 40 CP-33 contracts per state were randomly selected from all CP-33 contracts. Buffered treatment fields within these contracts were selected for monitoring of several priority species. Here, we analyse indigo bunting data.

During the breeding seasons of 2006-2007, point transect surveys were conducted from one point per field located in the buffer at the edge of the field. Unbuffered control points on the edge of fields of the same agricultural use, located 1-3km away, were surveyed concurrently to ensure similar conditions for observing and calling rates of birds. Each pair of adjacent treatment and control points was considered a site, and

each site was surveyed between 1-4 times per survey year. The objective was to evaluate whether buffers result in higher bird densities.

Observers recorded all male indigo buntings (all singles) detected visually or aurally in a 10-minute period in predetermined intervals (0-25, 25-50, 50-100, 100-250, 250-500, >500m). Information on wind speed (in km/hr), fog (scaled 0-2 with increasing amounts of fog) and cloud cover (as a percentage) were collected simultaneously. We assume that indigo buntings distribute themselves independently of point locations. Only those sites surveyed at least once in each of the two survey years were included. An additional criterion was that each state included in the analysis contained >50 detections. The 446 sites satisfying these criteria were located in nine states (Georgia, Illinois, Indiana, Kentucky, Missouri, Mississippi, Ohio, South Carolina and Tennessee).

2.3.2   Analysis using the two-stage approach

The first stage involved fitting a collection of detection function models to the distance data and selecting the best by minimum AIC. As distances were collected in intervals, we used the multinomial likelihood given in eqn (2.5) to find parameter estimates for the respective models. Upon visual inspection of the detection functions fitted by Distance (Thomas et al., 2010), the data were truncated at 100m as detection probabilities were generally below the recommended 0.1 beyond this distance regardless of the choice of model.

Hence, with only three distance intervals left, and allowing a degree of freedom for

assessing model fit, we only considered one parameter models for the detection function. These included the half-normal or hazard-rate models where, for the latter, the shape parameter was fixed (see below). For the same reason, modelling heterogeneity in detection probabilities was limited to post-stratification with *year*, *type* (i.e. control or treatment point) and *state* as potential covariates. For the stratified models, the multinomial likelihood for interval distance data from eqn (2.5) changed to:

$$L_{yG_{strat}}(\boldsymbol{\theta}) = \prod_{s=1}^{S} \frac{n_s!}{\prod\limits_{i=1}^{I} n_{s_i}!} \prod_{i=1}^{I} f_{s_i}^{n_{s_i}} \tag{2.11}$$

where n_s is the total number of detections in stratum s and n_{s_i} the number of detections in the ith interval of stratum s. The f_{s_i} represent the cell probabilities for interval i in stratum s and were obtained by integrating the conditional $f(y)$ for each stratum s between cutpoints of the intervals (see eqn (2.4) on page 15 for details).

To determine an appropriate value for the shape parameter, model fit and AIC values were compared for each stratum using three different values: 2.0, 2.5 and 3.0. Lower and higher values were considered in preliminary analyses. Lower values were excluded as detection functions were spiked near distance zero with a rapid decline in detection probabilities as distances increased. Higher values were excluded, although AIC scores were lower in two cases, as under these models detection probabilities were nearly uniform out to unreasonable distances and standard errors for the scale parameter increased.

Using these three values for the shape parameters, 2.0 was selected for the global and for both strata of each of the detection models stratified by *type* or *year*. For the *state*-stratified model, 2.0 was also selected for six states, while 2.5 and 3.0 were

chosen for two and one state, respectively.

Overall, post-stratification by *state* using a hazard-rate key function returned a lower AIC value than other models. Using the estimates for $\boldsymbol{\theta}$, the effective area was calculated using $\nu_{jpr} = 2\pi \int_0^w yg(y)dy$ for each of the nine states and incorporated into the second-stage count model as an offset.

For the second stage count model, eqn (2.7) was maximised using the *glmer* function in R. Here, R_j ranged from 2 to 8 visits per site. As each site comprised two points, one control and one treatment, p equalled 1 or 2. The argument $nAGQ$ was set to 10 for all models fitted with *glmer*. Potential fixed effects for the count model included the factor covariates *year*, *type* (control or treatment point) and *state* as well as the continuous variable *Julian day*. For the latter, we compared the fit of regression splines using the B-spline basis from the *splines* package in R with the fit of a one parameter linear term. The latter returned better AIC values and was hence used for formulating the contending models. A total of 16 combinations - all possible combinations of the four covariates - were included in the model selection. In these models, the parameter of interest was the covariate *type*. A significant *type* term in the model would indicate a difference in bird densities between the control and treated plots. The random effects term b_j was assumed normal with $b_j \sim N(0, \sigma_{b_j}^2)$.

A non-parametric bootstrap using site as the resampling unit as described in section 2.2.3 was conducted to obtain precision estimates for all parameters. The strategy followed here was to take the best models identified for the real data and fit these to bootstrapped data. Hence, precision estimates are conditional on these models and inference does not include model selection uncertainty.

2.3.3 Results

During the two survey years included in this study, 2006 and 2007, a total of 2924 counts at control or treatment points of 446 sites were made. During these counts, a total of 3785 indigo buntings were detected within the three innermost distance intervals. Parameter estimates as well as analytical (ASEs) and bootstrap standard errors (BSEs) for the best detection and count models are given in Table 2.1. Estimates for the scale parameter of the hazard-rate detection function stratified by *state* ranged between 21.08 (ASE=1.74, BSE=3.51, fixed shape parameter=2.0) for Tennessee and 56.30 (ASE=4.13, BSE=6.85, fixed shape parameter=3.0) for South Carolina.

To calulate a baseline expected number of male indigo buntings within the plot area a using the best model, we set the covariates to $type = $ Control, *Julian day* $= 174$ (the mid-point of all days surveyed) and $state = $ GA. We used the coefficients from Table 2.1 while applying the following transformations (reversing the log-link function of the Poisson model and converting $birds/m^2$ to $birds/a$): $\exp(-10.91 + 0.0046 \times 174 + 0.5 \times 0.49^2) \times a$. The last component inside the bracket represents the contribution of the random effects term and $a = \pi w^2$. The resulting expected baseline of indigo bunting numbers within the plot was 1.43 (BSE=0.59) (or 43.51 (BSE=18.91) birds per km^2).

The remaining fixed effect coefficients represent proportional changes with respect to this baseline. The *type* coefficient for the count model was 0.30 (ASE=0.03, BSE=0.04) indicating a 35% increase of bird densities on treated fields compared to control fields. For both models, BSEs were generally larger than ASEs except for the scale parameters of the detection function for two out of the nine states where they were smaller. These results are further discussed and compared with those using

an integrated likelihood approach in chapter 3.

Table 2.1: Maximum likelihood estimates (MLE), analytic (ASE) and bootstrap (BSE) standard errors for model parameters obtained by the two-stage approach for best models. Shape parameters for the one-parameter hazard-rate detection function were fixed.

Model Paramters	Two-stage			Shape
	MLE	ASE	BSE	
Detection model				
Scale State GA	37.27	7.72	8.12	2.0
Scale State IL	34.42	2.86	3.17	2.5
Scale State IN	24.34	2.35	4.92	2.0
Scale State KY	27.75	1.13	1.52	2.5
Scale State MO	37.78	3.14	2.86	2.0
Scale State MS	38.73	3.33	4.21	2.0
Scale State OH	24.59	1.97	1.94	2.0
Scale State SC	56.30	4.13	6.85	3.0
Scale State TN	21.08	1.74	3.51	2.0
Count model				
Random Effects				
Standard deviation	0.49	-	0.04	
Fixed Effects				
Intercept D_{jpr}	-10.91	0.29	0.43	
Type Treated	0.30	0.03	0.04	
Julian Day	0.0046	0.0017	0.0018	
State IL	1.20	0.18	0.38	
State IN	1.34	0.18	0.49	
State KY	1.79	0.17	0.36	
State MO	0.32	0.16	0.36	
State MS	0.91	0.17	0.37	
State OH	0.92	0.17	0.37	
State SC	0.28	0.18	0.39	
State TN	2.12	0.17	0.44	

2.4 Case study 2: point transect surveys of northern bobwhite coveys

2.4.1 The data

As part of a study to assess the potential benefits of herbaceous buffers around agricultural fields, Mississippi State University, Department of Wildlife, Fisheries, and Aquaculture set up a monitoring program using point transects in a number of Midwestern and Southeastern states in the US (Evans et al., 2013). Similar to case study 1, survey points located at the edge of the field were paired up: one point on a buffered treatment field and the other on a non-buffered control field of the same agricultural use and within $1 - 3$km of the treatment point. Each pair of points will be referred to as a site in the following. Repeat visits were made to each point during fall of three survey years (2006-2008), and each detected northern bobwhite covey was recorded along with their estimated radial distance to the point. To facilitate unbiased distance estimation, observers used satellite images of the point location and surroundings to mark each detected covey. As this survey did not include obtaining estimates of cluster size for each covey, we consider cluster densities (rather than densities of individuals).

Only those states were included in the analysis that contained more than 50 detections of coveys: Georgia, Illinois, Indiana, Iowa, Kentucky, Missouri, Mississippi, North Carolina, South Carolina, Tennessee and Texas. Within these states, 447 sites were visited between 1 and 3 times in each survey year. After defining a truncation distance of 500m following recommendations of Buckland et al. (2001), the analysed data included a total of 2545 detections with associated distances that were observed

during 2534 counts.

2.4.2 Analysis using the two-stage approach

As distance were measured exactly during the surveys, the first step included fitting a detection function to observed distances by maximising the likelihood in eqn (2.3). Preliminary investigation of the distance data indicated that the hazard-rate detection function provided a much better fit than the half-normal. In addition to the global model, we included seven different multiple covariate models where the scale parameter of the hazard-rate detection function was modelled as a function of one, two or three of the covariates, all possible combinations of including the covariates *state*, *year* and/or *type*. For these models, the likelihood changed to eqn (2.9) from page 18.

In a second step, the effective area was incorporated into the density model for λ_{jpr} for which the likelihood is given in eqn (2.7). Parameter estimates were obtained using the *glmer* function of the *lme4* package (Bates, 2009b) in R. The number of quadrature points was set to 10 using the argument $nAGQ$ of this function, following recommendations of Lesaffre and Spiessens (2001). We explored 16 models that included a fixed intercept and a random effect for site and combinations of the four covariates *state*, *type*, *year* and *Julian day*. Best fitting models for both steps were found by minimum AIC values.

As the effective area represents an estimate but is included in the model as if it was a known constant, non-parametric bootstrapping was used to estimate uncertainty (bootstrap standard errors (BSE) and 95% confidence intervals) of parameter

estimates. To implement a non-parametric bootstrap routine with 999 repeats, an automatic model selection was set up in R that included calls to the MCDS engine from the Distance software (Thomas et al., 2010) for the first step. For each bootstrap iteration, sites were resampled with replacement until the original number of sites was obtained (Buckland et al., 2009). To include model uncertainty in inference, the strategy followed here was to select best fitting models based on minimum AIC values for each bootstrap iteration (Buckland et al., 1997).

2.4.3   Results

The preferred detection model from the analysis of the original data was the hazard-rate function that included the covariates *year*, *type* and *state* in the model for the scale parameter. The preferred count model included all available covariates, i.e. *year*, *type*, *Julian day* and *state*. The same models were preferred for the bootstrap analysis although three other models were selected for both the detection and the count model with smaller probabilities (Table 2.2). Maximum likelihood estimates as well as bootstrap standard errors (BSEs) and 95% confidence intervals are given in Table 2.3. The *type* coefficient was 0.63 (ASE=0.12) indicating an 88% increase in covey densities on treated fields compared to control fields. These results will be discussed further and compared to those from an RJMCMC algorithm in chapter 4.

Table 2.2: Models and their probabilities resulting from bootstrap analysis. Each count model included a fixed effect intercept and a random effect for site in addition to shown covariates (JD = Julian day). Model probabilities refer to the percentage of times the respective models were chosen during 999 bootstrap iterations.

Model	Probability
Detection Function	
MCDS: State	0.01
MCDS: Year + State	0.16
MCDS: Type + State	0.02
MCDS: Year + Type + State	0.81
Count	
Type + State	0.003
Year + Type + State	0.01
Type + JD + State	0.10
Year + Type + JD + State	0.89

Table 2.3: Maximum likelihood estimates (MLE), bootstrap standard errors (BSE) and 95% confidence intervals (CI) using the two-stage approach for the models with the highest probabilities (see Table 2.2 for model probabilities). Units of measurements were metres for the detection function model and square metres for the count model.

	Two-stage		
	MLE	BSE	95% CI
Detection function: fixed effects			
Scale Intercept	138.59	16.13	112.26, 163.79
Shape	3.01	0.27	2.68, 3.41
Scale: Year 2006	0.10	0.06	-0.05, 0.14
Scale: Year 2007	-0.15	0.05	-0.25, -0.1
Scale: Type Control	0.15	0.05	0.05, 0.23
Scale: State GA	0.42	0.17	0.05, 0.54
Scale: State IA	0.21	0.16	-0.12, 0.30
Scale: State IL	0.70	0.17	0.35, 0.76
Scale: State IN	0.66	0.14	0.34, 0.72
Scale: State KY	0.64	0.12	0.35, 0.68
Scale: State MO	0.69	0.09	0.46, 0.71
Scale: State MS	0.61	0.10	0.37, 0.64
Scale: State NC	0.66	0.12	0.35, 0.70
Scale: State SC	3E-5	0.14	-0.29, 0.12
Scale: State TN	0.47	0.12	0.19, 0.54
Count model: random effects			
Standard deviation	0.78	0.04	0.69, 0.81
Count model: fixed effects			
Intercept Density	-13.23	0.33	-13.91,-12.87
Year 2007	0.17	0.13	-0.16, 0.37
Year 2008	0.17	0.11	-0.12, 0.31
Type Treatment	0.63	0.12	0.36, 0.71
Julian Day	-0.01	3E-3	-0.02, -0.01
State IA	-0.74	0.44	-1.65, -0.24
State IL	-0.53	0.38	-1.25, -0.07
State IN	-1.18	0.41	-1.99, -0.70
State KY	-0.44	0.34	-1.07, -0.02
State MO	0.05	0.34	-0.63, 0.46
State MS	-0.37	0.34	-1.04, 0.05
State NC	-1.31	0.36	-1.99, -0.87
State SC	0.08	0.42	-0.76, 0.56
State TN	-1.03	0.38	-1.80, -0.60
State TX	1.46	0.29	0.99, 1.81

Chapter 3

An integrated likelihood approach for modelling distance sampling data with mixed effects

3.1 Introduction

In this chapter, we deal with some of the shortcomings of the two-stage approach presented in chapter 2 and provide an alternative method, the integrated likelihood approach. As for the two-stage approach, the motivation for the integrated likelihood approach also was to replace the design-based component of conventional distance sampling with a model where animal densities or counts are related to covariates such as habitat.

However, the shortcoming of the two-stage approach described by Buckland et al. (2009), and equivalently of the extended version described in chapter 2, is that it treats the offset, and hence the detection function from the first stage, as known.

Hence, nonparametric bootstrapping needs to be used to quantify precision of parameter estimates, to allow uncertainty from fitting the detection function to propagate into the second stage.

Royle et al. (2004), on the other hand, developed an integrated likelihood for point transect data where distances were measured in intervals. These authors combined a covariate model for the latent variable N_p (the true but unknown abundance of animals at the point) with the cell probabilities f_i (derived from the detection function) to model the observed counts n_{pi} in the ith distance interval at the pth point. An advantage of the approach of Royle et al. (2004) is that all model parameters for both the N_p and the f_i are estimated in one step. However, Royle et al. (2004) only assumed a global half-normal detection function where the distance information was pooled across all points for the respective distance intervals.

Here, we extend the approach of Royle et al. (2004) in the following ways. We model heterogeneity in detection probabilities and include model selection for the f_i model as well as for the N_p. We also extend Royle et al.'s Poisson model by including a random effect for site in the abundance model to account for correlated counts at the same sites. This was motivated by our case study 1, the point transects of indigo buntings, where a large number of sites was included in the analyses and the number of repeat visits to the same site varied.

In the following we begin by describing our extended version of the integrated likelihood of Royle et al. (2004) for both line and point transects (section 3.2), analyse our case study of indigo buntings using this integrated likelihood and contrast results with those using the two-stage approach from chapter 2 (section 3.3), and discuss further applications (section 3.4).

3.2 Integrated likelihood

As in the previous chapter, we begin by considering a wildlife study that was carried out at a number of sites, at each of which point or line transects were placed according to some design and that each site was surveyed at least once following a distance sampling protocol (see Buckland et al., 2001 or chapter 2 for further details).

If all animals within the search area were detectable with certainty, then observed counts within the search distance around the line or point would equal the true number of animals within the search distance around the line (point) and could be modelled via a log-link using a generalised linear mixed model (GLMM) with a Poisson error structure ($E(N_{jpr}) = \lambda_{jpr}$) where N_{jpr} is the total number of animals present within the search area a_{jpr} at visit r to line (point) p of site j. The search area a_{jpr} equals $2wl_{jpr}$ for lines (l_{jpr} = length of the respective line) and πw^2 for points; in both cases w represents the truncation distance. Combining adjacent lines or points as sites and including site as a random effect allows (repeat) counts from the same site to covary without causing bias for the remaining parameters in the model. The λ_{jpr} may then be modelled by a linear predictor via a log-link function using:

$$\lambda_{jpr} = \exp\left(\beta_0 + b_j + \sum_{k=1}^{K} x_{kjpr}\beta_k\right). \tag{3.1}$$

Here, β_0 represents the fixed effect intercept, b_j the random effect for site j with $b_j \sim N(0, \sigma_b^2)$, x_{kjpr} the observed covariate values for the $k = 1, 2, ..., K$ fixed effects and β_k the associated coefficients. In the following $\beta_0, ..., \beta_K$ may be summarised as $\boldsymbol{\beta}$. This is similar to chapter 2, however, in this formulation (eqn (3.1)), the combination of fixed and random effects represents a model for the true numbers on the plot

(which is identical to eqn (2.1) from chapter 2.2 page 13 which assumes perfect detection within the search area). In contrast, the combination of fixed and random effects of eqn (2.6) from chapter 2.2 page 16 represents a model for density. As detection is generally not perfect on the plot, we need a formulation to allow for detectability to decrease with distance from the line or point. Here, we employ the unconditional likelihood of observed distances which we denote with $f_u(y)$ (Royle et al., 2004) and contrast with the conditional formulation in the following section.

3.2.1 The unconditional likelihood of observed distances

The unconditional likelihood of observed distances is given by $f_u(y) = \pi(y)g(y)$ (Royle et al., 2004). As for the conditional $f(y)$ from chapter 2, $\pi(y)$ describes the expected distribution of animals (whether detected or not) with distance from the line or point, and $g(y)$ the probability of detecting an animal given that it is at distance y. As before, $\pi(y)$ is assumed to be known ($1/w$ for lines and $2y/w^2$ for points) and a detection function model is proposed for $g(y)$ (Buckland et al., 2001). Detection function parameters pertaining to the unconditional function are summarised as $\boldsymbol{\theta}_u$ in the following.

However, the difference between the conditional and the unconditional formulation for $f(y)$ is that the conditional likelihood conditions on the animal being at distance y and that it is detected while the unconditional only conditions on the animal being at distance y. Hence, the normalising constant in the denominator for the conditional $f(y)$ is $\int_0^w \pi(y)g(y)dy$ while for the unconditional $f_u(y)$ it is $\int_0^w \pi(y)dy$ (which always equals 1 for both lines and points under the assumed distributions for this

function described above in this section). It follows that for the conditional formulation $\int_0^w f(y)dy = 1$ (and this may be called a probability density function), while for the unconditional formulation $\int_0^w f_u(y)dy \neq 1$.

To illustrate the difference, one may think of the likelihood as a joint density of two variables, the animal being at distance y and whether the animal was detected δ given it was at distance y (where δ equals one if detected or zero if not). Hence, $\pi(y)g(y)$ can be rewritten as $\pi(y)p(\delta = 1|y)$ where $p(\delta = 1|y)$ is the probability that the animal is detected given it is at distance y. The conditional likelihood only pertains to those animals that were detected and $\delta = 1$ while the unconditional likelihood pertains to all animals that were available to be detected and $\delta = 1$ or $\delta = 0$.

For the integrated likelihood approach we assume that distances from the line or point are recorded by I distance intervals, or, if distances are recorded exactly, that these are binned into I intervals after the survey is completed. Using the unconditional formulation for interval distance data, the I areas under $f_u(y)$ between the cutpoints of the intervals c_i represent the I proportions of N_{jpr} recorded within the I intervals (as opposed to the proportions of n_{jpr} when using the conditional $f(y)$ from Buckland et al., 2001) and can be obtained using:

$$f_{u_i} = \int_{c_{i-1}}^{c_i} f_u(y)dy = \int_{c_{i-1}}^{c_i} \pi(y)g(y)dy. \tag{3.2}$$

Due to imperfect detection on the plot, we cannot assume that all N_{jpr} were detected. Hence, the sum of the I proportions f_{u_i} does not equal one, but $\sum_{i=1}^{I} f_{u_i} = E[n_{jpr}]/N_{jpr} = P_a$, where P_a is the average detection probability during the respective visit to the plot.

In Figure 3.1 we illustrated an example for the case with $I = 3$ distance intervals.

Here, the fourth cell probability in the graphs with the unconditional function represents the proportion of N_{jpr} that was not detected (i.e. $f_{u_4} = 1 - P_a$). Note that this cell probability cannot be estimated in the same manner as the I cell probabilities, i.e. using eqn (3.2). We address this issue below. The I cell probabilities from the conditional $f(y)$ can be converted into the f_{u_i} using $f_i * P_a = f_{u_i}$.

In the graphs from Figure 3.1, the horizontal line at $1/w$ for lines and the diagonal line defined by $2y/w^2$ for points are $\pi(y)$.

equivalent to what the observer would have seen had he or she not missed any animals. Consequently, using the unconditional $f_u(y)$, the four cell probabilities under the horizontal line at $1/w$ between 0 and w sum up to 1 for lines. Equivalently for points, the cell probabilities under $2y/w^2$ between 0 and w sum up to 1 using the unconditional formulation (note that in the formulations for $\pi(y)$ given above, $\int_0^w \pi(y)dy$ always equals 1 for both lines and points). In contrast, using the conditional $f(y)$, the sum of the cell probabilities under $f(y)$ between 0 and w equals 1 for both lines and points. Thus, for the conditional formulation, each of the I estimated cell probabilities represents an expected proportion of observed counts n_{jpr}, while for the unconditional formulation each of the I estimated cell probabilities represents an expected proportion of the true (but unknown) number of animals in the search area around the line or point N_{jpr}.

The additional cell probability for the unconditional formulation, i.e. the area between the horizontal line at $1/w$ and $f_u(y)$ from 0 to w for lines or between the diagonal line $2y/w^2$ and $f_u(y)$ from 0 to w for points, represents the proportion of animals within the search area that were missed by the observer (f_{u4} in Figure 3.1). For the conditional $f(y)$, this area is not considered for the likelihood formulation
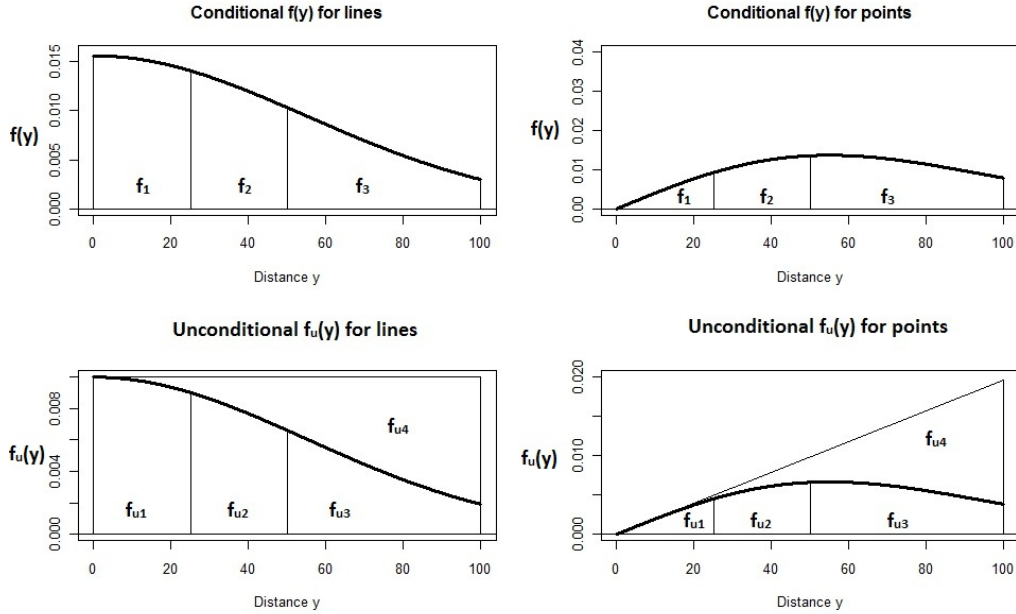
Figure 3.1: Examples for a conditional and unconditional likelihood with a half-normal model using the same scale parameter ($\sigma = 55$) plotted between 0 and $w = 100$ including three distance intervals. The f_1, f_2, f_3 and $f_{u_1}, f_{u_2}, f_{u_3}, f_{u_4}$ refer to the cell probabilities. For the conditional $f(y) : \sum_{i=1}^{3} f_i = 1$, while for the unconditional $f_u(y) : \sum_{i=1}^{4} f_{ui} = 1$.

(see section 2.2 chapter 2 pages 12-17 for reference) as here we condition on both that the animal is available for detection and that it is detected ($\delta = 1$) while for the unconditional $f_u(y)$ we do not condition on it being detected ($\delta = 1$ or 0). For the unconditional $f_u(y)$, however, this cell probability is considered, although there are no observed counts associated with this cell to estimate it. We address this issue in the following section.

3.2.2   Formulating the integrated likelihood

Royle et al. (2004) showed, that when combining the multinomial likelihood for the unconditional case with the Poisson likelihood of the true number of animals N_{jpr} at

the point (or sampler in general), the latter may be integrated out over all possible values for N_{jpr}, i.e. $n_{jpr}, ..., \infty$. This combined likelihood may then be reduced to a Poisson likelihood for observed counts n_{jpri} in the ith interval of the sampler where the expected value is the product of the covariate model for N_{jpr} and the cell probability f_{u_i}. This reduced likelihood no longer contains the cell probability for the animals that were missed (f_{u_4} from Figure 3.1 and the example above). Although Royle et al. (2004) did not present the mathematical derivation for this, we present it in Appendix A.

Hence, for the integrated likelihood approach we use the unconditional cell probabilities f_{u_i}. We divide the observed counts n_{jpr} at the line (point) into the counts made within each distance interval i and consider these counts n_{jpri} as a Poisson random variable, $n_{jpri} \sim Poisson(\lambda_{jpr} f_i)$. The integrated likelihood function, where the Poisson rate λ_{jpr} (eqn (3.1)) is adjusted for imperfect detectability using f_{u_i} (eqn 3.2), is then defined as:

$$
\begin{aligned}
&L_{y,n}(\boldsymbol{\beta}, \sigma_b, \boldsymbol{\theta}_u) = \\
&\prod_{j=1}^{J} \int_{-\infty}^{\infty} \left[\prod_{p=1}^{P_j} \prod_{i=1}^{I_j} \prod_{r=1}^{R_j} \frac{(\lambda_{jpr} f_{u_i})^{n_{jpri}} \exp\left(-\lambda_{jpr} f_{u_i}\right)}{n_{jpri}!} \right] \times \\
&\frac{1}{\sqrt{2\pi\sigma_b^2}} \exp\left(-\frac{b_j^2}{2\sigma_b^2}\right) db_j.
\end{aligned}
\tag{3.3}
$$

As before in chapter 2, J equals the total number of sites, P_j and R_j refer to the total number of lines (points) and visits to a line (point) for the jth site, respectively and may vary between different sites. I_j refers to the outermost distance interval and is generally the same for each site. The component inside the square brackets of the right hand side of eqn (3.3) pertains to the Poisson likelihood of the observed counts

n_{jpri} and the component to the right of the brackets to the normal densities of the random effect coefficients.

By maximising this likelihood function, e.g. by using the *optim* or *nlm* function in R, all parameters are estimated simultaneously. Although it is easier to maximise the log-likelihood, the likelihood values inside the integral cannot be converted onto the log scale before integration, so that the log of eqn (3.3) becomes:

$$
\begin{aligned}
\log L_{y,n}(\boldsymbol{\beta}, \sigma_b, \boldsymbol{\theta}_u) = \\
\sum_{j=1}^{J} \log \int_{-\infty}^{\infty} \left[\prod_{p=1}^{P_j} \prod_{i=1}^{I_j} \prod_{r=1}^{R_j} \frac{(\lambda_{jpr} f_{u_i})^{n_{jpri}} \exp\left(-\lambda_{jpr} f_{u_i}\right)}{n_{jpri}!} \right] \times \\
\frac{1}{\sqrt{2\pi\sigma_b^2}} \exp\left(-\frac{b_j^2}{2\sigma_b^2}\right) db_j.
\end{aligned}
$$

$$(3.4)$$

3.2.3 Modelling heterogeneity in detection probabilities

Note that in the formulations above (eqns (3.3) and (3.4)), detections from all points are pooled to obtain parameter estimates for one common $f_u(y)$ assuming no heterogeneity in detection probabilities between different lines (points) or different detections. As for the two-stage approach (chapter 2), heterogeneity in detection probabilities can be modelled using stratification or multiple covariate distance sampling (MCDS) (Buckland et al., 2001, p. 88-92). Here, the global f_{u_i} are replaced with stratum-specific or covariate-specific $f_{u_{jpri}}$ in eqns (3.3) and (3.4). The $f_{u_{jpri}}$ may require further breaking down in case strata or covariates differ between detections during the same visit to a line or point (e.g. male vs. female birds). See section 2.2.1 for details.

3.2.4   Model selection

The function value returned by optimising equation (3.4) is the log-likelihood (logL) of the model evaluated at the maximum likelihood estimates of the parameters. This can be converted into a model selection criterion, e.g. AIC where AIC = -2logL + $2p$ (p = the number of parameters) (Akaike, 1979). In cases where the number of possible models is too large to consider, stepwise model selection may be used where one covariate is added to or removed from the model at a time. To obtain model-averaged estimates for parameters of interest, a weighted average may be taken across the models using AIC weights (Buckland et al., 1997).

3.2.5   Estimate of precision

Standard errors can be obtained using the Hessian matrix, which is calculated by optimisation routines such as the *optim* and *nlm* commands in R.

3.3   Case study 1: point transects of indigo buntings

Here, we analyse the indigo bunting data from case study 1 presented in the previous chapter. For details on these data see section 2.3.1 page 20. Results from the integrated approach are compared with those from the two-stage approach analysed and presented in chapter 2.

3.3.1 Analysis using the integrated likelihood approach

As before, we found in preliminary modelling of the detection function, that estimated detection probabilities dropped below 0.1 beyond 100m regardless of model choice, and so following recommendations of Buckland et al. (2001), we limited the analysis to the three innermost distance intervals (0-25, 25-50 and 50-100m). With just three intervals, and allowing a degree of freedom for assessing model fit, we considered only one-parameter models for the f_{u_i}. Hence, model selection for f_{u_i} included the half-normal and hazard-rate detection functions, where for the latter, the shape parameter was fixed. For the same reason, modelling detection heterogeneity was limited to stratification using one of the available factor covariates *year* (2006 or 2007), *type* (treated or control field) and *state* (9 levels). The covariates *windspeed*, *cloudcover* and *fogscore* were not considered as these were continuous covariates and/or did not reveal any significant influence on detection probabilities in preliminary analyses. We determined appropriate values for the fixed shape parameters for the global or the stratified hazard-rate functions in preliminary analyses. See section 2.3.2 for details (page 21).

The λ_{jpr} were modelled using eqn (3.1), with *year*, *type*, continuous *Julian day* and *state* as possible covariates. In these models, the parameter of interest was the covariate *type*. A significant *type* term in the λ_{jpr} model would indicate a difference in bird densities between the control and treated plots. The random effects term b_j was assumed normal with $b_j \sim N(0, \sigma_b^2)$. Analytical standard errors (ASEs) were obtained from the Hessian matrix.

Since complex models can be difficult to fit, we used stepwise forward model selection as described in section 3.2.4. For each contending model, eqn (3.4) was maximized

using the *optim* function in R, where, as for the analysis of the data with the two-stage approach in chapter 2, the total number of sites was $J = 446$ and the total number of distance intervals was $I_j = 3$. R_j ranged from 2 to 8 visits per site. As each site comprised two points, one control and one treatment, $p = 1$ or 2. We analyse the same 2924 counts at control or treatment points of the 446 sites that were made during the two survey years. These included a total of 3785 indigo bunting detections in the three innermost distance intervals.

3.3.2   Results

In the following, the results from the integrated approach are given in more detail than for the two-stage approach as we were interested in the effect of incorporating heterogeneity in detection probabilities on abundance model parameters (and vice versa) for this approach.

3.3.2.1   Model selection

For the integrated approach, forward stepwise model selection was started with the half-normal detection function for the detection model and the $\beta_0 + b_j$ (intercept + random effect) model for λ_{jpr} (Table 3.1). Considering the global and stratified hazard-rate models next (stratified by either *year*, *type* or *state*) for the detection function indicated that a state-stratified hazard-rate model gave the lowest AIC values. With this model for the detection function, covariates were added to the abundance model one at a time and retained if inclusion lowered the AIC value. Here,

the best model by AIC included the covariates *type*, *Julian day* and *state*. We did not consider model averaging in this case as the difference in AIC values between the best and the second best model was 138, so the model-averaged estimates would be the same as the estimates under the best model. The equivalent models were selected for the two-stage approach although here separately for each of the two stages (see section 2.3.3 in chapter 2 page 24).

Table 3.1: Models included in the forward stepwise model selection for the integrated approach including the half-normal (HN) and the global and stratified hazard-rate (HR) detection functions for $f_{u_{jpri}}$ and the inclusion of four covariates for λ_{jpr} in addition to the intercept β_0 and the random effects b_j. $\triangle$AIC is given in relation to the overall best model (model 9). Improved? refers to whether in this iterative model selection process starting with model 1 the respective model yielded an improved AIC compared to the previous and whether it should be retained.

ID	$f_{u_{jpri}}$ Model	λ_{jpr} Model	Log-Lik	Parameters	$\triangle$AIC	Improved?
1	HN global	$\beta_0 + b_j$	-7327.74	3	375.92	NA
2	HR global	$\beta_0 + b_j$	-7296.49	3	313.41	yes
3	HR by year	$\beta_0 + b_j$	-7295.73	4	313.91	no
4	HR by type	$\beta_0 + b_j$	-7268.85	4	260.13	yes
5	HR by state	$\beta_0 + b_j$	-7248.12	11	232.68	yes
6	HR by state	$\beta_0 + b_j + year$	-7247.77	12	233.99	no
7	HR by state	$\beta_0 + b_j + type$	-7205.50	12	149.45	yes
8	HR by state	$\beta_0 + b_j + type + JD$	-7198.99	13	138.41	yes
9	HR by state	$\beta_0 + b_j + type + JD + state$	-7121.78	21	0	yes

3.3.2.2 Comparing contending models from the integrated approach

Parameter estimates with standard errors for each contending model for the integrated approach from Table 3.1 are shown in Table 3.2. Substantial differences in parameter

estimates between models were obtained by including *state* in the λ_{jpr} model (model 9). This resulted in an increase in detection function parameters for seven states and a decrease for the remaining two. Given the same truncation distance and fixed shape parameters, larger scale parameters of a hazard-rate detection function translate into larger estimates of $f_{u_{jpri}}$ (i.e. proportions of N_{jpr} that were detected) for the respective strata. Including *state* in the abundance model also led to a decrease in the random effect standard deviation. The change in detection function parameters was probably because with *state* in the abundance model, the state-specific $f_{u_{jpri}}$ represent proportions of the estimates of the expected N_{jpr} that are modelled as a function of *state* (as well as of *type* and *Julian day*), while before they represented proportions of the expected N_{jpr} that were not modelled as a function of *state*. In addition to a change in point estimates for parameters, the standard errors increased for all detection function parameters after including *state* in the λ_{jpr} model (model 9 compared to models 5-8). The decrease in the random effect standard deviation for model 9 indicated that the *state* covariate modelled part of the variation absorbed by the random effects coefficients in the λ_{jpr} models 1-8 (Table 3.2).

Table 3.2: Maximum likelihood estimates (MLE) and analytical standard errors (ASE) for parameters of contending models for the integrated approach (models 1 through 9, Table 1). For the $f_{u_{jpri}}$ model, HN and HR refer to the half-normal and hazard-rate detection functions respectively.

Model	1 MLE (ASE)	2 MLE (ASE)	3 MLE (ASE)	4 MLE (ASE)	5 MLE (ASE)	6 MLE (ASE)	7 MLE (ASE)	8 MLE (ASE)	9 MLE (ASE)
$f_{u_{jpri}}$ Model Parameters									
Scale HN	37.61 (0.46)								
Scale HR		26.81 (0.66)							
Scale HR Year 2006			27.22 (0.84)						
Scale HR Year 2007			26.45 (0.81)						
Scale HR Type CONTROL				24.13 (0.68)					
Scale HR Type TREAT				28.42 (0.76)					
Scale HR State GA					17.21 (1.71)	17.08 (1.61)	17.20 (1.68)	18.74 (1.81)	45.85 (9.33)
Scale HR State IL					35.20 (2.02)	35.34 (2.00)	35.20 (2.02)	35.29 (2.07)	36.03 (3.21)
Scale HR State IN					26.48 (1.67)	26.31 (0.02)	26.48 (1.68)	26.49 (1.69)	27.31 (2.66)
Scale HR State KY					31.19 (1.06)	31.61 (1.06)	31.19 (1.05)	31.54 (0.05)	29.63 (1.23)
Scale HR State MO					24.56 (1.23)	24.61 (1.19)	24.56 (1.23)	24.41 (1.33)	41.31 (3.03)
Scale HR State MS					34.71 (2.14)	35.13 (2.14)	34.71 (2.12)	35.44 (2.23)	38.50 (3.31)
Scale HR State OH					22.05 (1.14)	22.36 (0.07)	22.06 (1.14)	22.23 (0.03)	27.15 (2.16)
Scale HR State SC					43.38 (2.33)	43.14 (2.29)	43.37 (2.33)	43.21 (2.36)	57.79 (4.40)
Scale HR State TN					28.92 (1.53)	29.94 (1.56)	28.93 (0.03)	30.24 (1.63)	26.11 (0.04)
λ_{jpr} Model Parameters									
Random Effects									
SD	0.66 (0.03)	0.66 (0.03)	0.66 (0.03)	0.66 (0.03)	0.60 (0.01)	0.59 (0.01)	0.60 (0.01)	0.58 (0.01)	0.50 (0.02)
Fixed Effects									
Intercept	1.31 (0.04)	1.53 (0.05)	1.52 (0.05)	1.55 (0.04)	1.54 (0.02)	1.54 (0.0033)	1.38 (0.02)	0.39 (0.02)	-0.99 (0.28)
Year 2007						0.002 (0.018)			
Type TREAT							0.30 (0.02)	0.30 (0.02)	0.30 (0.02)
Julian Day								0.0056 (0.0003)	0.0053 (0.0005)
State IL									1.46 (0.32)
State IN									1.50 (0.32)
State KY									2.00 (0.29)
State MO									0.53 (0.29)
State MS									1.25 (0.31)
State OH									1.11 (0.30)
State SC									0.59 (0.31)
State TN									2.14 (0.28)

3.3.2.3 Comparing best models from the integrated and two-stage approach

For the best fitting model, estimates of the scale parameters of the hazard-rate detection function from the integrated approach ranged between 26.11 for Tennessee (ASE=0.04, fixed shape=2.0, Table 3.3) and 57.79 for South Carolina (ASE=4.40, fixed shape=3.0). They were generally larger for the integrated approach compared to the two-stage approach except for Mississippi where the estimate was slightly smaller. The discrepancy between parameter estimates in the two approaches was less than 10% in five states and up to 24% in the remaining four.

In the following we refer to baseline expected number of male indigo buntings within the plot area a for the values returned by the respective models after setting the covariates to $type =$ Control, $Julian\ day = 174$ (the mid-point of all days surveyed), $state =$ GA and incorporating a contribution of the random effects term using the mean of $\log N(0, \hat{\sigma}_b^2)$. To compare these baselines for the two approaches, we applied the following transformation: $\exp(\hat{\beta}_0 + \hat{\beta}_2 \times 174 + 0.5 \times \hat{\sigma}_b^2)$ using the coefficient values from Table 3.3 for the respective approaches (see eqns (3.1) page 33 and (2.6) page 16 for details), with $\hat{\beta}_0$ being the intercept estimate and $\hat{\beta}_2$ being the estimate for the $Julian\ day$ coefficient. For the two-stage approach the resulting value also needed to be multiplied by the search area of the plot ($a = \pi w^2 = 31416\text{m}^2$), as the intercept represents birds/m^2 (as opposed to birds per search area of the plot for the integrated approach). The estimates of the baseline expected numbers of male indigo buntings were 1.05 (ASE=0.29) and 1.43 (BSE=0.59) individuals per search area a (or 33.52 (ASE=9.12) and 43.51 (BSE=18.91) birds per km^2) for the integrated and two-stage approach, respectively.

Table 3.3: Maximum likelihood estimates (MLE), analytic (ASE) and bootstrap (BSE, two-stage approach only) standard errors for model parameters obtained by the integrated and the two-stage approach for best models. Shape parameters for the one-parameter hazard-rate detection function were fixed.

Model Paramters	Integrated likelihood		Two-stage			Shape
	MLE	ASE	MLE	ASE	BSE	
Detection model						
Scale State GA	45.85	9.33	37.27	7.72	8.12	2
Scale State IL	36.03	3.21	34.42	2.86	3.17	2.5
Scale State IN	27.31	2.66	24.34	2.35	4.92	2
Scale State KY	29.63	1.23	27.75	1.13	1.52	2.5
Scale State MO	41.31	3.03	37.78	3.14	2.86	2
Scale State MS	38.50	3.31	38.73	3.33	4.21	2
Scale State OH	27.15	2.16	24.59	1.97	1.94	2
Scale State SC	57.79	4.40	56.30	4.13	6.85	3
Scale State TN	26.11	0.04	21.08	1.74	3.51	2
Abundance model						
Random effects						
Standard deviation	0.50	0.02	0.49	-	0.04	
Fixed effects						
Intercept N_{jpri}	-0.99	0.28	-	-	-	
Intercept D_{jpr}	-	-	-10.91	0.29	0.43	
Type Treated	0.30	0.02	0.30	0.03	0.04	
Julian Day	0.0053	0.0005	0.0046	0.0017	0.0018	
State IL	1.46	0.32	1.20	0.18	0.38	
State IN	1.50	0.32	1.34	0.18	0.49	
State KY	2.00	0.29	1.79	0.17	0.36	
State MO	0.53	0.29	0.32	0.16	0.36	
State MS	1.25	0.31	0.91	0.17	0.37	
State OH	1.11	0.30	0.92	0.17	0.37	
State SC	0.59	0.31	0.28	0.18	0.39	
State TN	2.14	0.28	2.12	0.17	0.44	

The remaining fixed effects represent proportional changes compared to the respective baseline estimates. The *type* coefficient was the same for the two approaches (0.30, Table 3.3) with ASE=0.02 for the integrated approach and ASE=0.03 and BSE=0.04 for the two-stage approach. This indicated a 35.0% increase in density or abundance on the treated fields ($\exp(0.30) = 1.35$). For the remaining fixed effect coefficients in the λ_{jpr} model, parameter estimates were again larger for the integrated approach. The coefficient for the continuous covariate *Julian day* was 0.0053 (ASE=0.0005) for the integrated likelihood approach and 0.0046 for the two-stage approach (ASE=0.0017, BSE=0.0018).

Discrepancies in estimates for the *state* coefficients between the two approaches were more pronounced than for detection function parameters, ranging between 1 and 111%. Larger coefficients again translate into proportionately larger increases in λ_{jpr} for the respective factor levels compared to the baseline expected number of birds. The random effect standard deviation was slightly larger for the integrated likelihood approach (0.50, ASE=0.02 vs. 0.49, BSE=0.04).

Comparing ASEs between the two approaches, detection function parameters were smaller for the integrated approach for three states and larger for the remaining six, whereas ASEs from the integrated approach were smaller than BSEs from the two-stage approach in five states and larger in four. ASEs for the intercept and the coefficients for *type* and *Julian day* in the λ_{jpr} models were smaller for the integrated approach than ASEs and BSEs for the two-stage approach. For the *state* coefficients, ASEs from the integrated approach were always larger than ASEs from the two-stage approach but always smaller than BSEs.

3.4   Discussion

Designed experimental studies generally have an advantage over purely observational studies in that they allow inference on cause and effect of a treatment. Designed experiments allow attribution of the change in animal abundances directly to the treatment, while purely observational studies do not.

For the CP-33 Monitoring Program, the experimental design comprised sites, each with a pair of survey points, one in a buffered treatment field and one in a nearby unbuffered control field; repeat surveys were conducted concurrently at both points of each site. This study is possibly unique due to its scale (over 400 sites with repeat surveys each year at each site). Analysis of such data is complex and it is critical to attribute the causes for variations in observed counts to the correct sources by appropriate model specification together with objective model selection criteria.

The aim of our case study was to determine whether buffers improved habitat for birds which would be indicated by higher densities near buffered compared to unbuffered fields. Our analyses demonstrated that implementation of buffers resulted in an increase of indigo bunting densities by 35%. Previous studies have shown beneficial effects of such buffers for birds (e.g. Evans et al., 2013). Conover et al. (2011) showed that field buffers increased nesting activities along field margins for a range of birds, including indigo buntings. In contrast, Riddle and Moorman (2010) showed that implementing field borders had no beneficial effect on nesting success of indigo buntings. However, their effort was limited to 12 hog farms and his inference limited to breeding success. Besides the potential for additional breeding habitat, buffers may also provide new habitat for breeding and escape cover. While our results give evidence of larger densities of indigo buntings on buffered fields, behavioural data

would be necessary to make inference on how these birds make use of this habitat.

It is important to avoid false inference due to random variation. Hence, Buckland et al. (2009) recommended that repeat surveys be made on plots. This, however, possibly introduces correlation between counts at the same site which we accommodated by including a random effect for site in the abundance model. This is a new technique in the context of distance sampling analysis methods where covariate models for abundance have generally been limited to fixed effects. If sites are few and budget limitations allow for multiple repeats of counts at each site, site may indeed be included as a fixed effect in the abundance model, although inference is then restricted to the sites surveyed. For large-scale studies, such as our case study, this strategy would require the estimation of too many parameters. In any case, we wish to draw inference on the effects of field buffers generally, and not just on those field buffers in the survey, and inclusion of a random effect for site in the abundance model allows us to do this.

Potential correlations between counts at the same site may be accommodated by expanding the two-stage approach of Buckland et al. (2009) by including a random effect for site in the count model (as presented in chapter 2). However, the two-stage approach conditions on the first-stage detection function model for the second-stage count model and uncertainty from the first stage does not propagate into the second stage. For our case study, this was evident in artificially small analytical standard errors for *state* in the λ_{jpr} model obtained with the two-stage approach. Underestimation of standard errors may result in retaining the wrong covariates in the final model. This issue may generally be avoided using the integrated likelihood approach where all parameters are estimated simultaneously.

We considered a collection of abundance models and models for the detection function. One may argue that modelling raw counts (i.e. not adjusted for imperfect detection on the plot) in a GLMM without an offset might have been sufficient for inference on the parameter of interest. For our case study, inference on the parameter of interest, the *type* coefficient, would have remained the same regardless of choice of model (those including *type* in the plot abundance models, models 7-9, Table 3.2) or approach (Table 3.3). As the best detection function did not include the *type* covariate and *state* and *type* were not correlated (absolute correlation between *type* and any *state* coefficients was <0.01), inference on this parameter would have also remained the same if modelling raw counts in a GLMM without any offset. However, inference on parameter estimates for a covariate that is both in the detection function and the λ_{jpr} model may differ substantially between the two approaches (or when comparing either of these approaches to modelling raw counts). Dissimilarities between estimates for *state* both in the detection function and the abundance model likely resulted from estimating all parameters in one step for the integrated approach as opposed to two steps for the two-stage approach. Conceptually the difference between the approaches is that for the integrated approach, we assume that the patterns by which animals distribute themselves in the study area (and resulting densities) and the observation process influence each other, while they are considered as separate processes for the two-stage approach. We argue – along the lines of Royle et al. (2004) and Johnson et al. (2010) – that the former case is the more realistic assumption.

A source of unmodelled variation in large-scale studies often is the large number of observers. Ideally, the number of observers in a study is small and all observers are well trained in the distance sampling protocol and distance estimation. Inter-observer

differences might then be modelled using MCDS analyses for the detection function (see section 3.2.3 page 39). For our case study, the data were collected by 172 observers, so that too many parameters would be required to include observer as a fixed effect. Stratifying the detection function by *state* possibly captured some of the inter-observer variation. An alternative strategy is to include a random effect for observer in the detection function model, requiring the estimation of only one parameter, the random effect standard deviation (e.g. Yuan et al., unpublished manuscript). We address this issue in chapter 6. However, we demonstrated that modelling heterogeneity in detection probabilities may have a strong influence on parameters in the abundance model and that using a model selection routine is necessary to determine which parameters should be retained in the final model.

We expect designed distance sampling experiments to become widely used for assessing effectiveness of conservation measures, and for environmental impact studies. The use of random effects as described here allows correlations in multiple counts from a single sampling unit to be accommodated, and allows inference to be extended to a wider area for which the sites are a representative sample, thus strengthening the ability of wildlife and natural resource managers to evaluate the implications of changes in the environment.

Chapter 4

Building hierarchical models with an integrated likelihood for distance sampling data

4.1 Introduction

Bayesian methods are becoming increasingly popular for modelling wildlife populations and abundances (e.g. Buckland et al., 2000; Marcot et al., 2001; Durban and Elston, 2005; Schmidt et al., 2009; King et al., 2010). However, few distance sampling studies have taken a Bayesian approach. Karunamuni and Quinn (1995) developed a Bayes estimator for $f(0)$ using a half-normal detection function. Other studies have built upon this approach. Eguchi and Gerrodette (2009) extended this model by including a binomial likelihood for the encounter rate along the line and described a joint posterior distribution for the count model and effective strip width, Gimenez et al. (2009) implemented an estimator for $f(0)$ in WinBUGS software, while Zhang

(2011) developed an empirical Bayes estimator for $f(0)$. All these studies follow a similar approach in that they develop their methods for line-transect data using the half-normal detection function and use the Gibbs sampler to explore the parameter space.

Gibbs sampling is potentially easier to implement than the alternative, the Metropolis-Hastings (MH) update, when standard distribution functions are used. The joint posterior distribution, and consequently the set of full conditionals, however, may be difficult to obtain when using non-standard distributions. As the detection function for distance sampling data often involves such non-standard distributions (e.g. the hazard-rate key function) we use the MH algorithm.

Our Bayesian approach to density estimation from distance sampling data presented here is applicable to both line and point transect data and to any detection function. We use an integrated likelihood that combines the likelihood components of the detection and count models. For the latter, we use a Poisson likelihood for the distance sampling counts that incorporates a component corresponding to the detection function, thus allowing for imperfect detection at the line or point. In comparison to Eguchi and Gerrodette (2009) who use a binomial likelihood to scale up from density at the line to density in the study area, our Poisson model relates animal counts to covariates. This approach does not rely on random placement of samplers in the study area (Hedley and Buckland, 2004). We include a random effect for site in the Poisson model to accommodate correlated counts due to e.g. repeat counts at the same site. The parameter space is explored using a Metropolis-Hastings (MH) update, different prior distributions for the parameters are easily implemented, and a reversible jump

Markov chain Monte Carlo (RJMCMC) algorithm allows model uncertainty to be incorporated. This may include different key functions for the detection function model and different covariate combinations for both the detection function and the count models.

These developments were motivated by our case study 2, a large-scale experimental study to assess the effects of establishing conservation buffers along field margins on density of several species of conservation interest such as the northern bobwhite. Pairs of points were set up at the edge of fields in farmland in 13 states in the USA. These pairs of points consisted of one point on a buffered treatment field and one on a nearby non-buffered control field and will be referred to as sites in the following. Point transect surveys of coveys were conducted at least once but up to three times per year in autumn $2006 - 2008$. For details on the data see section 2.4.1 in chapter 2 page 26.

In the following we begin by developing the integrated likelihood (section 4.2). This integrated likelihood combines the two likelihood components of the two-stage approach presented in chapter 2, hence uses the conditional probability density function of observed distances (Buckland et al., 2001). In contrast, for the integrated likelihood from chapter 3 we use the unconditional formulation (Royle et al., 2004). We then describe the Bayesian approach (section 4.3), analyse bobwhite covey data using our Bayesian approach (and a maximum likelihood approach, the two-stage approach from chapter 2, for comparison) (section 4.4), before concluding with a discussion (section 4.5) where we contrast our Bayesian approach with existing studies dealing with distance sampling likelihoods (e.g. Buckland et al., 2004; Eguchi and Gerrodette, 2009, Johnson et al., 2010, Royle et al., 2004).

4.2 An integrated likelihood for distance sampling data

To obtain abundance estimates of a population of interest using distance sampling methods, lines or points may be placed in the study area according to some design (see Buckland et al., 2001, for details). Each line or point is surveyed at least once following the distance sampling protocol where the observer travels down the line (line transects) or remains at the point for a fixed amount of time (point transects). Detections are recorded along with the perpendicular distance from the line to the detection (or radial distance from the point to the detection). These distances may be recorded exactly or in predetermined distance bands. Thus, surveys of this type produce two types of data: firstly, the observed distances y_e with $e = 1, 2, 3, ..., n$ (n being the total number of detections).

Secondly, the observed counts n_p at point p or encounter rate along the line (i.e. observed counts along line p divided by its length: n_p/l_p) with $\sum_{p=1}^{P} n_p = n$. In the case that distances are recorded in intervals, the counts n_p at line or point p are divided into the number of counts in each of i distance intervals n_{pi} (instead of y_e exact distance measurements). For modelling the detection function, the latter may be pooled across all lines/points to form the n_i, i.e. the counts in the i distance interval from all surveyed lines/points.

In addition, if detections are made of groups of animals (rather than single individuals), a third type of data generated from a distance sampling survey is cluster size

s_e which represents the number of individuals within the eth detected group. For simplicity, we ignore cluster sizes for this study. Methods could, however, easily be extended to also include a model for cluster sizes.

In contrast to many existing covariate models for distance sampling data (e.g. Hedley and Buckland, 2004; Buckland et al., 2009), the proposed integrated likelihood deals with both components of the data simultaneously. It consists of the likelihood components for the detection function, which is denoted by $L_y(\boldsymbol{\theta})$ for exact distance data (see eqn (4.5) below for interval data), and the Poisson likelihood for observed counts, $L_n(\boldsymbol{\beta}|\boldsymbol{\theta})$. We use $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ to summarise the detection function and Poisson model parameters, respectively. These are defined in more detail below. The integrated likelihood is the product of the two components (modified from Buckland et al., 2004, ch. 2):

$$L_{n,y}\left(\boldsymbol{\beta}, \boldsymbol{\theta}\right) = L_y(\boldsymbol{\theta})L_n(\boldsymbol{\beta}|\boldsymbol{\theta}). \tag{4.1}$$

We consider each individual likelihood component in $L_{n,y}\left(\boldsymbol{\beta}, \boldsymbol{\theta}\right)$ and begin with $L_y(\boldsymbol{\theta})$. Here, we use the conditional formulation for $f(y)$, as opposed to the unconditional formulation from chapter 3. This allows us to define this likelihood for exact and interval distance data, while the formulation for the integrated likelihood from chapter 3 was limited to interval data. Let $f(y|\boldsymbol{\theta})$ denote the conditional probability density function of observed distances which is given as (Thomas et al., 2010):

$$f(y|\boldsymbol{\theta}) = \frac{\pi(y)g(y|\boldsymbol{\theta})}{\int_0^w \pi(y)g(y|\boldsymbol{\theta})dy}, \tag{4.2}$$

where y is the observed distance from the line (point) and w is the truncation distance (i.e. the furthest distance from the line or point included in the analysis). As before (chapters 2 and 3), $\pi(y)$ describes the expected distribution of animals with respect to the line ($\pi(y) = 1/w$) or point ($\pi(y) = 2y/w^2$). The detection function $g(y|\boldsymbol{\theta})$ may be modelled e.g. as half-normal ($g(y|\boldsymbol{\theta}) = \exp\left(-y^2/2\sigma^2\right)$, with $\boldsymbol{\theta} = \{\sigma\}$) or hazard-rate ($g(y|\boldsymbol{\theta}) = 1 - \exp\left(-(y/\sigma)^{-\tau}\right)$, with $\boldsymbol{\theta} = \{\sigma, \tau\}$). The likelihood, which is conditional on the number of detections n, may be expressed as (Buckland et al., 2004, p. 16):

$$L_y\left(\boldsymbol{\theta}\right) = \prod_{e=1}^{n} f\left(y_e|\,\boldsymbol{\theta}\right), \tag{4.3}$$

where y_e refers to eth detection.

When detections are recorded in distance intervals, let f_i denote the probability that a detected animal is in the ith interval which is delineated by the cutpoints c_{i-1} and c_i:

$$f_i(\boldsymbol{\theta}) = \frac{\int\limits_{c_{i-1}}^{c_i} f\left(y_i|\,\boldsymbol{\theta}\right)dy}{\int\limits_{0}^{w} f\left(y_i|\,\boldsymbol{\theta}\right)dy}, \tag{4.4}$$

where the truncation distance, w corresponds to the outermost cutpoint. Note that the f_i represent proportions of the observed counts on the plot (as opposed to proportions of the true number of animals on the plot when using the unconditional f_{u_i} from chapter 3). As a consequence, we do not consider the cell probability associated with the animals that were missed on the plot. Here, the sum of the I cell probabilities equals one ($\sum\limits_{i=1}^{I} f_i = 1$) and the multinomial likelihood L_{yG}, given in the following equation, replaces L_y in eqn (4.1):

$$L_{yG}(\boldsymbol{\theta}) = \left(\frac{n!}{\prod\limits_{i=1}^{I} n_i!} \right) \prod_{i=1}^{I} f_i(\boldsymbol{\theta})^{n_i}, \qquad (4.5)$$

where n_i is the number of detected animals in the ith interval.

Note that in eqns (4.3) and (4.5), detections from all sites are pooled in one detection function. For modelling heterogeneity using multiple covariate distance sampling (MCDS) methods, the scale parameter σ of the half-normal or hazard-rate detection function is modelled as a function of covariates ($\sigma(\boldsymbol{z}) = \delta_0 \times \exp(\sum_{q=1}^{Q} z_q \delta_q)$, where $\delta_q, q = 0, 1, 2, ..., Q$ replace σ in $\boldsymbol{\theta}$) (Marques and Buckland, 2003).

For the log-linear Poisson model, $L_n(\boldsymbol{\beta}|\boldsymbol{\theta})$ we begin by considering counts n_{jpr} at visit r to line or point p at site j as a Poisson random variable with mean λ_{jpr}. To adjust these counts for imperfect detection out to distance w, $f(y|\boldsymbol{\theta})$ from eqn (4.2) is used to estimate the effective area $\nu(\boldsymbol{\theta})$ which is defined as the area beyond which as many animals are seen as are missed within (Buckland et al., 2001). Consequently, dividing counts by the area effectively surveyed along the line (point) gives a valid estimator for density. For line transects, $\nu = 2l_p \int_0^w g(y|\boldsymbol{\theta})dy$, where l_p is the length of the line surveyed; for point transects $\nu = 2\pi \int_0^w yg(y|\boldsymbol{\theta})dy$. Note that again, these definitions for ν are given for the case where all detections are pooled in one detection function. When modelling heterogeneity, e.g. by using MCDS methods, the effective area may vary between different lines (points) and the global ν becomes ν_{jpr}. These may require further breaking down in case covariates are included that pertain to individual detections as opposed to individual visits to a line/point. By including the effective area as an offset, counts are divided by the effective area: $E[n_{jpr}]/\nu_{jpr} = \lambda_{jpr}/\nu_{jpr}$ or $E[n_{jpr}] = \lambda_{jpr}$ with:

$$\lambda_{jpr}\left(\boldsymbol{\beta}|\boldsymbol{\theta}\right) = \exp\left(\beta_0 + b_j + \sum_{k=1}^{K} x_{kjpr}\beta_k + \ln\left(\nu_{jpr}(\boldsymbol{\theta})\right)\right), \tag{4.6}$$

where β_0 is the fixed effect intercept, b_j the random effect for site j $(b_j \sim N\left(0, \sigma_b^2\right))$, x_{kjpr} the covariate values of the k fixed effect covariates measured during visit r to that line or point and β_k the associated coefficients. Vector $\boldsymbol{\beta} = \{\beta_0, \beta_1, \beta_2, ..., \beta_K, \sigma_b\}$ denotes the parameters associated with the covariates affecting densities and the random effect standard deviation. Here, the combination of fixed and random effects in (4.6) models density at the line (point) rather than counts while the observed counts remain the response variable.

Eqn (4.6) is given for the general case where lines or points that may produce correlated counts, due to closeness in space and/or due to repeated measurements at the same line (point), are grouped together as site j. The inclusion of a random effect for site accommodates covariances for these measurements. However, in cases where lines (points) follow a random survey design (Buckland et al., 2001) and each line (point) is surveyed only once, the random effect term may be omitted.

Using this model for λ_{jpr}, the likelihood for the count model, conditional on the estimate of the effective area, may be expressed as:

$$L_n(\boldsymbol{\beta}|\boldsymbol{\theta}) = \prod_{j=1}^{J} \int_{-\infty}^{\infty} \left(\prod_{p=1}^{P_j} \prod_{r=1}^{R_j} \frac{(\lambda_{jpr})^{n_{jpr}} \exp\left(-\lambda_{jpr}\right)}{n_{jpr}!}\right) \times \frac{1}{\sqrt{2\pi\sigma_b^2}} \exp\left(-\frac{b_j^2}{2\sigma_b^2}\right) db_j, \tag{4.7}$$

where J refers to the total number of sites, and P_j and R_j refer to the total number of lines (points) at and visits to the jth site, respectively. $L_n(\boldsymbol{\beta}|\theta)$ forms the second likelihood component in eqn (4.1). Note that in a maximum likelihood context the

likelihood function including a random effect (for which normality is assumed) is generally formulated with an integral as shown in eqn (4.7) as the random effect is integrated out analytically (or by approximation) and the individual coefficients b_j are not estimated (e.g. McCulloch and Searle, 2001). In the Bayesian context, however, the random effect is not integrated out analytically. Here, we use a data augmentation scheme where the individual coefficients b_j are included in the model specification and the updating process (see below).

4.3 The Bayesian approach

4.3.1 Hierarchical models

Using a Bayesian approach, random effects models can be implemented using hierarchical models where the standard deviation of the random effect (σ_b from eqn (4.7)) is considered to have a distribution rather than a fixed value (which is also true for the remaining fixed effect coefficients) (Davison 2003). Individual random effects coefficients (the b_j from eqns (4.6) and (4.7)) are fitted in the model and updated at each iteration of the chain (see below).

Prior beliefs regarding the parameters such as knowledge obtained from a different study previously conducted, may be included in the current study via the prior distribution. This may allow inference on model parameters in cases where too little data exists in the current study to obtain maximum likelihood estimates with great precision (e.g. Eguchi and Gerrodette, 2009). However, in the following we assume the case where no prior information exists and place uniform priors on all parameters $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ (eqns (4.1)-(4.7)).

4.3.2   MCMC algorithm

A Markov chain Monte Carlo algorithm is used to explore the posterior distribution of the parameters given the data and obtain summary statistics of interest. Commonly used MCMC methods are the Gibbs sampler and Metropolis-Hastings (MH) update. The Gibbs sampler samples directly from the conditional posterior distribution of the respective parameter. In particular, during each iteration t, each parameter is updated consecutively by drawing a sample from the posterior distribution conditional on the current state of the other parameters. This approach is relatively convenient to use when using standard distributions (e.g. Spiegelhalter et al., 2003; King et al., 2010), e.g. by using the user-friendly software WinBUGS. However, when using non-standard distributions such as the half-normal detection function these methods become more complex. For example, Gimenez et al. (2009) implemented the 'zero-trick' in WinBUGS to obtain the posterior distribution for estimating animal density from distance sampling data. For this trick they considered a set of n detections with associated perpendicular distances y_e ($e = 1, 2, 3, ..., n$) where the likelihood contribution of each detection equals a likelihood term L_e. This likelihood contribution is calculated using the half-normal detection function. The 'zero-trick' implies using a set of n zeros, each with an assumed Poisson distribution, $P(\phi_e)$. The essential part of this trick is that the Poisson likelihood with expected value ϕ of a zero observation equals $\exp(-\phi)$. Hence, they then set $\phi_e = -\log(L_e)$ where L_e represents the contribution of the eth observed perpendicular distance to the likelihood. Then, $\sum_{e=1}^n \log(L_e) = \sum_{e=1}^n -\phi_e$.

However, we were interested in developing methods that do not rely on tricks or transformations and are easily implemented for a range of detection function models. Hence, we focus on the MH update (Hastings, 1970; Metropolis et al., 1953) in the following, as some of the likelihood functions that may be used to form the posterior conditional distributions of parameters are non-standard (e.g. half-normal or hazard-rate detection function that may include a covariate model for the scale parameter). In particular, we use a random walk single-update MH algorithm with normal proposal density where we cycle through each parameter in $L_{n,y}(\boldsymbol{\beta}, \boldsymbol{\theta})$. To use a simple scenario, assume $\boldsymbol{\beta} = \{\beta_0, \sigma_b\}$. Then, e.g. for parameter β_0 with current value β_0^t we propose to move to a new state, β_0', with $\beta_0' \sim N\left(\beta_0^t, \sigma_{\beta_0}^2\right)$ (e.g. Hastings, 1970; Davison, 2003). This newly proposed state is accepted as the new state with probability $\alpha(\beta_0'|\beta_0^t)$ given by (King et al., 2010):

$$\alpha(\beta_0'|\beta_0^t) = min\left(1, \frac{L_{n,y}(\beta_0', \sigma_b^t, \boldsymbol{\theta}^t)p(\beta_0')q(\beta_0^t|\beta_0')}{L_{n,y}(\beta_0^t, \sigma_b^t, \boldsymbol{\theta}^t)p(\beta_0^t)q(\beta_0'|\beta_0^t)}\right). \tag{4.8}$$

Here, $q(\beta_0'|\beta_0^t)$ denotes the proposal density of β_0' given the current state is β_0^t. We note that the terms $q(\beta_0^t|\beta_0')$ and $q(\beta_0'|\beta_0^t)$ cancel in the acceptance probability since we use a symmetrical proposal distribution. The analogous MH updates are used for random effect coefficients. Proposal variances are chosen via pilot-tuning (Gelman et al., 1996).

4.3.3 Model selection: reversible jump MCMC

To discriminate between competing models, we treat the model itself as a parameter and form the joint posterior distribution over both parameters and models which is

given (up to proportionality) by:

$$\pi_{n,y}\left(\boldsymbol{\beta}_m, \boldsymbol{\theta}_m, m\right) \propto L_{n,y}(\boldsymbol{\beta}_m, \boldsymbol{\theta}_m, m)p(\boldsymbol{\beta}_m, \boldsymbol{\theta}_m|m)p(m), \qquad (4.9)$$

where $L_{n,y}(\boldsymbol{\beta}_m, \boldsymbol{\theta}_m, m)$ denotes the probability density function of the data given current parameter values $\boldsymbol{\beta}_m$ and $\boldsymbol{\theta}_m$ and model m, $p(\boldsymbol{\beta}_m, \boldsymbol{\theta}_m|m)$ the prior distribution for model parameters $\boldsymbol{\beta}_m$ and $\boldsymbol{\theta}_m$ and $p(m)$ the prior probability of model m. As this distribution is too complex to sample from directly, an RJMCMC algorithm is used to move within both parameter and model space simultaneously (Green, 1995).

For this RJMCMC algorithm each iteration involves two steps; step 1: update parameters given the current model using the MH algorithm (within model move) as described above in section 4.3.2; step 2: update the model using a reversible jump (RJ) algorithm (between model move). During RJ step, model m conditional on the current parameter values is updated. This move involves a proposal to update the model itself; suppose the chain is in model m and we propose to move to model m'. A bijective function describes the relationship between the current and proposed parameters and is used to convert parameters from model m to parameters for model m'. In a simple scenario, say, where model m contains parameters $\boldsymbol{\beta} = \{\beta_0, \beta_1\}$ and model m' contains parameters $\boldsymbol{\beta}' = \{\beta_0', \beta_2'\}$, the bijective function might be expressed as an identity function where:

$$\beta_0' = \beta_0$$
$$u' = \beta_1$$
$$\beta_2' = u. \qquad (4.10)$$

Here u and u' are random samples from some proposal distributions for the respective parameters. The acceptance probability may then be expressed as:

$$A = \frac{\pi_{n,y}(\boldsymbol{\beta}', m')P(m|m')q'(u')}{\pi_{n,y}(\boldsymbol{\beta}, m)P(m'|m)q(u)} \, |J| \,, \tag{4.11}$$

where $P(m'|m)$ denotes the probability of proposing to move to model m' given that the chain is in model m, $q(u)$ and $q'(u')$ are the proposal densities of u and u' and $|J|$ is the Jacobian (which equals one if the bijective function is the identity function). For the RJ step in general, two main strategies may be followed. In cases where models differ only in the combination of the same set of covariates, a single RJ step may involve going through each covariate and proposing to delete or add it depending on whether it is in the current model or not. This involves generating a value for the new parameter from a proposal distribution (if we propose to add it) or setting it to zero (if we propose to delete it) and calculating the acceptance probability each time we propose to add or delete a parameter.

In those cases where all parameters of the newly proposed model change, one RJ step involves generating new values for all parameters of the new model and accepting or rejecting the new model based on the above acceptance probability. A proposed move from a half-normal detection function model to a hazard-rate model represents a simple example for this scenario.

Posterior model probabilities are estimated as the proportion of time the chain spent in a particular model after the burn-in.

4.4 Case study 2: point transect surveys of northern bobwhite coveys

Here, we analyse the northern bobwhite covey data from our case study 2 presented in chapter 2 using an RJMCMC algorithm. For details on the data see section 2.4.1 page 26. Results from the Bayesian approach are compared with those from the two-stage approach analysed and presented in chapter 2.

4.4.1 Analysis using the Bayesian approach

We used eqns (4.3) and (4.7) to form the integrated likelihood function as shown in (4.1). Potential covariates included in the models for $L_y(\boldsymbol{\theta})$ and $L_n(\boldsymbol{\beta}|\boldsymbol{\theta})$ were the factor covariates *year* (three levels: 2006, 2007, 2008), *type* (two levels: Control or Treatment plot), *state* (11 levels) and the continuous covariate *Julian day* which was centred around its mean before the analyses (for the $L_n(\boldsymbol{\beta}|\boldsymbol{\theta})$ model only as it did not reveal any influence on detection probabilities during preliminary analyses). The covariates *wind speed*, *cloud cover*, *atmospheric pressure*, *month* and *habitat* were tested in preliminary analyses but did not reveal any effect on detection probabilities or counts. Uniform priors were placed on all parameters. Lower and upper bounds for these are given in Table 4.1. To make summary statistics of parameters directly comparable to the maximum likelihood approach (see section 2.4 in chapter 2), the highest covariate levels (in numerical or alphabetic order) of detection function parameters were absorbed in the intercept to follow formatting inherent in Distance software, while the lowest levels were absorbed in the intercept for the count model

to follow formatting inherent in the *glmer* function from the *lme4* package in R.

Table 4.1: Lower and upper bounds for uniform prior distributions for all model parameters. The different *states* included GA, IA, IL, IN, KY, MO, MS, NC, SC, TN and TX.

Parameters	Lower	Upper
Detection Function		
Scale Intercept:	1	100 000
Shape:	1	20
Year levels: 2006, 2007	-3	3
Type level: Control	-2.5	2.5
State levels: GA:TN	-2.5	2.5
Counts		
Intercept:	-20	-7
Year levels: 2007, 2008	-1	1
Type level: Treatment	0	1
Julian Day:	-0.1	0.1
State level: IA:TX	-3	3
Random effect standard deviation	0	2

Preliminary investigation of the distance data indicated that the hazard-rate detection function provided a much better fit than the half-normal. Hence, we included eight different hazard-rate models as choices for the probability density function of observed distances $f(y|\boldsymbol{\theta})$ in $L_y(\boldsymbol{\theta})$ during the RJ step: one global (with no covariates) and seven multiple covariate models. For the global model, the only two parameters that required estimation were the scale and the shape parameters (see section 4.2 for details). The multiple covariate models contained additional parameters as the scale parameter was modelled as a function of one, two or three of the covariates.

For $L_n(\boldsymbol{\beta}|\boldsymbol{\theta})$, λ_{jpr} from eqn (4.6) was modelled including a fixed effect intercept and combinations of the four fixed effect covariates (16 different combinations) as well as a random effect for site.

The chain was started without any covariates for either the detection function or count model. During a single RJ step of each iteration, each of the covariates was proposed to be added or deleted depending on whether it was in the current model or not. Values u for the new parameters contained in the new model were drawn from parameter-specific proposal distributions shown in Table 4.2. These were initially defined as normal distributions with mean and standard deviation equal to the maximum likelihood estimates and standard errors from the full models. However, we adjusted means (by averaging estimates across different models for the respective parameters) and standard deviations during pilot-tuning to improve model mixing.

To move from e.g. a global hazard-rate model to a model including a covariate, the global scale parameter σ was converted into $\delta_0 \times \exp(z_1\delta_1)$ with $\sigma = \delta_0$ and $u = \delta_1$, where δ_1 is the coefficient associated with covariate z_1. The bijective function in this case (as well as in all the other possible model moves) was the identity function similar to the example shown in section 4.3.3. Therefore, the Jacobian $|J|$ (from eqn (4.11)) equalled one. We assume that all models were equally likely a priori, hence the probability of moving to model m conditional on the chain being in model m', $P(m|m')$ was equal to $P(m'|m)$ and vice versa for all possible model moves and cancelled when calculating the acceptance probability (see eqn (4.11) in section 4.3.3 page 65).

Proposal distributions for the MH step were normal where the mean was the current value of the parameter and the standard deviation was parameter-specific. The RJ

Table 4.2: Mean and standard deviation (SD) of Normal proposal distributions for parameters proposed to be added or deleted during the RJ step of the RJMCMC algorithm. All parameters were categorical, except for continuous Julian day.

Parameters	Mean	SD
Detection Function		
Year level 2006	0.11	0.10
Year level 2007	-0.15	0.10
Type level: Control	0.50	0.10
State level: GA	0.42	0.10
State level: IA	0.21	0.10
State level: IL	0.70	0.10
State level: IN	0.67	0.10
State level: KY	0.64	0.10
State level: MO	0.69	0.10
State level: MS	0.61	0.10
State level: NC	0.66	0.10
State level: SC	0.03	0.10
State level: TN	0.47	0.10
Count		
Year level: 2007	0.16	0.05
Year level: 2008	0.08	0.05
Type level: Treatment	0.42	0.10
Julian Day:	-0.01	0.01
State level: IA	0.71	0.24
State level: IL	-0.49	0.24
State level: IN	-1.16	0.23
State level: KY	-0.41	0.22
State level: MO	0.01	0.20
State level: MS	-0.38	0.22
State level: NC	-1.36	0.23
State level: SC	0.07	0.22
State level: TN	-1.05	0.23
State level: TX	1.77	0.21

and MH step together completed one iteration. A total of 100 000 iterations was carried out where the first 10 000 were considered the burn-in period and were ignored for obtaining model probabilities and summary statistics for parameters.

4.4.2 Results

For the Bayesian approach, the preferred detection function model included the covariates *year*, *type* and *state* in the model for the scale parameter of the hazard-rate

key function (probability = 1.00 to two decimal places, Table 4.3). Two other models were chosen with probabilities of < 0.001 that included two covariates (*type* and *state*) or one covariate only (*type*).

Table 4.3: Models and their probabilities resulting from RJMCMC and bootstrap analyses. Each count model included a fixed effect intercept and a random effect for site in addition to shown covariates (JD = Julian day). Model probabilities refer to the percentage of times the respective models were chosen during 90 000 iterations (after 10 000 iterations of burn-in) for RJMCMC and during 999 bootstrap iterations.

Model	RJMCMC	Two-stage
Detection Function		
MCDS: Type	< 0.001	-
MCDS: State	-	0.01
MCDS: Year + State	-	0.16
MCDS: Type + State	< 0.001	0.02
MCDS: Year + Type + State	1.00	0.81
Count		
Type + State	-	0.003
Year + Type + State	-	0.01
Type + JD + State	0.89	0.10
Year + Type + JD + State	0.11	0.89

The same model including the covariates *year*, *type* and *state* was the preferred model for the two-stage approach having been selected by AIC for 81% of bootstrap resamples. Three other models were selected: one with covariates *year* and *state* (16%), one with *type* and *state* (2% probability) and one with *state* alone (1% probability). For the count model, two models dominated the RJMCMC algorithm, the model with covariates *type*, *Julian day* and *state* as the preferred model (0.89 probability) and the full model (*year + type + Julian day + state*, 0.11 probability, Table 4.3). For the bootstrap the latter was the preferred model selected in 89% of resamples, while

the former was the second most frequently chosen model (10%). Two other models were chosen during the bootstrap including the covariates *year*, *type* and *state* (1%) and the model including covariates *type* and *state* (0.3%).

For the parameters of the detection function model, the posterior means of the parameters in the preferred model were in most cases similar to the maximum likelihood estimates resulting from the two-stage analysis on the original data (Table 4.4). The intercept for the scale parameter and the shape parameter were larger for the Bayesian approach while the coefficients for the scale parameter were on average smaller. Interestingly, measures of uncertainty were mostly smaller for the Bayesian approach despite the fact that both stages from the two-stage approach were combined in one. The posterior standard deviations were smaller than the bootstrap standard errors for all detection function parameters. 95% credible intervals were narrower than the 95% confidence intervals for all but four detection function parameters (state coefficients IN, MS, NC and TN). Intervals from the two approaches overlapped in all cases for the detection function parameters.

Table 4.4: Mean, standard deviation (SD) and 95% credible intervals (CRI) from the RJMCMC analysis along with maximum likelihood estimates (MLE), bootstrap standard errors (BSE) and 95% confidence intervals (CI) using the two-stage approach for the models with the highest probabilities (see Table 4.3 for model probabilities). Units of measurements were metres for the detection function model and square metres for the count model.

	RJMCMC			Two-stage		
	Mean	SD	95%CRI	MLE	BSE	95% CI
Detection function: fixed effects						
Scale Intercept	152.96	8.97	135.47, 170.12	138.59	16.13	112.26, 163.79
Shape	3.30	0.16	3.00, 3.63	3.01	0.27	2.68, 3.41
Scale: Year 2006	0.06	0.03	0.01, 0.12	0.10	0.06	-0.05, 0.14
Scale: Year 2007	-0.11	0.04	-0.19,-0.05	-0.15	0.05	-0.25, -0.1
Scale: Type Control	0.13	0.04	0.05, 0.20	0.15	0.05	0.05, 0.23
Scale: State GA	0.38	0.07	0.23, 0.52	0.42	0.17	0.05, 0.54
Scale: State IA	0.17	0.10	-0.01, 0.36	0.21	0.16	-0.12, 0.30
Scale: State IL	0.66	0.09	0.48, 0.85	0.70	0.17	0.35, 0.76
Scale: State IN	0.62	0.10	0.43, 0.81	0.66	0.14	0.34, 0.72
Scale: State KY	0.58	0.08	0.43, 0.74	0.64	0.12	0.35, 0.68
Scale: State MO	0.62	0.06	0.51, 0.73	0.69	0.09	0.46, 0.71
Scale: State MS	0.55	0.07	0.41, 0.70	0.61	0.10	0.37, 0.64
Scale: State NC	0.60	0.09	0.43, 0.79	0.66	0.12	0.35, 0.70
Scale: State SC	0.01	0.08	-0.14, 0.16	3E-5	0.14	-0.29, 0.12
Scale: State TN	0.44	0.10	0.25, 0.63	0.47	0.12	0.19, 0.54
Count model: random effects						
Standard deviation	0.82	0.05	0.73, 0.91	0.78	0.04	0.69, 0.81
Count model: fixed effects						
Intercept Density	-13.10	0.18	-13.43, -12.73	-13.23	0.33	-13.91,-12.87
Year 2007	-	-	-	0.17	0.13	-0.16, 0.37
Year 2008	-	-	-	0.17	0.11	-0.12, 0.31
Type Treatment	0.62	0.07	0.48, 0.75	0.63	0.12	0.36, 0.71
Julian Day	-0.01	2E-3	-0.02, -0.01	-0.01	3E-3	-0.02, -0.01
State IA	-0.81	0.29	-1.38, -0.23	-0.74	0.44	-1.65, -0.24
State IL	-0.59	0.27	-1.12, -0.06	-0.53	0.38	-1.25, -0.07
State IN	-1.24	0.27	-1.79, -0.71	-1.18	0.41	-1.99, -0.70
State KY	-0.47	0.25	-0.98, 0.03	-0.44	0.34	-1.07, -0.02
State MO	0.01	0.22	-0.44, 0.42	0.05	0.34	-0.63, 0.46
State MS	-0.43	0.25	-0.92, 0.05	-0.37	0.34	-1.04, 0.05
State NC	-1.39	0.26	-1.88, -0.87	-1.31	0.36	-1.99, -0.87
State SC	0.01	0.27	-0.53, 0.53	0.08	0.42	-0.76, 0.56
State TN	-1.10	0.28	-1.65, -0.57	-1.03	0.38	-1.80, -0.60
State TX	1.74	0.18	1.33, 1.99	1.46	0.29	0.99, 1.81

For the count model, means and intervals were again similar between the two approaches. The mean and standard deviation of the random effects standard deviation were slightly larger for the Bayesian approach (0.82, SD=0.05) compared to the two-stage approach (0.78, BSE=0.04) (Table 4.4). The fixed effect intercept was slightly larger for the Bayesian approach (-13.10, SD=0.18) compared to the two-stage approach (-13.23, BSE=0.33). However, fixed effect coefficients that were in the preferred models for both approaches were generally smaller for the Bayesian approach, except for *Julian day* where the means were equal and state coefficient TX where the mean was larger for the Bayesian approach.

Again, measures of uncertainty were mostly smaller for the Bayesian approach: standard deviations from the Bayesian approach were smaller for all fixed effect covariates in the count model compared to bootstrap standard errors. 95% credible intervals were narrower for all coefficients of the count model compared to 95% confidence intervals, except for the covariate *Julian day* where they were equal. 95% credible and confidence intervals overlapped for all count model parameters. The only covariate selected for the preferred count model for the two-stage approach that was not also in the preferred model for the Bayesian approach was *year*. 95% confidence intervals for both *year* coefficients included zero indicating that this covariate might have been negligible for the count model.

The parameter of interest in these models was the coefficient for the level Treatment of the *type* covariate in the count model. This was 0.62 (SD=0.07) and 0.63 (BSE=0.12) for the Bayesian and the two-stage approach, respectively, indicating an increase in covey densities by 86% (exp(0.62)=1.86) or 88% (exp(0.63)=1.88) by the respective methods. A positive coefficient for the level Control of the *type* covariate

in the detection function indicated that detection probabilities were slightly larger on this type of field compared to level Treatment resulting in larger values for the offset (ν_{jpr} from eqn (4.6)). However, the resulting differences in the offset between control and treatment fields did not exceed 0.23 for any year or state including both methods and could not have artificially created a positive coefficient for treatment fields in the count model as large as in the present models.

4.5 Discussion

There are two main aspects described in this chapter that are relatively innovative and deserve comparison to existing methods. We present a novel approach for combining the likelihood functions for analysing distance sampling data in section 4.2. We also present a Bayesian approach for analysing distance sampling data of multiple types in a straightforward manner.

Bayesian methods have been used before for analysing line transect data with a global half-normal detection function (e.g. Eguchi and Gerrodette, 2009; Gimenez et al., 2009; Zhang, 2011) or a half-normal with covariates (Moore and Barlow, 2011). We used the hazard-rate detection function and included model selection between different covariate combinations for the scale parameter.

In section 4.2, we summarised the equations needed to compose this likelihood for line transects, point transects, exact and interval distance data and for including covariates in the scale parameter model (as described by Marques and Buckland, 2003). Different key functions, e.g. the half-normal, hazard-rate or others, may easily be implemented. It may also be extended to include adjustment terms (added to the

half-normal or hazard-rate model, Buckland et al. (2001)) or covariates in the shape parameter.

The log-linear Poisson model for densities described in section 4.2 may be used when information for covariates is available. It does not depend on a random survey design in contrast with the conventional distance sampling approach and data arising from surveys conducted from platforms of opportunity may be used (Hedley and Buckland, 2004). In addition, it allows identification of relationships between abundance or density and parameters of interest, such as the *type* covariate in our case study. It is different from the approaches described by Hedley and Buckland (2004) or Buckland et al. (2009) in that these authors analyse their data in two stages. In their second stage count model, they condition on the estimate of the effective area which is derived from the first stage detection function model. Our integrated likelihood approach estimates all parameters simultaneously allowing to quantify the precision of the parameters in the count model while taking proper account of the estimation of detection function parameters.

Using the Poisson model including a random effect for estimating densities as defined in eqn (4.6) also allows us to accommodate correlated measurements due to closeness in space and/or time, for example as occurs when there are repeat counts at the same line or point. This is different from the integrated likelihood described by Royle et al. (2004). These authors considered the true but unknown abundances at the site as a random effect with a Poisson distribution (in their notation $N_i \sim Poisson(\lambda_i)$) and integrated it out. Hence, they derived a Poisson likelihood for the observed counts with expected value equal to $\lambda_i \pi_k(\theta)$, where $\pi_k(\theta)$ describes the probability that an animal occurs and is detected in the kth distance band. In contrast, we

consider variations in the observed counts between the different sites as a normally distributed random effect with mean zero, hence accounting for correlations between measurements at the same sites. We used this approach to obtain unbiased estimates of coefficients retained in the count model which may then be used to predict total abundance at the line or point. In addition, the approach of Royle et al. is limited to interval distance data.

Similar to Hedley and Buckland (2004) and Buckland et al. (2009), our likelihood may be extended to include smooth functions for continuous covariates, e.g. by fitting regression splines using the B-spline basis, or the Poisson likelihood may be replaced with a negative binomial likelihood if more appropriate, e.g. in case overdispersion in the count data is present. In cases where no covariates are available and the survey followed a random design, our covariate model for $L_n(\boldsymbol{\beta}|\boldsymbol{\theta})$ in eqn (4.1) may be replaced with a binomial likelihood to estimate abundance in the covered region or the survey area (Buckland et al., 2004, eqns (2.33) and (2.34), respectively). For the latter, the average inclusion probability is defined as the product $P_c \hat{P}_a$, where P_c is the probability that an animal is covered (i.e. within truncation distance w, known from the survey design) and $\hat{P}_a$ is the average detection probability in the covered region. This approach allows direct estimation of total abundance in the study area. To estimate total abundance in the study area using our approach requires making predictions of abundance in the entire area including those subareas that were not covered during the survey. This, in turn, requires knowledge of covariates retained in the preferred model(s) for these areas. While the binomial model only allows estimation of total abundance in the study area, our approach allows us to estimate local abundances on a smaller scale within the study area.

Alternatives to both our Poisson model and the binomial model of Buckland et al. was presented by Johnson et al. (2010) and Niemi and Fernández (2010). These authors modelled distance sampling data as a thinned spatial point process. While Niemi and Fernández (2010) conditioned on a 'known' detection function, Johnson et al. (2010) estimated parameters of the detection model and the intensity parameters simultaneously. While the underlying theory may differ, the Poisson model from our integrated likelihood and the spatial point process model of Johnson et al. have similar assumptions, i.e. perfect detection on the line (point), animals follow certain distribution patterns which may be captured by measurable covariates, the observation process is a snapshot, and animals distribute themselves independently from each other and from the line (point). However, the approach of Johnson et al. does not require defining a truncation distance for each line which might make it favourable for study areas with complex boundaries, e.g. narrow bays. However, while our approach requires defining a truncation distance, it may vary between different transects or transect segments.

The comparison of summary statistics for model parameters from the Bayesian approach with parameter estimates from the two-stage approach revealed some differences in means and point estimates (Table 4.4) which cannot be due to prior sensitivity as we used uniform priors on all parameters for the Bayesian approach. We assume these differences may have been due to the fact that - as opposed to the two-stage approach - the likelihoods for both components of our model are combined for the integrated likelihood and influence each other. We argue, in concurrence with Johnson et al. (2010), that simultaneous estimation of all parameters in one stage represents a more realistic model without having to rely on the assumption of a true detection

function model.

Model uncertainty for the detection model might have been captured better with the two-stage approach where - besides the preferred model - three other models were selected during the bootstrap with probabilities of up to 0.16 as opposed to two others with probabilities of $< 0.1\%$. However, for the count model, results from the two-stage approach were slightly ambiguous in that the analysis of the original data (and the bootstrap) favoured a model that included the covariate *year* (in addition to covariates *type*, *Julian day* and *state*) where coefficients for both levels were significant at the 0.001% level. However, 95% confidence intervals for both *year* coefficients resulting from the bootstrap overlapped zero, suggesting that the *year* effect may not be different from zero (Table 4.4). This ambiguity was not present in the results from the Bayesian approach as the preferred count model did not include *year*.

Overall, our Bayesian approach delivered valid results. Besides the often stated benefits for Bayesian analyses, e.g. allowing for prior information to be included, it provided a particular benefit for using the integrated likelihood defined in section 4.2: it might be challenging in some cases, such as our case study, to find the maximum likelihood estimates for all parameters in one step. The covey data included a total of 2545 observed distances during 2534 counts and the full model included 31 parameters with a random effect (447 sites). Using maximum likelihood methods, the random effect is integrated out. However, due to the integrated nature of the detection and count models, functions such as *glmer* from the *lme4* package in R may not be used as these treat the offset as a constant. Using the hierarchical model set up for the Bayesian approach where the random effect coefficients are included in the model specification and updated during each iteration, offers a straightforward technique to

explore the parameter space. In comparison to the RJMCMC algorithm, an equivalent model selection routine using a maximum likelihood approach that considered all possible model combinations would have required maximising $L_{y,n}(\boldsymbol{\beta}, \boldsymbol{\theta})$ for 128 models (possible combinations of eight detection functions and 16 count models). Hence, the RJMCMC algorithm provided a very efficient option for incorporating model uncertainty.

Chapter 5

Using hierarchical centering to facilitate a reversible jump MCMC algorithm for random effects models

5.1 Introduction

For Bayesian analyses, for a given model, the posterior distribution of the parameters is formed by combining the likelihood of the data with the prior distributions of the parameters. An MCMC algorithm is often used to sample from this posterior distribution to obtain inference on the parameters of interest. In the presence of model uncertainty, the posterior distribution can be extended to be defined jointly over both parameter and model space. This posterior distribution is often explored using the

reversible jump Markov chain Monte Carlo (RJMCMC) algorithm (Green, 1995). However, the art of setting up an RJMCMC algorithm can be challenging on multiple levels. The objective is generally to construct a chain that moves freely between models, efficiently exploring model and parameter space simultaneously. The RJMCMC algorithm entails iteratively updating the parameters conditional on the model (i.e. within-model move) and then updating the model (and corresponding model parameters) conditional on the current parameters (i.e. between-model move). See chapter 4 for details.

Mixing problems for the within-model moves are often due to high autocorrelation within the constructed Markov chain. Improvements for mixing within a given model have been investigated in the framework of MCMC with the aim of reducing posterior correlations and increasing the effective sample size by reparameterisation. In this context, Browne (2004) and Browne et al. (2009) showed that hierarchical centering (first described by Gelfand et al., 1995) can significantly reduce the autocorrelation within the MCMC algorithm. The use of hierarchical centering in the presence of random effects refers to exchanging the zero mean of the random effect component, typically assumed to be of normal form, with a model consisting of an intercept and one or more fixed effect covariates. This will be described in detail in section 5.2. Papaspiliopoulos et al. (2007) investigated the circumstances when hierarchical centering performs well in comparison to noncentering.

Other methods for improving mixing of an MCMC algorithm include parameter expansion, which refers to augmenting the model with additional parameters to form an expanded model (Browne, 2004). The original model is embedded in the expanded one and parameters from the original model can be constructed with parameters from

the expanded model. Vines et al. (1995) describe a method of reparameterisation for random effects models called *sweeping* which is suitable also for models with multiple sets of random effects in a GLMM framework. The idea consists of adding the mean of the random effects coefficients to the intercept of the fixed effects while subtracting the same quantity from each random effect coefficient.

For the between-model move in an RJMCMC algorithm (the RJ step), the current model is updated by proposing to move to an alternative model (with given parameters) and accepting this move with some probability. Mixing problems for these between-model moves may arise for multiple reasons, e.g. due to difficulties in finding proposal distributions and updating procedures that produce suitable acceptance probabilities. Besides careful pilot-tuning of proposal distributions, several methods for improving the acceptance rate for the reversible jump step have been proposed. For example, Green and Mira (2001) proposed delayed rejection, where after initial rejection a second attempt to jump is made with samples generated from a new distribution that may depend on the rejected proposal. Brooks et al. (2003) assumed a family for the proposal distribution, where the proposal parameters are chosen to maximise (in some form) the acceptance probability. Al-Awadhi et al. (2004) demonstrated that increasing acceptance probabilities may be achieved by using a secondary Markov chain with a fixed number of steps that serves to move the value of an RJMCMC proposal closer to a mode before calculating the acceptance probability for the proposed move. Papathomas et al. (2011) proposed that model mixing for generalised linear models may be improved by using proposal densities that draw samples from parameter subspaces of competing models. Forster et al. (2012) used the Laplace approximation to integrate out the random effects and orthogonal projections of the

current linear predictor onto the proposed linear predictor to produce effective proposals for generalised linear mixed models.

While these previous approaches describe strategies to improve the acceptance rate for RJ steps in general, they can be quite complex to implement. We propose an approach using hierarchical centering that is relatively straightforward to implement for random/mixed effects models. A particular problem that one may encounter with random effects models is that the random effects coefficients may begin absorbing the effect of one or more fixed effect covariates if these are not present in the model at times during the Markov chain. The inclusion of such effects into the model may then be unlikely as they are already accounted for within the random effects. We will demonstrate below that using hierarchical centering provides a simple way of reparameterising the model that will prevent this problem and improve the between-model mixing.

Hierarchical centering was initially described by Gelfand et al. (1995) as a method to improve convergence for mixed models using MCMC methods. Here we extend the ideas to improve mixing in an RJMCMC algorithm. In particular, we consider the case for a log-linear Poisson model with fixed effects and a normally distributed random effect, where the overall likelihood combines the Poisson likelihood for each observation and the normal density for each random effect coefficient. We demonstrate how the Poisson likelihoods and the normal densities are affected differently during a proposal to add a covariate for a regular RJMCMC algorithm and one including hierarchical centering.

We demonstrate the improved model mixing using our case study 1, point transects of indigo buntings. This is the same data as described in chapter 2 and analysed with

the two-stage and the integrated approach in chapters 2 and 3. In brief, to study the effect of establishing conservation buffers along margins of agricultural fields on density of several species of conservation interest, pairs of points were set up at the edge of fields in a number of states in the USA. These pairs of points consisted of one point on a treatment field and one on a nearby control field without a buffer and will be referred to as sites in the following. We use a combined likelihood including the likelihoods for the detection function and the log-linear Poisson model where counts are adjusted for imperfect detection within the search area around the point (as described in chapter 4). A random effect for site is included in the Poisson model to accommodate correlated counts between different sites.

In the following we begin by describing how to implement hierarchical centering for RJMCMC, describe the effects on the dynamics of the algorithm, and present updating methods for the RJ step using hierarchical centering (section 5.2). We then apply the methods to our case study (section 5.3) and discuss our findings (section 5.4).

5.2   Hierarchical centering

The hierarchical centering described in this chapter refers to mixed effect models where a normal distribution is assumed for the random effect. Other distributions may be assumed for the random effect (e.g. Komárek and Lesaffre, 2008) to which these methods can be applied but we focus on the normal distribution for simplicity. We describe the case for a generalised linear mixed model with a Poisson error structure. Here, the expected value λ is modelled via a log-link function with a common intercept, β_0 and random effect coefficients b_j for group j are included for which normality is assumed.

For a mixed effect model without hierarchical centering, the random effect is incorporated into the model under the assumption of a global zero-mean and unknown standard deviation, σ_b, i.e. $b_j \sim N\left(0, \sigma_b^2\right)$ (e.g. Bates, 2009b). Let us assume we have a set of K covariates, x_k (and associated coefficients, β_k) that can be incorporated as fixed effects. The expected value for the full model including all covariates may then be expressed as:

$$\lambda_j = \exp\left(\beta_0 + \sum_{k=1}^{K} x_k \beta_k + b_j\right), b_j \sim N\left(\mu = 0, \sigma_b^2\right). \tag{5.1}$$

Different models correspond to the combinations of covariates present in the model (i.e. non-zero β_k values). During a between-model move of an RJMCMC algorithm using this scenario, the proposal to delete or add one (or more) of the covariates alters the formula for λ while the distribution for the random effects terms b_j remain the same (see chapter 4 for details on the RJ step).

In hierarchical centering, the mean of the random effect is modelled using a combination of the intercept and one or more covariates that are "pulled from" the λ model from eqn (5.1) (Gelfand et al., 1995). In the case that the intercept and covariate x_1 are used for centering, the full model from eqn (5.1) becomes:

$$\lambda_j = \exp\left(\sum_{k=2}^{K} x_k \beta_k + b_j\right), b_j \sim N\left(\mu = \beta_0 + x_1 \beta_1, \sigma_b^2\right). \tag{5.2}$$

The proposal to delete or add x_1 from the model involves altering the distribution for b_j, while the proposal to delete or add any other covariates remains the same as before (altering the formula for λ).

In the case that all k covariates are included in the centering, the full model from eqn

(5.1) becomes:

$$\lambda_j = \exp\left(b_j\right), b_j \sim N\left(\mu = \beta_0 + \sum_{k=1}^{K} x_k \beta_k, \sigma_b^2\right). \tag{5.3}$$

In this scenario, the formula for λ remains unchanged during the proposals to delete or add any of the covariates, while the distribution for b_j changes for each proposed model move.

These methods may also be extended to estimate the random effect standard deviation as a function of covariates, for example if different levels of variability are present for different levels of a factor covariate. We refrain from this to keep the example relatively simple. However, it is essential that only those covariates are included in the centering (i.e. x_1 in eqn (5.2) or x_k with $k=1,...,K$ in eqn (5.3)) that have consistent values for all observations within a group (Browne et al., 2009). We refer to a group in terms of the grouping unit for the random effect, i.e. all observations belonging to the same group j are modelled with the same random effects coefficient b_j in the equations given above. The grouping for the random effect should generally occur to account for intra-group dependence (Davison, 2003). In the case, for example, where a study consists of a large number of sites and repeat measurements were taken at the same sites, we expect the measurements from the same sites to be correlated.

However, for hierarchical centering to be applicable, the grouping for the random effect needs to be so that for any of the covariates included in the centering, values for the respective covariate are the same within a group. If, for example, the grouping unit for a study is *site*, then the covariate *state* (*state* as in the geographical governed entity) can be included in the centering as each site only belongs to one state. Hence, all observations for a given site belong to the same state regardless of how many

times the measurements were repeated. Conversely, *Julian day* could not be included as different values are possible for different observations within a given site due to repeating the measurements on different days. As long as this condition holds, any combination of covariates may be included.

Hierarchical centering relies on the fact that the random effect coefficients pick up the effect of the covariates included in the centering (given that they have an effect) as they are updated during the within-model move of each iteration of the RJM-CMC. Running separate MCMC algorithms (without between-model moves) on the full models from eqns (5.1), (5.2) or (5.3) would probably result in the same posterior distribution and nearly identical summary statistics for the covariates if the chain was run long enough - although mixing might be different for these different parameterisations. However, when including the between-model moves in an RJM-CMC algorithm, mixing problems may become more severe, potentially leading to the wrong conclusions. Here, results may be different depending on which scenario was used. If, e.g. under the scenario of eqn (5.1), the random effect coefficients absorb the effect of covariate x_1, the chain will likely get stuck in models that do not include x_1. For the scenarios of eqns (5.2) and (5.3), moves to models including covariate x_1 would be favoured if the random effect coefficients absorbed the effect of x_1 as then the coefficients will be closer to their modelled means. We will show below, that this is due to the fact that here different parts of the likelihood are affected by a proposed model move compared to eqn (5.1).

5.2.1 Effects of hierarchical centering on RJMCMC dynamics

Using either one of the models for λ from above (eqns (5.1), (5.2), or (5.3)), the likelihood of the log-linear Poisson model, $L_n(\boldsymbol{\beta}, \sigma_b)$, with a normally distributed random effect may be formulated as:

$$L_n(\boldsymbol{\beta}, \sigma_b) = \prod_{j=1}^{J} \left(\prod_{r=1}^{R_j} \frac{(\lambda_{jr})^{n_{jr}} \exp\left(-\lambda_{jr}\right)}{n_{jr}!} \right) \times \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left(-\frac{(b_j - \mu)^2}{2\sigma_j^2}\right), \qquad (5.4)$$

where vector $\boldsymbol{\beta}$ contains the coefficients for covariates included in the models and n_{jr} are the observed measurements. The indices $j = 1, 2, 3, ..., J$ represent the groups for the random effects and $r = 1, 2, 3, ..., R_j$ indices the different measurements taken for the jth group. Hence, for each of J groups of observations the probability of observing n_{jr} under the log-linear Poisson model with expected value of λ_{jr} is multiplied for all observations within that group, which is then multiplied by the normal density of the random effect coefficient b_j. The only coefficients that influence both parts of this likelihood, i.e. the Poisson likelihood for the observations and the normal densities, are the random effect coefficients, regardless of which scenario is used from the previous section.

Using a random walk single-update Metropolis-Hastings algorithm for the within-model move, the probability of accepting the newly proposed state for, say parameter β_1 at time t is calculated with:

$$\alpha(\beta_1'|\beta_1^t) = min\left(1, \frac{L_n(\beta_1', \boldsymbol{\beta}_{-1}^t, \sigma_b^t)p(\beta_1')q(\beta_1^t|\beta_1')}{L_n(\beta_1^t, \boldsymbol{\beta}_{-1}^t, \sigma_b^t)p(\beta_1^t)q(\beta_1'|\beta_1^t)}\right), \qquad (5.5)$$

where $L_n(\beta_1', \boldsymbol{\beta}_{-1}^t, \sigma_b^t)$ is the probability density function of the data conditional on the newly proposed parameter value β_1' (Brooks and Gelman, 1998; King et al., 2010). For our example it is calculated using eqn (5.4). The prior distribution is denoted by $p()$ and $q(\beta_1'|\beta_1^t)$ is the proposal density of β_1' given the current state is β_1^t. From eqn (5.5) we can deduce that on average, $L_n(\boldsymbol{\beta}^t, \sigma_b^t)$ will improve with each iteration in particular if the posterior is not too sensitive to prior probabilities and if symmetrical proposal distributions are used (and $q(\beta_1^t|\beta_1')$ and $q(\beta_1'|\beta_1^t)$ cancel in eqn (5.5)). Furthermore, combining what we know from eqns (5.1) to (5.4), it is evident that the Poisson likelihoods within $L_n(\boldsymbol{\beta}^t, \sigma_b^t)$ will improve if the variation that is not accounted for by the fixed effect coefficients is picked up by the random effect coefficients. On the other hand, the normal densities will return higher values for random effect coefficients close to their mean values.

For the RJ step, the dynamics are more complex. Here, the calculation of the acceptance probability to move to another model also includes the proposal densities for the newly proposed parameter values, a priori model probabilities and the Jacobian (which is derived from the bijective function). For details see eqn (4.11) in chapter 4 page 65.

Hence, for any RJMCMC algorithm, the MH step will on average improve the likelihood of the current model with each iteration as long as the chain remains in the same model. Current parameter values including those of the random effect coefficients will be adjusted in such a manner that they, on average, produce λ_{jr} that return higher likelihood values for n_{jr} under the current model.

Intuitively, one may think that without hierarchical centering a problem arises for a between-model move (using models from eqn (5.1)) when a covariate, say x_1, may

have an effect but is not included in the current model. Then, the random effect coefficients may begin to absorb this effect and, in this manner, adjust the value for λ_{jr} to improve the likelihood. This may result in a "tug-of-war" between the Poisson likelihood trying to adjust the coefficients in such a manner that the effect of x_1 is accounted for and, on the other hand, the normal densities trying to keep the coefficients close to zero. This will typically also result in an inflated random effect standard deviation since the random effects coefficients are replacing some unexplained variability attributable to x_1. If this has indeed occurred, an acceptance of x_1 into the model during a between-model move step may become very unlikely as its effect is already accounted for by the random effect coefficients. Hence, during a proposal to add x_1, the new model with x_1 will create inferior λ_{jr}. These will then return decreased likelihood values even if the randomly drawn value(s) for x_1 would produce a larger likelihood under circumstances before the effect has been absorbed by the random effect coefficients.

This issue may be addressed using hierarchical centering since proposing to add x_1 into the model will not change λ_{jr} (and the Poisson likelihood). Here, the random effects coefficients absorb the effects of the covariates included in the model within the mean of the random effect distribution (in addition to the intercept β_0). Using eqn (5.2) this would be covariate x_1 (as in our example), using eqn (5.3) this would be covariates x_k with $k = 1, 2, 3..., K$. The only part of the likelihood that is affected when updating this/these covariate(s) (for within-model and between-model moves) are the normal densities from eqn (5.4). It is likely that, on average, the normal densities improve for the individual random effects coefficients as these will on average be closer to their assumed mean. As λ_{jr} remains the same, likelihood values returned

by the Poisson part of eqn (5.4) remain the same.

5.2.2 RJ updating methods using hierarchical centering

To demonstrate how to implement hierarchical centering, we continue with our simple example from the previous section of including covariate x_1 into an intercept-only model, say model m. Suppose that at a given iteration the current state of the chain is model m, where $\lambda = \exp(b_j)$ with $b_j \sim N(\mu = \beta_0, \sigma_b^2)$ from eqn (5.3) (although if x_1 is the only covariate available, $K = 1$ and eqns (5.2) and (5.3) are equivalent). We then propose to move to model m' by adding covariate x_1. Hence, model m' is defined as $\lambda' = \exp\left(b_j'\right)$ with $b_j' \sim N\left(\mu' = \beta_0' + x_1\beta_1', \sigma_b'^2\right)$. Let us assume that covariate x_1 is a categorical covariate with L levels (individual levels are denoted with l) and that all measurements within a group j belong to the same level of x_1. In the following, we describe two different ways for implementing the RJ step. The difference between them lies in the definition of the proposal distributions for the new parameters for the between-model move, and, hence, should only have an influence on the acceptance probability of this move. It should not have an influence on summary statistics of the parameters in the final model given that the chain had an adequate burn-in.

5.2.2.1 Hierarchical centering using predefined proposal distributions

For this method, we define proposal distributions for the intercept and the L levels of covariate x_1 before starting the chain (as for the other parameters in the model). If, for example, normal proposal densities are used, we define the proposal density for the intercept β_0 as $\beta_0 \sim N(\mu_0, \sigma_0^2)$ and for coefficient β_{1l}' as $\beta_{1l}' \sim N\left(\mu_{1l}', \sigma_{1l}'^2\right)$.

The proposal to add x_1 to model m during the RJ step involves drawing random samples from these proposal distributions that are used as coefficients in the model for the random effect means μ, and calculating the acceptance probability for this move (based on eqn (4.11) in chapter 4).

5.2.2.2 Hierarchical centering using updated proposal distributions

Here, the proposal distributions of both the global random effects mean, $\mu = \beta_0$ of model m and the covariate-specific means, $\mu' = \beta_0' + x_1\beta_1'$ of model m' are updated before the RJ step during each iteration of the RJMCMC algorithm. The covariate-specific means for our example with covariate x_1 with L levels can also be expressed as $\mu_l' = \beta_0' + x_1\beta_{1l}'$. To update $\mu = \beta_0$ at iteration $t+1$, we take the overall mean of the current values of all random effect coefficients (i.e. $\mu^{t+1} = \bar{b}_j^t$). To update $\mu_l' = \beta_0' + x_1\beta_{1l}'$, we take the L means of the coefficients belonging to the respective levels of covariate x_1, i.e. $\mu_l'^{t+1} = \bar{b}_{jl}'^t$, where the $b_{jl}'^t$ are those coefficients belonging to the lth level of x_1 at iteration t.

5.3   Case study: point transects of indigo buntings

5.3.1   Data and Methods

Here we analysed the point transect data of indigo buntings from chapter 3 again. This data is described in detail in section 2.3.1. As the models from eqns (5.1) to (5.3) assume perfect detection on the plot, we needed to supplement these with a model to adjust counts for imperfect detection. Thus, we fitted a detection function to the observed distances of individual detections which was then used to estimate

the effective area, ν. The effective area was incorporated into the log-linear Poisson model for λ_{jr} as an offset as described in detail in chapter 4. With the offset included, the full model without hierarchical centering from eqn (5.1) becomes:

$$\lambda_{jr} = \exp\left(\beta_0 + \sum_{k=1}^{K} x_k\beta_k + b_j + \ln(\nu)\right), b_j \sim N\left(\mu = 0, \sigma_b^2\right). \qquad (5.6)$$

Site was used as the grouping factor for the random effect. Available covariates were *state* with nine levels (x_1), *year* (2006, 2007, x_2), *Julian day* (x_3) and the parameter of interest, covariate *type* with two levels: control or treatment plot (x_4). We refrained from modelling heterogeneity in detection probabilities to keep this example relatively simple. Hence, the effective area ν from eqn (5.6) was assumed to be the same for all observed counts.

As *state* was the only covariate that had consistent values for all measurements within a given site, we were limited to using only one covariate within the hierarchical centering. With hierarchical centering using the *state* covariate, x_1, the full model from eqn (5.2) becomes:

$$\lambda_{jr} = \exp\left(\sum_{k=2}^{K} x_k\beta_k + b_j + \ln(\nu)\right), b_j \sim N\left(\mu = \beta_0 + x_1\beta_1, \sigma_b^2\right). \qquad (5.7)$$

To estimate parameters of both the detection function and the count model in one step, we combined the likelihood components pertaining to the respective models using the integrated likelihood, $L_{n,y}(\boldsymbol{\beta}, \sigma_b, \boldsymbol{\theta}) = L_{yG}(\boldsymbol{\theta})L_n(\boldsymbol{\beta}, \sigma_b|\boldsymbol{\theta})$ described in chapter 4. $L_n(\boldsymbol{\beta}, \sigma_b|\boldsymbol{\theta})$ is equivalent to eqn (5.4); however, it becomes conditional on detection function parameters $\boldsymbol{\theta}$ when including the effective area as an offset in eqns (5.6) or (5.7). The maximum number of visits to a site R_j ranged from 4-16 between sites as

each site consisted of two points and was visited $1 - 4$ times in each of the two survey years.

As distances were recorded in intervals (rather than exact distances), the likelihood for the detection function component, $L_{yG}(\boldsymbol{\theta})$ was defined as the multinomial likelihood given in eqn (4.5) of chapter 4. As before, we only analysed data from the three innermost distance intervals (see chapter 2 for details). For the detection function models, we considered the half-normal and hazard-rate key functions as the two (non-nested) model options (Buckland et al., 2001). For the count model, we considered all possible combinations of the covariates *year*, *type*, *Julian day* and *state*. We ran two different analyses on the same data: firstly, regular RJMCMC with a global zero-mean random effect (as shown in eqn (5.6)) which we refer to as the global zero-mean analysis (GZM).

To implement hierarchical centering, we pulled the intercept β_0 and covariate *state* from the λ_{jpr} model to include them in the model for the random effect mean (as shown in eqn (5.7)). This analysis will be referred to as HC in the following. We used predefined proposal distributions for all parameters. These were the same for both analyses (see Table 5.1). A priori model probabilities were considered equal and the identity function used for the bijective function; hence, the Jacobian for calculating the acceptance probability for the between-model move given in eqn (4.11) from chapter 4 equalled one. For both analyses, we placed the same set of uniform priors on the parameters (Table 5.1).

For each analysis, the chain was started with the most parsimonious models: the half-normal detection function and a density model containing the fixed effect intercept and a random effect for site. However, we ran additional chains for each type

of analysis (GZM and HC) that included *state* in the initial density model which we refer to as GZM-state and HC-state in the following. We ran 200 000 iterations for each analysis, the first 20 000 were considered as the burn-in phase. The effective sample size was calculated for each parameter using the function *effectiveSize* from the R package *coda*.

Table 5.1: Means and standard deviations (SD) of normal proposal distributions for model parameters as well as their lower and upper boundaries for uniform prior distributions. HN and HR refer to the half-normal and the hazard-rate detection functions respectively.

Parameters	Proposal Distributions		Uniform Priors	
	Mean	SD	Lower	Upper
Detection Function				
Scale HN:	37	2	10	99
Scale HR:	28	2	10	99
Shape HR:	2	1	1	10
Density				
Random effect SD	-	-	0	1
Intercept:	-	-	-20	-7
Year level: 2007	0.05	0.2	-1	1
Type level: Treated	0.3	0.1	0	1
Julian Day:	0.0055	0.003	-0.1	0.1
State level: IL	0.4	0.5	-2.5	2.5
State level: IN	0.3	0.5	-2.5	2.5
State level: KY	0.7	0.5	-2.5	2.5
State level: MO	0	0.5	-2.5	2.5
State level: MS	0.5	0.5	-2.5	2.5
State level: OH	0	0.5	-2.5	2.5
State level: SC	0.2	0.5	-2.5	2.5
State level: TN	0.8	0.5	-2.5	2.5

5.3.2 Results

The preferred detection model was the hazard-rate function with posterior probability of 1.00 for all analyses (Table 5.2). Probabilities for the density models differed between the methods. When the chain was initialised without *state* in the model, the preferred model for density from the GZM analysis included the covariates *type* and *Julian day* with probability 0.85. The alternative model included the additional covariate *year* and was selected during the remaining 15% of the iterations. The covariate *state* was never included in any of the models for this analysis.

By contrast, all other analyses, including GZM-state, HC and HC-state, included

Table 5.2: Posterior model probabilities for the analyses of the indigo bunting data. GZM and the HC analyses did not include *state* in the initial model. GZM-state and HC-state did include *state* in the initial model.

Analysis	GZM	GZM-state	HC	HC-state
Detection Function Model				
CDS: Hazard-rate key	1.000	1.000	1.000	1.000
Density Model				
Type + JD	0.851	–	–	–
Year + Type + JD	0.149	–	–	–
Type + State	–	0.735	–	0.066
Year + Type + State	–	0.004	–	0.004
Type + JD + State	–	0.259	0.946	0.876
Year + Type + JD + State	–	0.002	0.054	0.054

state with probability = 1.00. The preferred model for GZM-state included *type* and *state* (0.74 probability) (Table 5.2). The second most preferred mode included the additional covariate *Julian day* (0.26 probability). Two more models were selected

including *type*, *year* and *state* (< 0.01 probability) and *year*, *type*, *year* and *state* (< 0.01 probability).

The preferred model for the HC analysis included *type*, *Julian day* and *state* (0.95 probability) and the second most preferred model included *type*, *Julian day*, *year* and *state* (0.05 probability) (Table 5.2). Similarly, the preferred model for HC-state included *type*, *Julian day* and *state* (0.88 probability). Three other models were selected including *type* and *state* (0.07 probability), *year*, *type*, *Julidan day* and *state* (0.05 probability) and *year*, *type* and *state* (< 0.01 probability).

When comparing the probabilities of being in the model for different covariates between analyses, we found the largest discrepancies for covariates *state* and *Julian day* (Table 5.2). The probability of *state* being in the model was 0.00 for GZM and 1.00 for all other analyses. This discrepancy caused us to believe that the chain for the GZM analysis had not converged. For covariate *Julian day*, the probability of being in the model was 0.26 for the GZM-state analysis, while it was 0.93 for HC-state and 1.00 for the remaining two. Again, this caused us to believe that the chain of the GZM-state analysis had not converged.

Summary statistics for the parameters of the preferred models resulting from the four analyses are given in Table 5.3. Means and 95% credible intervals (CRI) were nearly identical between analyses for detection function parameters of the hazard-rate detection function.

Table 5.3: Mean and 95% credible intervals for models with highest posterior support from the respective analyses. *State* level GA is absorbed in the intercept.

Analysis	GZM	GZM-state	HC	HC-state
Detection Function Parameters				
Scale σ	28.20	28.16	28.05	28.24
	(25.03,31.25)	(24.91,31.21)	(25.00,31.04)	(25.00,31.04)
Shape τ	2.08	2.08	2.08	2.09
	(1.92,2.26)	(1.91,2.25)	(1.92,2.25)	(1.92,2.26)
Density: Random Effects Parameters				
Standard deviation σ_b	0.77	0.58	0.51	0.51
	(0.65,0.91)	(0.49,0.68)	(0.45,0.57)	(0.45,0.57)
Density: Fixed Effect Parameters				
Intercept β_0	-10.62	-9.85	-10.44	-10.63
	(-11.21,-10.13)	(-10.12,-9.59)	(-10.97,-10.01)	(-11.22,-10.01)
Type level: Treated β_4	0.31	0.31	0.30	0.31
	(0.24,0.37)	(0.24,0.37)	(0.24,0.37)	(0.24,0.37)
Julian Day β_3	0.008	–	0.004	0.005
	(0.006,0.012)	–	(0.002,0.007)	(0.002,0.009)
State level: IL $\beta_{1_{IL}}$	-	1.04	0.97	0.95
	-	(0.70)	(0.63,1.32)	(0.60,1.31)
State level: IN $\beta_{1_{IN}}$	-	0.85	0.79	0.77
	-	(0.49,1.16)	(0.45,1.14)	(0.42,1.12)
State level: KY $\beta_{1_{KY}}$	-	1.35	1.24	1.21
	-	(1.03,1.66)	(0.90,1.57)	(0.87,1.55)
State level: MO $\beta_{1_{MO}}$	-	0.43	0.35	0.33
	-	(0.15,0.72)	(0.04,0.67)	(0.01,0.66)
State level: MS $\beta_{1_{MS}}$	-	1.22	0.97	0.96
	-	(0.89,1.56)	(0.64,1.31)	(0.61,1.30)
State level: OH $\beta_{1_{OH}}$	-	0.56	0.39	0.36
	-	(0.25,0.86)	(0.07,0.72)	(0.03,0.69)
State level: SC $\beta_{1_{SC}}$	-	0.78	0.68	0.66
	-	(0.45,1.12)	(0.32,1.04)	(0.30,1.02)
State level: TN $\beta_{1_{TN}}$	-	1.45	1.38	1.35
	-	(1.13,1.78)	(1.04,1.72)	(1.01,1.70)

By contrast, summary statistics of the random effects standard deviation of the density model were very different between the analyses. In particular, the mean of this parameter was larger for the GZM analysis and CRIs did not overlap those from either analysis involving hierarchical centering. For the GZM-state analysis, the mean and CRIs for this parameter were closer to those from the HC and HC-state analyses than to the GZM analysis. Although surprisingly similar, the intercept of the density model cannot be compared between the GZM and the remaining three analyses as it represents a global mean for GZM and the mean for state GA only for the others. The mean for this parameter from GZM-state was larger which was likely due to the fact that here the preferred model did not include *Julian day*. For the *type* covariate, mean and CRIs were almost identical. For covariate *Julian day*, the mean and limits for CRIs were larger for GZM compared to HC or HC-state. For the coefficients of the different *state* levels, means were all greater than zero, indicating that the state absorbed in the intercept, GA, had the lowest bird densities.

The effective sample sizes for detection function parameters and for the intercept and *type* covariate of the density model were similar for all four analyses (Table 5.4). However, for the GZM-state analysis, the effective sample sizes for the *state* coefficients were consistently larger than those from HC or HC-state. On the other hand, effective sample sizes for the random effects standard deviation were consistently larger for analysis with hierarchical centering than for those without.

Table 5.4: Effective sample sizes for model parameters from the four RJMCMC analyses.

Parameter	GZM	GZM-state	HC	HC-state
Detection function				
Scale	856	916	853	741
Shape	1045	1160	997	934
Density: random effects				
Standard deviation	249	324	1408	1259
Density: fixed effects				
Intercept Density	78	243	184	82
Type Treatment	9516	8846	11152	9484
Julian Day	83	–	47	22
State IL	–	540	113	49
State IN	–	483	120	52
State KY	–	368	76	32
State MO	–	447	51	22
State MS	–	497	102	45
State OH	–	430	71	31
State SC	–	559	122	55
State TN	–	500	94	43

5.4 Discussion

The purpose of incorporating random effects in count models is generally to model variation that is otherwise unaccounted for. When using RJMCMC methods, the danger exists that the random effect coefficients account for too much of the variation and prevent the inclusion of a fixed effect covariate into the model. We demonstrated this case with our GZM analysis that was initiated without *state* in the model. Due

to lack of convergence, the covariate *state* was never selected and the resulting random-dom effect standard deviation was much larger compared to the GZM-state, HC or HC-state analyses. The HC analysis was also initiated without *state* in the model but revealed posterior probabilities of *state* being in the model of 1.00. However, for both analyses that were initiated without *state* in the model, GZM and HC (as well as for HC-state), the random effect coefficients absorbed the effect of the *state* covariate. For GZM, this prevented the inclusion of this parameter into the model. For HC, this favoured the inclusion of *state* into the model as here this covariate was part of the model for the random effect mean.

Similarly, the chain from the GZM-state analysis revealed different probabilities of being included in the model for *Julian day* (total probability of being included in any density model = 0.26) compared to the other analyses (total probability of being included in any model = 1.00, 1.00 and 0.93 for GZM, HC and HC-state, respectively). As this covariate was not part of the preferred model for the GZM-state analysis, this likely caused means and CRIs of the intercept and *state* coefficients to be different compared to the other analyses. The mean and CRIs for this parameter were also larger for GZM than for HC or HC-state. It is likely that for the latter two, the inclusion of the *state* covariate in the density model caused this change rather than any change in analysis method or resulting dynamics. On the other hand, implementing hierarchical centering had no effect on detection function parameters or the *type* covariate from the density model. For these, summary statistics were nearly identical between all analyses.

We could not confirm the findings of Browne (2004), that implementing hierarchical centering would improve the effective sample size for the covariate involved in the

centering. He compared the effective sample sizes for the same covariate in two different MCMC chains, one with hierarchical centering and one without. Our example was different in that we used RJMCMC chains. However, for our case study, effective sample sizes were similar for all parameters between the analyses except for the random effect standard deviation and the *state* covariate. For the former, the effective sample size increased using hierarchical centering while for the latter effective sample sizes decreased using hierarchical centering – the latter being a contradiction to Browne (2004).

Overall we showed that the main benefit of implementing hierarchical centering lies in improving mixing between-models and, hence, in improving inference on model parameters. For our case study, the parameter of interest was *type*, i.e. how densities differed between control and treatment points. For this parameter, model probabilities, summary statistics and hence inference were nearly identical between the analyses. However, had we been interested in how densities changed between different states, the parameter of interest would have been *state*. Inference on this covariate using the GZM analysis could potentially have led us to believe falsely that this covariate had no effect on densities. Similarly, had our interest been temporal variation in densities of indigo buntings throughout the year, inference on the *Julian day* covariate from the GZM-state analysis could have falsely lead us to believe that this covariate had no effect on densities.

Chapter 6

Incorporating random effects in the detection function for line transect data

6.1 Introduction

Distance sampling is a commonly used tool for wildlife studies where the interest lies in obtaining estimates $\hat{N}$ of the number of animals N in the study area (e.g. Archer et al., 2008; Buckland et al., 2001; De Segura et al., 2007; Edwards and Kleiber, 1989; Palacios et al., 2012). The most common form of distance sampling is line transect sampling, in which lines are laid out in the study area according to some design and observations of animals are made along these lines during the survey(s). To scale up from the observed number of animals during the survey(s) n to $\hat{N}$, a Horvitz-Thompson-like estimator is used (Borchers et al., 1998; Yuan, 2012) where the number of observed animals is divided by the inclusion probability, assuming that

this inclusion probability is the same for all animals. Hence, this method relies on obtaining an asymptotically unbiased estimate of the inclusion probability from additional data collected during the survey. In the simplest case, this information consists of the distances to the detections.

For line transect data the inclusion probability consists of two components, the average detection probability P_a within the search area and the probability that the animal was within the covered area π_a (i.e. the area that the observer(s) searched in) (Buckland et al. 2001, Thomas et al., unpublished manuscript). Then, the estimator for the total number of animals $\hat{N}$ is given by $\hat{N} = n/(\pi_a \hat{P}_a)$. The proportion of area surveyed π_a ($\pi_a = \textit{covered area/survey area} = a/A$) assumes that the covered area is a good representation of the study area which relies on an appropriate survey design (e.g. Buckland et al., 2001; Strindberg and Buckland, 2004). To obtain an estimate of the average detection probability $\hat{P}_a$ within the search area, we need a flexible model for the detection function to capture the decay in detection probabilities with increasing distance from the line. Using conventional distance sampling (CDS) methods it is generally assumed that all animals on the line are detected with certainty while with increasing distances, detection probabilities may decrease (Buckland et al., 2001). Flexible detection functions that capture the shape of this decay well, are the half-normal and hazard-rate detection functions in combination with adjustment terms. An estimator for animals within the covered region is then given by

$$\hat{N}_{P_a} = n/\hat{P}_a. \tag{6.1}$$

It has been shown that this estimator yields approximately unbiased estimates, even though different animals may have different detection probabilities, e.g. due to different properties that makes them easier/harder to detect or different observation conditions at different times of the survey. This quality is generally referred to as the pooling robustness criterion (Burnham et al. 1980; Buckland et al. 2004), which, for line transects, may be expressed as:

$$\frac{n}{\hat{P}_a} \approx \sum_{e=1}^{n} \frac{1}{\hat{P}_{a_e}}. \tag{6.2}$$

On the right hand side of this equation, the estimate of the average detection probability $\hat{P}_{a_e}$ is given for each of $e = 1, 2, ..., n$ individuals. Despite the fact that pooling robustness generally holds, there may be an interest in modelling the P_{a_e}. This interest may arise, e.g. when abundance estimates are desired for subareas of the study region (e.g. different habitats) or during different survey years. It has also been shown that this approach may increase precision of the abundance estimate (Marques et al., 2007). The P_{a_e} are generally estimated by including covariates in the detection function model which allow the detection function for the eth detection to be adjusted depending on the covariate values associated with the detection (Marques and Buckland, 2003, 2004).

An approach for fitting flexible detection functions which allows but does not rely on the use of covariates was presented by Miller and Thomas (unpublished manuscript), who proposed the use of mixture models. Here, the detection function is composed of (in their notation) J mixture components (each a detection function that may contain adjustment terms and/or covariates) which are scaled by mixing proportions ϕ_j. However, we investigate the case where heterogeneity in detection probabilities exists

but requires too many parameters to be estimated using fixed effect covariates or a small number of mixture components. We build upon the multiple covariate distance sampling (MCDS) approach where the fixed effect covariates enter the detection function via the scale parameter of the half-normal or hazard-rate key function (Marques and Buckland, 2003). We use a half-normal detection function and model the scale parameter using an intercept common to all detections and random effects coefficients for the different detections for which we assume normality with a zero-mean (Laake and Skaug, unpublished manuscript). This allows fitting a flexible detection function at the cost of only one additional parameter, the random effects standard deviation. This model may also be extended to include fixed effect covariates. We present the likelihood for this detection function as well as two estimators for abundance and associated variances. One estimator is a function of the expected average detection probability estimated from a detection function with a random effect. The second estimator is a function of the expected value of the reciprocal of the average detection probability and also estimated from a detection function with a random effect. Both estimators use simulated values from the estimated distribution of the random effects coefficients to calculate the expected values.

In the following sections we revise the likelihood and estimators for abundance using CDS methods (section 6.2). We then develop the likelihood and estimators for incorporating random effects in the detection function (section 6.3). We compare the performance of the new estimators to CDS estimators using a simulation study (section 6.4). Finally, we discuss the consequences of these findings (section 6.6).

6.2 The detection function without random effects - conventional distance sampling methods

We begin by reviewing the likelihood for the probability density function of observed distances $f(y)$ and estimators for the average detection probability using CDS methods. The probability density function is given by $f(y) = \pi(y)g(y)/\int_0^w \pi(y)g(y)dy$ (Buckland et al., 2001), where w is the truncation distance (i.e. the furthest distance included in the analysis). The y refer to perpendicular distances from the line; $\pi(y)$ describes the expected distribution of animals with respect to increasing distances from the line out to w. As is generally done for CDS methods, we assume that lines were placed in the study area independently of the distribution of animals, hence on average resulting in a uniform distribution for this function, $\pi(y) = 1/w$.

For CDS methods, we also assume that all animals on the line are detected with certainty. Then, the detection function $g(y)$ can be modelled using e.g. the half-normal $(\exp(-y^2/2\sigma^2)$ or the hazard-rate model $(1 - exp(-(y/\sigma)^\tau))$. In the case that distances are measured exactly (as opposed to in distance intervals) the likelihood of the parameter given the data is defined as:

$$L_y(\boldsymbol{\theta}) = \prod_{e=1}^n f(y_e), \tag{6.3}$$

where y_e is the distance of the eth detection from the line, $e = 1, 2, ..., n$. Maximising this likelihood returns the values for the detection function parameters that yields the best fit to the observed data given the detection function. Generally, we maximise the log of eqn (6.3). AIC values may be used to compare the fits of competing models. We use $\hat{g}(y)$ to denote the best fitting detection function.

To obtain an estimate of the average detection probability P_a in the covered area, we use the following relationship:

$$\hat{P}_a = \frac{\hat{\mu}}{w} = \frac{\int_0^w \hat{g}(y)dy}{w}, \tag{6.4}$$

that expresses the average detection probability as the effective strip half width $\hat{\mu}$ divided by the truncation distance. The effective strip half width is defined as the perpendicular distance from the line beyond which as many animals were detected as were missed within and is estimated by integrating $\hat{g}(y)$ between zero and the truncation distance. We divide the number of detections n by the estimated average detection probability to obtain an estimate of the number of animals in the covered region $\hat{N}_{P_a}$ as in eqn (6.1) and refer to this as the P_a estimator. We then divide this estimate by π_a to scale up from the estimated numbers of animals in the covered area to the estimated number of animals in the study area:

$$\hat{N} = \frac{\hat{N}_{P_a}}{\pi_a}. \tag{6.5}$$

6.2.1 Estimating the variance

Maximising the likelihood given in eqn (6.3) using e.g. the *optim* command in R, produces a Hessian matrix $\boldsymbol{H}$ which can be used to estimate the variance of the detection function parameter estimates $\hat{\boldsymbol{\theta}}$. The variance associated with the ith element of $\hat{\boldsymbol{\theta}}$ is the ith element of the main diagonal of the inverse Hessian, $\boldsymbol{H}^{-1}$.

To estimate the variance of the average detection probability using the Hessian, we use the first derivative of eqn (6.4) with respect to the individual parameters contained in $\boldsymbol{\theta}$ evaluated at $\hat{\boldsymbol{\theta}}$ (Borchers et al., 2002). An estimate of the variance associated

with $\hat{P}_a$ is given by:

$$\widehat{var}(\hat{P}_a) = \frac{1}{w^2} \left[\frac{\partial \hat{\mu}}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \right]^T \boldsymbol{H}^{-1} \left[\frac{\partial \hat{\mu}}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \right]. \tag{6.6}$$

These first derivatives can be difficult to obtain analytically. Hence, we use finite differences instead (see Appendix C.1 for details).

The estimate of the variance of $\hat{P}_a$ in combination with an estimate of the variance of the encounter rate (Fewster et al., 2009) along the lines can be converted into a variance estimate of the number of animals in the study area (abundance) using the delta method (Buckland et al., 2001). The variance of abundance can be used to create log-based confidence intervals for the abundance estimate. These methods are detailed in Appendix B.

6.3   The half-normal detection function with random effects

For this approach we assume that heterogeneity in detection probabilities exists between different detections due to differences in detection functions. As before, we use the conditional probability density function of observed distances, $f(y)$. However, now we introduce a random effect in the model for the scale parameter of the half-normal detection function which is given for the eth detection as:

$$\sigma_e = \beta_0 \times \exp(b_e). \tag{6.7}$$

Here, β_0 is the intercept common to all detections and b_e is the random effect coefficient for the eth detection. We assume these coefficients are normally distributed with a zero-mean and unknown standard deviation σ_b ($b_e \sim N(0, \sigma_b^2)$). For simplicity, we focus on the half-normal key function without adjustment terms, although methods could be extended to the hazard-rate model and to include adjustment terms. The latter, however, might not be necessary as this function with the random effects may provide sufficient flexibility.

The likelihood of the parameters ($\boldsymbol{\theta} = \beta_0, \sigma_b$) given the data is then defined as (Laake and Skaug, unpublished manuscript):

$$L_y(\boldsymbol{\theta}) = \prod_{e=1}^{n} \frac{\int_{-\infty}^{\infty} g(y_e|\beta_0, b_e) \times \pi(b_e) db_e}{\int_{-\infty}^{\infty} \int_0^w g(y|\beta_0, b_e) dy \times \pi(b_e) db_e} \qquad (6.8)$$

where y_e refers to the distance of the eth detection from the line (with $e = 1, 2, ..., n$). The component $\pi(b_e)$ is the normal density for the eth random effect coefficient. The estimates of the detection function parameters are obtained by maximising the log of this likelihood. The random effect coefficients are not estimated individually but are integrated out.

To obtain an estimate of the number of animals in the covered area, two estimators are available that include a random effect (Potts, 2011). Both estimators use simulated values from the distribution of the random effects coefficients. These and their respective variance estimators are discussed in the following.

6.3.1   The P_r estimator

This approach uses the expected value of the average detection probability $\hat{P}_r$ (defined in the following eqn) to scale up from the number of detected animals n to an estimate

of the number of animals in the covered area, denoted as $\hat{N}_{P_r}$:

$$\hat{N}_{P_r} = \frac{n}{\hat{P}_r} = \frac{n}{\int_{-\infty}^{\infty} \hat{\pi}(b_e)\hat{P}_e db_e} \tag{6.9}$$

where $\hat{\pi}(b_e)$ denotes the normal density of the random effects coefficient b_e (with $b_e \sim N\left(0, \hat{\sigma}_b^2\right)$). The estimate of the average detection probability for the eth detection $\hat{P}_e$ is given by $\hat{P}_e = \int_0^w \hat{g}(y, \hat{\beta}_0, b_e)dy/w = \int_0^w \exp\left(\frac{-y^2}{2(\hat{\beta}_0 \times \exp(b_e))^2}\right) dy/w$ and hence is a function of the random effects coefficients b_e (as well as of the fixed intercept $\hat{\beta}_0$) (modified from Potts, 2011). Here, $\hat{g}()$ denotes the half-normal detection function for which the scale parameter is modelled using $\hat{\beta}_0$ and b_e (see eqn (6.7)). Similar to the P_a estimator from section 6.2, we will refer to eqn (6.9) as the P_r estimator in the following.

In practice, the random effect of eqn (6.9) can be integrated out via simulation. For this we combine eqns (6.4) and (6.7) and simulate values for b_e by drawing $E = 10000$ random samples from $N(0, \hat{\sigma}_b^2)$. For each sampled b_e, a new value is calculated for the average detection probability which we denote $\hat{P}_{e_{sim}}$. The expected value of the average detection probability $P_{r_{sim}}$ is estimated using:

$$\hat{P}_{r_{sim}} = \frac{\sum_{e=1}^{E} \hat{P}_{e_{sim}}}{E} = \frac{\sum_{e=1}^{E} \hat{\mu}_{e_{sim}}}{E \times w} = \frac{\sum_{e=1}^{E} \int_0^w \hat{g}(y, \beta_0, b_e)dy}{E \times w}, \tag{6.10}$$

which replaces the $\hat{P}_r$ in eqn (6.9) for obtaining an estimate of number of animals in the covered area. To scale up from numbers of animals in the covered area to number of animals in the study area we use π_a as before:

$$\hat{N} = \frac{\hat{N}_{P_r}}{\pi_a}. \tag{6.11}$$

6.3.1.1   Estimating the variance

Again, we use the inverse of the Hessian $\boldsymbol{H}$ returned from the optimisation algorithm. The main diagonal of $\boldsymbol{H}^{-1}$ gives the variance estimates of the parameters ($\hat{\boldsymbol{\theta}} = \hat{\beta}_0, \hat{\sigma}_b$). For converting this Hessian into a variance estimate for $\hat{P}_r$, we use the derivative of eqn (6.10) with respect to the parameters $\boldsymbol{\theta}$ evaluated at $\hat{\boldsymbol{\theta}}$. The estimate of the variance is then given by:

$$\widehat{var}(\hat{P}_r) = \frac{1}{w^2}\left[\frac{\partial \hat{\mu}_r}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}\right]^T \boldsymbol{H}^{-1}\left[\frac{\partial \hat{\mu}_r}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}\right], \tag{6.12}$$

where $\hat{\mu}_r = \int_{-\infty}^{\infty} \hat{\pi}(b_e)\int_0^w \hat{g}(y, \hat{\beta}_0, b_e)dy db$. As before, the derivative is obtained using finite differences. However, here we evaluate this derivative for $E = 10000$ simulated values for b_e and take the average of the resulting values (for details see Appendix C.2). As before, we combine the variances of the estimated average detection probability and encounter rate using the delta method to obtain an estimate of the variance of abundance in the study area (Appendix B).

6.3.2   The $(1/P_r)$ estimator

For this approach we use the expected value of the reciprocal of the average detection probability $(\widehat{1/P_r})$ (defined in the following eqn). The estimator for number of animals in the covered area for this approach, $\hat{N}_{1/P_r}$ is given by:

$$\hat{N}_{1/P_r} = n \times \left(\widehat{\frac{1}{P_r}}\right) = n \times \int_{-\infty}^{\infty} \frac{\hat{\pi}(b_e)}{\hat{P}_e}db_e \tag{6.13}$$

(modified from Potts, 2011). As before, $\hat{\pi}(b_e)$ denotes the normal density of the random effects coefficients, $b_e \sim N(0, \hat{\sigma}_b^2)$. The average detection probability $\hat{P}_e$ is

calculated using $\hat{P}_e = \int_0^w \hat{g}(y, \hat{\beta}_0, b_e)dy/w = \int_0^w \exp\left(\frac{-y^2}{2(\hat{\beta}_0 \times \exp(b_e))^2}\right) dy/w$ for varying values of b_e.

Again, the random effect of eqn (6.13) is integrated out via simulation by drawing $E = 10000$ random samples from $N(0, \hat{\sigma}_b^2)$. For each sampled b_e, we calculate a new value for $(1/\hat{P}_{e_{sim}})$. The expected value of the reciprocal of the average detection probability $(1/\widehat{P_{r_{sim}}})$ is then given by:

$$\left(\widehat{\frac{1}{P_{r_{sim}}}}\right) = \frac{1}{E}\sum_{e=1}^{E}\frac{1}{\hat{P}_{e_{sim}}} = \frac{1}{E}\sum_{e=1}^{J}\frac{w}{\hat{\mu}_{e_{sim}}} = \frac{1}{E}\sum_{e=1}^{E}\frac{w}{\int_0^w \hat{g}(y, \hat{\beta}_0, b_e)dy}. \tag{6.14}$$

For this approach $(1/\widehat{P_{r_{sim}}})$ is substituted for $(1/\widehat{P_r})$ in eqn (6.13). Similar to above, we use the following to scale up from numbers of animals in the covered area to number of animals in the study area using π_a:

$$\hat{N} = \frac{\hat{N}_{1/P_r}}{\pi_a}. \tag{6.15}$$

6.3.2.1   Estimating the variance

The variance estimator associated with the $(1/\widehat{P_r})$ estimator is given by:

$$\widehat{var}\left(\widehat{\frac{1}{P_r}}\right) = w^2 \left[\frac{\partial\left(\widehat{\frac{1}{\mu_r}}\right)}{\partial\boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}\right]^T \boldsymbol{H}^{-1} \left[\frac{\partial\left(\widehat{\frac{1}{\mu_r}}\right)}{\partial\boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}\right], \tag{6.16}$$

where $(1/\widehat{\mu_r}) = \int_{-\infty}^{\infty}\frac{\hat{\pi}(b_e)}{\int_0^w \hat{g}(y, \hat{\beta}_0, b_e)dy}db$. As parameter estimates for this approach are obtained by maximising the same likelihood as for the P_r estimator, we use the same Hessian as in section 6.3.1.1. However, now we use the derivative of eqn (6.14) with respect to the detection function parameters $\boldsymbol{\theta} = \beta_0, \sigma_b$ evaluated at $\hat{\boldsymbol{\theta}}$. As before,

the derivative is obtained using finite differences (see Appendix C.3). The variance of the total abundance in the study area is obtained by combining $\widehat{var}\left(\frac{1}{P_r}\right)$ and the encounter rate variance using the delta method. Log-normal confidence intervals are constructed using the methods described in Appendix B.

6.4 Simulation study

6.4.1 Generating simulated data

We conducted a simulation study to evaluate the bias and precision of estimators for P_a (eqn (6.4)), P_r (eqn (6.10)) and $(1/P_r)$ (eqn (6.14)) and respective variance estimators (eqns (6.6), (6.12) and (6.16)) for different degrees of heterogeneity in detection probabilities between detections.

Each simulation consisted of the following steps. We used the R package *wisp* (Zucchini et al., 2007) to create a new study area of known dimensions and abundance of animals that were evenly distributed throughout the study area. Animals occurred as individuals as opposed to in clusters and had equal exposures (exposures allow the user of *wisp* to make some animals more detectable than others). In this study area, we randomly placed a number of lines of the same length along which detections were made within a predefined search area (out to truncation distance w) during a simulated survey. Detection probabilities were certain on the line and decreased with increasing distance from the line according to a half-normal detection function model.

Using the *wisp* package, the only option to manipulate the scale parameter of the half-normal detection function directly is by modelling this parameter individually

for each line in the study area with the *setpars.survey.lt* function. The arguments *disthalf.min* and *disthalf.max* of this function allow specifying the perpendicular distances, at which detection probability is 0.5 for animals with minimum and maximum exposure, respectively. For a half-normal detection function, the relationship between either of these distances and the scale parameter is given by $dishalf.min = c * scale$ with $c = \sqrt{-2\log(0.5)}$ or equally $dishalf.max = c * scale$. Setting $disthalf.min = disthalf.max$ for each line, all animals detected along the same line shared a common detection function.

Using this relationship between *disthalf.min* and the scale parameter, heterogeneity in detection probabilities between individual lines was introduced using eqn (6.7) where the scale parameter of the half-normal detection function for each of $k = 1, 2, ..., K$ lines was modelled as $\sigma_k = \beta_0 \times exp(b_k)$ (with $b_k \sim N(0, \sigma_b^2)$). The parameters β_0 and σ_b were known and a random sample was drawn for the b_k of each line. Detections that were made along these lines were extracted from the *wisp* objects and analysed with each of the estimators (see next section for analysis details).

This completed one simulation. For each simulation we specified a different seed for the steps involving the placement of animals in the study area, the drawing of the random effect coefficients and placement of the lines to ensure that each simulation was unique. The simulations were repeated 1000 times for each set of known parameters. These settings included the number of lines, the number of animals in the study area N, the truncation distance w, the detection function parameters β_0 and σ_b and the size of the study area (and hence the proportion of the study area covered π_a, Table 6.1). The differences in numbers of lines, N and detection probabilities between the different sets resulted in differences in the average total number of detections, the

average number of detections per line and the number of lines that had detections.

Table 6.1: Settings for the four sets of simulations: K and N respectively refer to the total number of lines and animals in the study area, w is the truncation distance, β_0 and σ_b are the detection function parameters and π_a is the proportion of the study area covered. Also shown are the resulting means (and standard deviations) of the total number of detections, detections per line and lines with detections. Note that the size of the study area A varied between sets with $A = 2wK/\pi_a$.

	$\sigma_b = 0.2$		$\sigma_b = 0.5$	
Set	**1**	**2**	**3**	**4**
K	200	50	200	50
N	10 000	10 000	20 000	8 000
w	5	5	18	18
β_0	1.70	1.70	5.10	5.10
π_a	0.2	0.125	0.072	0.18
Total detections	862.25 (31.65)	536.93 (27.82)	552.70 (28.82)	553.30 (42.39)
Detections per line	4.39 (2.17)	10.74 (3.83)	3.10 (1.91)	11.09 (5.89)
Lines with detections	196.30 (1.89)	49.99 (0.08)	178.07 (4.43)	49.90 (0.30)

6.4.1.1   Visualising heterogeneity in detection probabilities

To illustrate the effect of normally distributed random effects, i.e. $b_e \sim N(0, \sigma_b^2)$, entering the detection function via the scale parameter using the model from eqn (6.7), we plotted these for the cases where β_0 was fixed and $b_e \sim N(0, 0.2^2)$ or $b_e \sim N(0, 0.5^2)$ in Figure 6.1. For each of the two cases, 200 random samples for b_e were drawn and detection functions calculated of which we plotted those with the minimum, maximum, mean, 2.5 and 97.5 percentile values of b_e. In addition, the mean of the 200 detection functions was calculated and plotted in the same Figure.

Naturally, heterogeneity increased with increase in random effects standard deviation. However, the difference between the detection function where the scale parameter was modelled using the mean of the random effects coefficients ($mean(b_e)$ from Figure 6.1) and the mean of all 200 detection functions ($mean$ from Figure 6.1) also increased with increase in random effects standard deviation. For the $mean$ of these functions, detection probabilities were smaller for small distances but larger for larger distances when compared with the detection function corresponding to the mean of the random effect. This was an indication that a half-normal model without adjustment terms or random effects might not be sufficiently flexible to capture the decay in overall detection probabilities for this type of heterogeneity, if we rely on pooling robustness and do not model the heterogeneity.

Hence, the four sets of simulations with two different values for the random effect standard deviation will allow us to assess the effects of increasing amounts of heterogeneity in detection probabilities on the performance of the estimators described above.

6.4.2   Analysis

For each simulation, we extracted the detections from the *wisp* objects and fitted the probability density functions both without and with random effects by maximising the log of the respective likelihoods given in eqns (6.3) and (6.8). For the models without the random effect we included a selection routine where the best fitting model was chosen between a set of contending models. This was done to ensure that
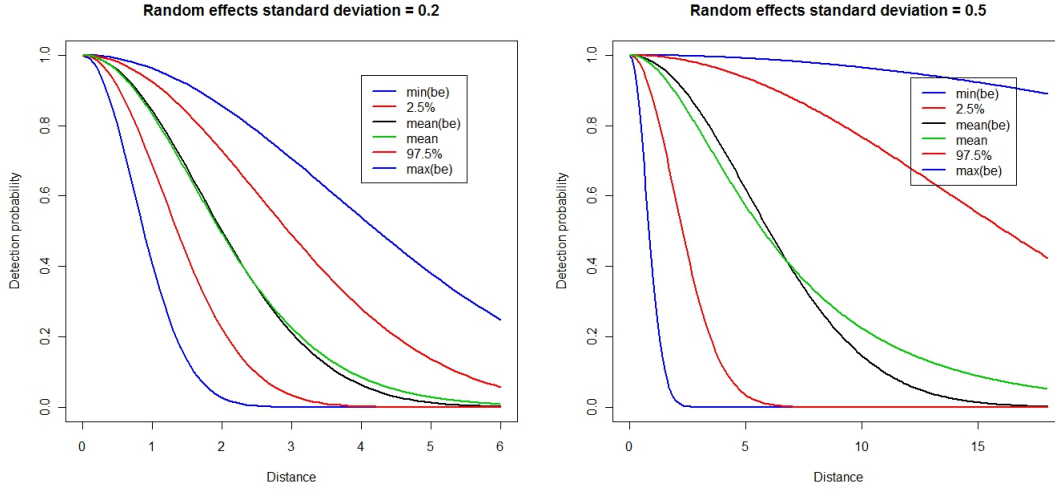
Figure 6.1: Half-normal detection functions for which the scale parameter was modelled with random effects. Shown are the functions resulting from the minimum, maximum, 2.5 and 97.5 percentiles and the mean of randomly sampled 200 coefficients b_e. In addition, the mean of all detection functions is plotted in green.

a sufficiently flexible model was fitted to the detections so that pooling robustness would hold.

We tested the half-normal and hazard-rate key functions and without adjustment terms or with cosine adjustment terms of orders up to 5. However, the hazard-rate function and the adjustment terms of orders higher than 3 were omitted as best bias and coverage results were obtained by using only the half-normal with adjustment terms of orders up to 3. Hence, models included were the half-normal key function either without adjustment terms or with cosine adjustment terms of orders up to 3. For the models fitted without the random effect, we distinguished between estimators with and without model selection in the following where $P_a HN$ refers to the P_a estimator where only the half-normal model without adjustment terms was considered, while $P_a *$ refers to the estimator with model selection.

We then used the estimators for the average detection probability without a random effect (P_aHN estimator from eqn (6.4) and $\widehat{var}(\hat{P}_aHN)$ from eqn (6.6)) to obtain an estimate of abundance in the covered area and scaled up using methods described in section 6.2 to obtain an estimate of abundance in the study area $\hat{N}$ with associated 95% log-normal confidence intervals for each simulation. The same was repeated for the P_a* estimator.

For the estimators with a random effect we only included the half-normal model without adjustment terms. Even though the data was generated using the same random effect coefficient for all detections made along the same lines (i.e. $b_k \sim N(0, \sigma_b^2)$ for $k = 1, 2, ..., K$ lines), we used the model given in eqn (6.8) where the random effects coefficients were assumed to differ between different detections (i.e. $b_e \sim N(0, \sigma_b^2)$ for $e = 1, 2, ..., n$ detections). One could argue that in a real life situation, heterogeneity in detection probabilities exists between groups of detections (e.g. those made along different lines, by different observers, different observation conditions, etc. or a combination of different factors), but too many different groups exist to model them as fixed effects. In addition, as it is often impossible to define these groups that share a common detection function in a real life situation, we wanted to investigate how the estimators from above perform when ignoring these groups and assuming individual random effects.

We obtained estimates of abundance in the covered area $\hat{N}_{P_r}$ and $\hat{N}_{1/P_r}$ from the resulting parameter estimates where the likelihood included a random effect, using the P_r and $(1/P_r)$ estimators (eqns (6.10) and (6.14)). Associated variances $\widehat{var}(\hat{P}_r)$ and $\widehat{var}(\widehat{1/P_r})$ were estimated using eqns (6.12) and (6.16). The estimates of abundance in the covered area were then scaled up to estimated abundance in the study area

including associated 95% log-normal confidence intervals using the methods described above (sections 6.3.1 and 6.3.2).

The performance of the estimators was evaluated by examining the bias of the resulting abundance estimates and whether 95% log-normal confidence intervals produced the appropriate coverage rate. Percent bias was estimated by taking the mean and standard error of $100 \times (\hat{N} - N)/N$ of the 1000 simulations ($\hat{N}$ and N are the estimated and true abundances in the study area, respectively). Coverage was considered appropriate if anywhere between 936 and 962 confidence intervals from a set of 1000 simulations covered the true value for abundance. (1000 binomial trials, each with p=0.95, generate between 936 and 962 successes 95% of the time.)

To evaluate the performance of the variance estimators for $\hat{P}_a HN$, $\hat{P}_a*$, $\hat{P}_r$ and $(\widehat{1/P_r})$, we compared the standard deviation of the 1000 abundance estimates to the mean of the 1000 associated standard errors for the respective simulation sets. These should approximately be the same.

6.4.3   Results

6.4.3.1   Parameter estimates

We begin by summarising the parameter estimates obtained for the simulations by maximising the likelihoods using a half-normal detection function without and with normally distributed random effects (without adjustment terms). Mean and standard deviations of these estimates are shown in Table 6.2. The estimates for β_0 and σ_b were approximately unbiased when using the likelihood with the random effects. The fact that β_0 estimates were larger than truth for the estimator without the random effect

was expected as the mean of all detection functions from the respective simulation was larger than a detection function using β_0 only (i.e. $1/K \sum_{k=1}^{K}(\beta_0 \times \exp(b_k)) > \beta_0$).

Table 6.2: Mean (and standard deviation) of parameters estimates obtained from 1000 simulations using the half-normal detection function with and without random effects (RE).

Set	1		2	
Parameter	$\hat{\beta}_0$	$\hat{\sigma}_b$	$\hat{\beta}_0$	$\hat{\sigma}_b$
without RE	1.81 (0.05)	-	1.79 (0.08)	-
with RE	1.72 (0.08)	0.19 (0.10)	1.68 (0.11)	0.20 (0.11)
Truth	1.70	0.20	1.70	0.20
Set	3		4	
Parameter	$\hat{\beta}_0$	$\hat{\sigma}_b$	$\hat{\beta}_0$	$\hat{\sigma}_b$
without RE	6.71 (0.33)	-	6.66 (0.48)	-
with RE	5.10 (0.57)	0.50 (0.12)	5.08 (0.64)	0.50 (0.12)
Truth	5.10	0.50	5.10	0.50

6.4.3.2 Bias in abundance estimates using estimated parameter values

The P_aHN estimator yielded negatively biased abundance estimates for all simulation sets, with bias of less than 5% for sets 1 and 2 with the smaller amount of heterogeneity and between 16 and 18% for sets 3 and 4 with the larger random effects standard deviation. Including model selection improved bias for CDS methods. Negative bias using the P_a* estimator was less than 2% for sets 1 and 2 and between 2 and 4% for sets 3 and 4. The P_r estimator on the other hand provided abundance estimates that were on average unbiased for each set (with bias of less than 1%). The $(1/P_r)$

estimator yielded positively biased abundance estimates for all sets, between 3 and 7% for sets 1 and 2 and between 28 and 30% for sets 3 and 4.

Table 6.3: Average bias of abundance estimates (and standard errors) yielded by different estimators. The estimators involving P_a did not include a random effect in the detection function, those involving P_r did.

Set	1	2	3	4
P_a HN	-4.44% (0.13)	-3.57% (0.17)	-17.07% (0.15)	-16.38% (0.16)
P_a *	-1.84% (0.18)	-1.07% (0.22)	-3.46% (0.22)	-2.74% (0.22)
P_r	-0.50% (0.16)	-0.97% (0.20)	0.40% (0.27)	0.95% (0.26)
$(1/P_r)$	3.84% (0.24)	6.06% (0.31)	29.30% (0.71)	28.32% (0.75)

6.4.3.3 Coverage rates

Assessing the coverage rate of 95% log-normal confidence intervals, we found substantial differences between estimators (Table 6.4). As mentioned above, we considered the coverage rates between 93.6 and 96.2% to be appropriate. The $P_a HN$ estimator performed worst where coverage rates were too low for all simulation sets. Here, coverage rates decreased down to only 9% with increase in random effects standard deviation. For the $P_a *$ estimator, coverage rates were appropriate for set 2, too low for sets 1 and 3 and too high for set 4. For the P_r estimator, coverage rates were appropriate for sets 1 and 3 and slightly too high for sets 2 and 4. Using the $(1/P_r)$ estimator, coverage rates were appropriate for set 1, too high for set 2 and too low for sets 3 and 4.

Table 6.4: Coverage rates of log-normal 95% confidence intervals around estimates of abundance in the study area yielded by the different estimators.

Set	1	2	3	4
P_a HN	86.1%	93.5%	9.9%	37.3%
P_a *	90.3%	93.6%	92.5%	98.3%
P_r	96.2%	97.7%	94.2%	97.3%
$(1/P_r)$	95.7%	96.7%	70.1%	76.4%

6.4.3.4   Performance of variance estimators

The $var(\hat{P}_aHN)$ and $var(\hat{P}_r)$ estimators consistently overestimated the variance. The $var(\hat{P}_a*)$ estimator underestimated variance for sets 1 and 3 and overestimated variance for sets 2 and 4. The $var(\widehat{1/P_r})$ estimator overestimated the variance for sets 1 and 2 and underestimated it for sets 3 and 4. Variance estimates were generally smallest for the $var(\hat{P}_aHN)$ estimator and increased by including model selection or random effects. Variance estimates were highest for the $var(\widehat{1/P_r})$ estimator.

6.5   Simulations without random effects

In addition to the above simulations, we ran four sets of simulations (1000 iterations each) where the number of lines and animals in the study area, the truncation distance, the scale parameter of the half-normal detection function and the proportion of the study area that was covered were the same as in Table 6.1. The random effects standard deviation, however, was set to zero. These simulations served two purposes.

Table 6.5: Assessing performance of variance estimators by comparing the standard deviation of the 1000 abundance estimates (sd) to the mean of the standard errors associated with the individual abundance estimates ($\bar{se}$).

Set	1		2		3		4	
Estimator	sd	$\bar{se}$	sd	$\bar{se}$	sd	$\bar{se}$	sd	$\bar{se}$
$var(\hat{P}_a)$ HN	399.89	428.03	531.45	579.87	961.47	1023.87	411.06	549.37
$var(\hat{P}_a)$ *	577.42	500.50	657.47	703.38	1415.11	1389.81	540.08	700.08
$var(\hat{P}_r)$	494.62	540.62	641.52	724.13	1679.46	1711.96	651.56	825.81
$var(\widehat{1/P_r})$	762.34	808.44	975.78	1072.72	4610.80	4092.37	1906.73	1723.42

Firstly, we wanted to verify that this simulation technique using the *wisp* package (from creating a study area to surveying along the randomly placed lines) produced detections that could be analysed reliably using CDS methods. Hence, we evaluated bias and coverage rates for the $\hat{P}_aHN$ estimator. No model selection was included here as we knew the true model the data originated from was the half-normal key function without adjustment terms. Secondly, we wanted to test how the estimators from section 6.3 performed when the random effect was fitted despite the fact that all detections originated from one common detection function.

Bias in abundance estimates was generally low (i.e. less than 2%) for the P_aHN and the P_r estimator, while it was slightly larger for the $(1/P_r)$ estimator (Table 6.6). Coverage rates were high enough for all estimators, except they were larger than what we defined as appropriate above for the P_aHN estimator for sets 2 and 4, where they were 100% and slightly too large for the $(1/P_r)$ estimator for set 2.

Hence, we were able to draw the following conclusions: 1. The detections can be analysed reliably using CDS methods, although some of the coverage rates using the

$P_a HN$ estimators were too high. 2. The P_r estimator produces reliable results even though a random effect was fitted to detections that shared a common detection function. The positive bias in the estimates from the $(1/P_r)$ estimator was moderate.

Table 6.6: Bias (and standard errors) in abundance estimates as well as coverage rates of 95% log-normal confidence intervals for the four sets of simulations without random effects yielded by the estimators in the left column.

Set	1	2	3	4
		Bias		
$P_a HN$	0.88% (0.13)	-1.16% (0.26)	-0.22% (0.05)	0.35% (0.17)
P_r	0.64% (0.15)	1.66% (0.31)	0.38% (0.05)	1.52% (0.18)
$(1/P_r)$	2.13% (0.19)	4.89% (0.44)	0.50% (0.06)	2.44% (0.21)
		Coverage		
$P_a HN$	94.9%	100.0%	96.1%	100.0%
P_r	95.2%	96.0%	95.6%	95.1%
$(1/P_r)$	95.6%	96.9%	96.0%	95.9%

6.6 Discussion

It is generally assumed that, due to pooling robustness, CDS methods provide unbiased estimates of abundance, despite potential heterogeneity in detection probabilities, as long as a sufficiently flexible detection function is fitted to the data (e.g. Buckland et al., 2001, 2004; Burnham et al., 1980, Thomas et al., unpublished manuscript).

If pooling robustness held for the generated data of our simulation study, we would have expected that resulting abundance estimates were on average unbiased and that coverage rates were appropriate for the P_a* estimator, i.e. using CDS methods (without a random effect in the detection function) and including model selection. Model selection should ensure that a sufficiently flexible detection function was chosen for the data that allowed unbiased estimation of the average detection probability. However, negative bias increased to up to 3.46% for this estimator with increase in the amount of heterogeneity that was introduced in the data.

When modelling the detection function with a random effect, however, the P_r estimator performed consistently well without adjustment terms or model selection. This estimator yielded abundance estimates with an average bias of less than 1% regardless of the amount of heterogeneity in detection probabilities. Using the $(1/P_r)$ estimator on the other hand, positive bias increased substantially with increase in heterogeneity. The P_r estimator also returned the best coverage rates among all tested estimators. Coverage rates using the P_a* estimator were too low for simulation sets 1 and 3. For all simulations, this was due to underestimating the upper boundary of the confidence interval. Interestingly, it was also sets 1 and 3 where the variance of abundance was underestimated. To obtain coverage rates that were high enough to be appropriate using this estimator, the data needed to be truncated at 2.0m and 4.5m for sets 1 and 3, respectively. This truncation disposed of the long tail of detections at distances where average detection probabilities were low. For comparison, the P_r estimator returned appropriate coverage rates even when including this tail. Here, the truncation distance was set to 5.0m and 18.0m for sets 1 and 3, respectively. Hence, this estimator allowed inclusion of distances with low detection probabilities that needed

to be excluded from the analysis for producing reliable results using CDS methods. Interestingly, the P_r estimator returned abundance estimates that were nearly unbiased and appropriate coverage rates for all four sets of simulations where no heterogeneity existed between detections. Here, bias was just slightly larger than for the P_aHN estimator. Hence, we conclude that the P_r estimator produces reliable results for both cases investigated in our simulation study, i.e. with and without heterogeneity in detection probabilities. However, the only type of heterogeneity tested in this study was with normally distributed random effect coefficients in the scale parameter of the detection function. Further simulations are needed to investigate how the P_r estimator performs in the case where these coefficients have a non-normal distribution, e.g. positively skewed. We also conclude that the $(1/P_r)$ estimator produces positively biased abundance estimates where the amount of bias depends on the amount of heterogeneity in the detection probabilities.

Chapter 7

Final discussion

7.1   General discussion

In this thesis we investigated four different methods of incorporating random effects into models for analysing distance sampling data. Using random effects is a new technique in the context of distance sampling analysis methods where covariate models for both the detection function and abundance have, until recently, generally been limited to fixed effects. For each method we presented the likelihood and methods for obtaining estimates of abundance in the covered area.

Firstly, we extended the two-stage approach, described by Buckland et al. (2009) for fixed effect models, by incorporating random effects in the second-stage count model (chapter 2). We also presented two integrated likelihood approaches for analysing distance sampling data which combine the two analysis stages and allow estimation of parameters from both stages simultaneously. For one formulation we extended the approach of Royle et al. (2004) by modelling heterogeneity in detection probabilities

and including a random effect for site in the abundance model (chapter 3). This approach uses the unconditional likelihood of observed distances and requires observed distances to be in intervals (rather than exact). Our second integrated likelihood approach uses the conditional probability density function of observed distances, and is applicable to both exact and interval distance data (chapter 4). Again, the random effect is incorporated in the count model to account for correlated measurements at the same sites. Furthermore, we developed estimators of abundance in the covered area that incorporate random effects in the detection function (chapter 6).

In addition, we proposed a novel Bayesian approach to analysing distance sampling data which uses a random walk single-update Metropolis-Hastings algorithm for updating model parameters and an RJMCMC algorithm for incorporating model uncertainty (chapter 4). Lastly, we proposed using hierarchical centering as a novel technique for improving model mixing and hence facilitating an RJMCMC algorithm for mixed effect models (chapter 5). These proposed methods are discussed further in the following.

7.1.1 Relaxing the assumption of independent counts for covariate models

Using a model-based approach for scaling up from encounter rate or density in the covered area to encounter rate or density in the study area, relaxes the assumption of random placement of samplers from conventional distance sampling (CDS) methods (see assumption II.A from chapter 1 on page 2 or Buckland et al., 2001 for details) and data arising from surveys conducted from platforms of opportunity may be used

(Hedley and Buckland, 2004). Similar to the methods described by Hedley and Buckland (2004) and Buckland et al. (2009), our covariate models from chapters 2, 3 and 4 may be extended to include smooth functions for continuous covariates, e.g. by fitting regression splines using the B-spline basis.

However, when using a covariate model without random effects (e.g. Hedley and Buckland, 2004 or Buckland et al., 2009) each count at a line/point is considered independent. Conceptually this means that each count is considered to contribute the same amount of new information to the analysis. In the case that counts are correlated, this is no longer the case. For both our case studies, counts were correlated due to repeated counts at the same sites. Repeated counts from the same site were likely to be more similar to each other than a count from a different site and, hence, contributed less new information to the analysis. By incorporating random effects into the covariate models, the similarities between counts at the same sites are modelled with the random effect coefficients. Hence, our methods relaxed the assumption of independent counts.

While mixed effect models can accommodate this type of positive correlations, they require specifying a grouping structure. In the case of a single random effect this involves defining the groups within which we allow counts to be correlated while still assuming inter-group independence. For our case study, this grouping structure consisted of the different sites. In other cases, it may be more difficult to specify grouping structures, e.g. if surveyed lines were broken up into small segments and counts from each segment considered a new observation. Here, we expect the correlations between counts at segments of the same line to decay with increasing distance along the line. In this case, it may be more appropriate to model the correlation structure using

generalised estimating equations (GEEs) (e.g. Halekoh et al., 2006; Peterson et al., 2011).

However, Bolker et al. (2008) recommended the use of GLMMs as, compared to a GLM, the inclusion of random effects in a GLMM allows biologists to generalise their conclusions to new times, places and species or, as in our case studies, to field buffers in general and not just those from the surveyed sites. This, however, is conditional on that the defined correlation structure is reasonable.

7.1.2   Covariate models for designed distance sampling experiments

Besides relaxing the assumptions of random survey design and independent counts mentioned above (section 7.1.1), the covariate models presented in chapters 2, 3, and 4 allow identification of relationships between abundance or density and covariates. This can be of particular interest for designed distance sampling experiments such as our case studies 1 and 2. For both these studies, the parameter of interest was the *type* covariate (with two levels: buffered treatment or unbuffered control fields) which allowed us to establish that the treatment applied to the fields had the desired effect. For both studies, the coefficient for treated fields was positive, indicating an increase in bird numbers on treated compared to untreated fields.

Designed experimental studies in general have an advantage over purely observational studies in that they allow inference on cause and effect of a treatment. The difference between observational and designed experimental studies in ecology lies in that for the latter, one part of the environment that may have an effect on numbers of the

species of interest is actively altered at the treated study plot(s) while all other parameters that may have an effect on animal density or detectability preferably remain the same in comparison to the control plot(s). In this manner, designed experiments allow attribution of any potential change in animal abundances between treated and control plots directly to the treatment, while purely observational studies do not. For our case studies, this was accomplished by pairing up each survey point from a treated field with a survey point from a nearby control field and conducting repeat surveys concurrently at both points to ensure similar observation conditions. The surveys were repeated at each site to avoid false inference due to random variation. The use of mixed effect models where site was the grouping factor for the random effect allowed us to use the individual counts from repeat visits to the points as the response variable. This made our methods very suitable for these large scale studies which are possibly unique due to their scale (over 400 sites with repeat surveys each year at each site).

Overall, we expect designed distance sampling experiments to become widely used for assessing effectiveness of conservation measures, and for environmental impact studies. One could argue that it might be sufficient to model raw counts rather than counts that were adjusted for imperfect detection. However, we demonstrated how model parameters may change depending on the detection model used (Table 3.2) and that model selection for both the detection and the abundance or count models is an important part of inference (chapters 3 and 4). As described above, the inclusion of site random effects accommodates correlations in multiple counts from a single site and allows inference to be extended to a wider area for which the sites are a representative sample, thus strengthening the ability of wildlife and natural resource

managers to evaluate the implications of changes in the environment.

7.1.3 Integrated likelihood methods for distance sampling data

Integrated modelling is becoming increasingly popular for ecologists (e.g. Besbeas et al., 2002; McCrea et al., 2010). It refers to simultaneous modelling of data from different sources, which may have been collected independently from each other, with the aim to improve the respective models. Within the distance sampling framework, integrated approaches have been proposed e.g. by Royle et al. (2004) and Johnson et al. (2010) or, using Bayesian methods, e.g. by Eguchi and Gerrodette (2009) and Moore and Barlow (2011). For integrated modelling of distance sampling data, however, the data sources to which the likelihood components of the integrated likelihood pertain (i.e. distances to the detections and number of detections at the line/point) are not independent but are collected simultaneously.

The motivation for developing our integrated likelihood methods arose from the shortcoming of the two-stage approach (Buckland et al., 2009) which requires using non-parametric bootstrapping to allow uncertainty from the first-stage detection function to propagate into the second-stage count model. For the two-stage approach, analytical standard errors for count model parameters may be artificially small. This was evident for our case study 1 in chapter 3, where these were particularly small for the *state* covariate in comparison to the bootstrap standard errors (or compared to the equivalent analytical standard errors from the integrated likelihood approach) (Table 3.3).

Underestimated standard errors may even lead to retaining the wrong covariates in the final model. Evidence for this was seen in the two-stage analysis of our case study 2 in chapter 4. Here, *year* was part of the best count model for the original data and both coefficients (both 0.17, Table 4.4) significant at the 0.001 level. Analytical standard errors were again artificially small at 0.049 and 0.050 (not shown) for years 2007 and 2008, respectively. The equivalent bootstrap standard errors were 0.13 and 0.11 and 95% confidence intervals obtained from the bootstrap included zero as a plausible value for these coefficients (Table 4.4). This issue may generally be avoided using the integrated likelihood approach where all parameters are estimated simultaneously.

In addition to improving analytical standard errors for count model parameters, estimating all parameters simultaneously may affect the parameter estimates themselves. We demonstrated this for our case study 1 in chapter 3 where detection function parameter estimates for *state* changed after including *state* in the count model (Table 3.2).

These differences may be caused by the difference in underlying concepts for the integrated vs. the two-stage approach. For the integrated approach, we assume that the observation process and the patterns by which animals distribute themselves in the study area influence each other. For the two-stage approach, these are considered separate processes. We argue that the integrated approach is the more realistic concept.

However, our integrated likelihood methods are also advantageous for our Bayesian approach for analysing distance sampling data, as they allow us to define one joint posterior distribution encompassing the detection and count models.

7.1.4   Bayesian analysis of distance sampling data

The novelties of our proposed Bayesian approach for analysing distance sampling data are the use of a Metropolis-Hasting (MH) algorithm for updating model parameters and the use of an RJMCMC algorithm for updating models. The few Bayesian distance sampling studies that exist to date, have used the Gibbs sampler for updating model parameters and either ignored model uncertainty (e.g. Eguchi and Gerrodette, 2009; Zhang, 2011) or used the Deviance Information Criterion (DIC) (Moore and Barlow, 2011). The latter study, however, conducted model selection separately for the different components of the analysis (detection, density and group size models). Gimenez et al. (2009) describe RJMCMC methods in their paper, but do not apply them in their line transect case study.

Our integrated likelihood methods allow all components of the analysis to be modelled simultaneously. The use of the MH updating allows easy implementation of different detection functions which may include adjustment terms, covariates and/or stratification. This variety of detection function models may be difficult to implement when using the Gibbs sampler, as this method requires sampling directly from a joint posterior distribution and this may be difficult to obtain when using non-standard distributions. We demonstrated in chapters 4 and 5 how RJMCMC may be used to move between different key functions with different covariate combinations for the detection model and between different covariate models for the Poisson model. In this fashion, our methods using RJMCMC provide a very efficient solution for exploring the model space including a large number of contending models. In comparison, using the DIC for model selection requires setting up and completing a separate MCMC chain for each contending model.

In contrast to maximum likelihood methods, our Bayesian approach provided additional benefits for our integrated likelihood methods. Firstly, for some studies, it might be challenging to find the maximum likelihood estimates for all parameters in one step, in particular if the number of observations is large and the models are rich in parameters. Using maximum likelihood methods, the random effect is integrated out which might be analytically challenging. For the integrated likelihood methods, functions such as *glmer* from the *lme4* package in R may not be used as they generally only deal with the count model component of the likelihood. Using a hierarchical model set up as in our Bayesian approach from chapter 4, the random effect coefficients are included in the model specification and updated during each iteration, which offers a straightforward technique to explore the parameter space.

Secondly, by combining the likelihood components, the number of possible models increases substantially making model selection quite elaborate when trying to fit each contending model separately to the data using maximum likelihood methods. Our RJMCMC methods allow us to explore the model space of all contending models simultaneously. It does, however, require the definition of proposal distributions and bijective functions that allowed easy jumps between models.

We found that the most problematic part of the RJMCMC algorithm for our hierarchical models was setting it up in such a manner that the chain indeed moved freely within the model space. The problems we encountered with our case study 1 consisted mostly in that the random effects coefficients absorbed the effect of the *state* covariate and prevented the inclusion of this covariate into the model as its effect was already accounted for. Model probabilities and summary statistics of parameters

would have falsely led us to the conclusion that *state* did not have an effect. In chapter 5, we proposed using hierarchical centering as a novel technique for solving such mixing problems. Here, the mixed model is reparameterised: the generally assumed zero-mean of the random effect is replaced with a model incorporating the intercept and one or more covariates from the Poisson model. In this formulation, the random effects coefficients are supposed to absorb any potential effects of the covariates included in the centering, and models with these covariates are favoured over those without.

7.1.5   A new method for fitting flexible detection functions

In chapter 6, we proposed a new method for fitting flexible detection functions which incorporate random effects in the model for the scale parameter of the half-normal key function. We tested this function using a simulation study and compared the performance of the new estimators to CDS estimators applied to the same data.

The simulation sets with heterogeneity consisted of 4 x 1000 simulations with two different amounts of heterogeneity which was introduced by modelling the scale parameter of the half-normal detection function using a common intercept and normally distributed random effects (i.e. in the same manner as the proposed detection function models the random effect). The amount of heterogeneity was determined by the size of the random effects standard deviation. For the CDS estimators we included a model selection routine to ensure that a sufficiently flexible detection function would be fitted to satisfy the pooling robustness criterion.

Average negative bias in abundance estimates using the CDS estimators was <2%

for the simulations with the smaller amount of heterogeneity but increased to up to 3.46% for those with the larger amount of heterogeneity. It is expected that, with further increase in heterogeneity, average negative bias will increase even further. In contrast, one of the new estimators with the random effects in the detection function (the P_r estimator) performed consistently well with an average bias of <1% regardless of the amount of heterogeneity in the data. This estimator also delivered the best coverage rates that were either appropriate or just too high. For those simulations where coverage rates were too low using the CDS estimators, the data needed to be truncated substantially to obtain high enough coverage rates. Hence, for these simulations, the P_r estimator delivered the best results without having to truncate the long tail of detections with low detection probabilities.

In addition, we conducted a simulation study without heterogeneity in detection probabilities. Here, the CDS estimators performed slightly better than our proposed P_r estimator. However, for the latter, average bias also remained less than 2% and coverage rates were appropriate for all four sets of simulations.

The second proposed estimator (the $(1/P_r)$ estimator) delivered positively biased abundance estimates where bias increased substantially with increase in the amount of heterogeneity. Coverage rates for this estimator were generally appropriate or just too high for the simulations without and with the smaller amount of heterogeneity but deteriorated for the larger amount of heterogeneity.

It is generally assumed that, due to pooling robustness, CDS methods provide unbiased estimates of abundance, despite potential heterogeneity in detection probabilities, as long as a sufficiently flexible detection function is fitted to the data (e.g. Buckland et al., 2001, 2004; Burnham et al., 1980, Thomas et al., unpublished manuscript).

Results from our simulation studies could not confirm the validity of this criterion without truncating the data substantially, as coverage rates were too small for two of the four sets of simulations with heterogeneity and bias in abundance estimates increased with increasing amount of heterogeneity. The performance of the P_r estimator on the other hand, was unaffected by the amount of heterogeneity in detection probabilities. One of the caveats of this study is, however, that we only introduced heterogeneity in the same manner as was assumed under the newly proposed methods, i.e. that the random effects coefficients are normally distributed. Further studies are needed to investigate how the newly proposed estimators perform in the case of non-normal distributions for the random effect.

Appendix A

Deriving the integrated likelihood for chapter 3

Royle et al. (2004) combined the multinomial likelihood (inside the large round brackets in eqn (A.1)) for observed counts y_{ik} in the k distance intervals at the ith point and unobserved counts $(N_i - y_{ik})$ with the Poisson likelihood (to the right of the $\times$ symbol in the same equation) for the N_i, the true number of animals at the ith point. Integrating this combined likelihood over all possible values for N_i gives (eqn (4) from Royle et al., 2004):

$$
L(\boldsymbol{\alpha}, \theta | \boldsymbol{y_i}) = \sum_{N_i=y_{i.}}^{\infty} \left(\frac{N_i!}{\left(\prod_k y_{ik}! \right) (N_i - y_{i.})!} \left[\prod_k \pi_k(\theta)^{y_{ik}} \right] \times \left[1 - \sum_k \pi_k(\theta) \right]^{N_i - y_{i.}} \right)
$$
$$
\times \frac{\exp(-\lambda_i(\alpha))\lambda_i(\alpha)^{N_i}}{N_i!}, \tag{A.1}
$$

where the $\pi_k(\theta)$ are the cell probabilities estimated using the unconditional likelihood of observed distances and θ the detection function parameters. The covariate model

for the expected number of N_i includes parameters $\boldsymbol{\alpha}$. This likelihood can be reduced to $\prod\limits_k Poisson[y_{ik}; \lambda_i(\alpha)\pi_k(\theta)]$ (Royle et al., 2004). Although these authors do not provide the proof for this, we prove it here. Eqn (A.1) can be rewritten as:

$$L(\boldsymbol{\alpha}, \theta | \boldsymbol{y_i}) = \sum_{N_i = y_{i.}}^{\infty} \left[\frac{\left(1 - \sum_k \pi_k(\theta)\right)^{N_i - y_{i.}}}{(N_i - y_{i.})!} \right] \times \lambda_i(\alpha)^{N_i} \exp(-\lambda_i(\alpha))$$
$$\times \frac{1}{\left(\prod_k y_{ik}!\right)} \left[\prod_k \pi_k(\theta)^{y_{ik}} \right] \tag{A.2}$$

Or as:

$$L(\boldsymbol{\alpha}, \theta | \boldsymbol{y_i}) = Term_1 \times Term_2 \tag{A.3}$$

where:

$$Term_1 = \sum_{N_i = y_{i.}}^{\infty} \left[\frac{\left(1 - \sum_k \pi_k(\theta)\right)^{N_i - y_{i.}}}{(N_i - y_{i.})!} \right] \times \lambda_i(\alpha)^{N_i} \exp(-\lambda_i(\alpha)) \tag{A.4}$$

and

$$Term_2 = \frac{1}{\left(\prod_k y_{ik}!\right)} \left[\prod_k \pi_k(\theta)^{y_{ik}} \right] \tag{A.5}$$

We then use the following relationship:

$$\sum_{x=0}^{\infty} \frac{a^x}{x!} = \exp(a), \tag{A.6}$$

where we set:

$$a = \lambda_i(\alpha) \left[1 - \sum_k \pi_k(\theta) \right] \tag{A.7}$$

$$x = N_i - y_{i.}. \tag{A.8}$$

Then, $Term_1$ can be rearranged in the following manner:

$$
\begin{aligned}
Term_1 \quad &= \sum_{N_i=y_{i.}}^{\infty} \left[\frac{\left(1 - \sum_k \pi_k(\theta)\right)^{N_i-y_{i.}}}{(N_i - y_{i.})!} \right] \times \lambda_i(\alpha)^{N_i} \exp(-\lambda_i(\alpha)) \\
&= \sum_{N_i-y_{i.}=0}^{\infty} \left[\frac{\left[\lambda_i(\alpha)\left(1 - \sum_k \pi_k(\theta)\right)\right]^{N_i-y_{i.}}}{(N_i - y_{i.})!} \right] \times \lambda_i(\alpha)^{y_{i.}} \exp(-\lambda_i(\alpha)) \\
&= \exp\left(\lambda_i(\alpha)[1 - \sum_k \pi_k(\theta)]\right) \times \lambda_i(\alpha)^{y_{i.}} \exp(-\lambda_i(\alpha)) \\
&= \exp\left(-\lambda_i(\alpha) \sum_k \pi_k(\theta)\right) \times \lambda_i(\alpha)^{y_{i.}}
\end{aligned}
$$

$$(A.9)$$

Also, $\lambda_i(\alpha)^{y_{i.}} = \lambda_i(\alpha)^{\sum_k y_{ik}} = \prod_k \lambda_i(\alpha)^{y_{ik}}$ and $\exp(-\lambda_i(\alpha) \sum_k \pi_k(\theta)) = \prod_k \exp(-\lambda_i(\alpha)\pi_k(\theta))$.

Hence, when recombining $Term_1$ and $Term_2$, we obtain:

$$
\begin{aligned}
L(\boldsymbol{\alpha}, \theta|\boldsymbol{y_i}) \quad &= \exp\left(-\lambda_i(\alpha) \sum_k \pi_k(\theta)\right) \times \lambda_i(\alpha)^{y_{i.}} \frac{1}{\left(\prod_k y_{ik}!\right)} \left[\prod_k \pi_k(\theta)^{y_{ik}}\right] \\
&= \prod_k \frac{\exp\left(-\lambda_i(\alpha)\pi_k(\theta)\right) \times (\lambda_i(\alpha)\pi_k(\theta))^{y_{ik}}}{y_{ik}!}
\end{aligned}
$$

$$(A.10)$$

Which is equivalent to $L(\boldsymbol{\alpha}, \theta|\boldsymbol{y_i}) = \prod_k Poisson[y_{ik}; \lambda_i(\alpha)\pi_k(\theta)]$.

Appendix B

Methods for building 95% log-normal confidence intervals around abundance estimates for chapter 6

B.1   Estimating the variance of encounter rate

For CDS methods, the average encounter rate of detections is generally expressed as the number of detections n divided by the total line length that was surveyed L. The variance of the encounter rate is given by:

$$\widehat{var}\left(\frac{n}{L}\right) = \frac{K}{L^2(K-1)} \sum_{k=1}^{K} l_k^2 \left(\frac{n_k}{l_k} - \frac{n}{L}\right)^2, \tag{B.1}$$

where K is the number of lines, the l_k refer to the length of the kth line (with $\sum_{k=1}^{K} l_k = L$) and the n_k refer to the number of detections of the kth line (with $\sum_{k=1}^{K} n_k = n$) (Fewster et al., 2009).

B.2   Estimating the variance of abundance

For detections of single objects the estimate of the variance of abundance is given by (Buckland et al., 2001):

$$\widehat{var}(\hat{N}) = \hat{N}^2 \times \left(\frac{\widehat{var}(\hat{P}_a)}{\hat{P}_a^2} + \frac{\widehat{var}(n/L)}{(n/L)^2} \right), \tag{B.2}$$

where $\hat{N}$ is the abundance estimate, n/L is the encounter rate and $\widehat{var}(n/L)$ the associated variance from above in this section.

When using the P_a estimator, we use eqns (6.6) to estimate the associated variance $\widehat{var}(\hat{P}_a)$ as shown in eqn (B.2). In the case that we model the detection function with a random effect, we use the P_r or the $(1/P_r)$ estimators instead which are given in sections 6.3.1 and 6.3.2, respectively. Then, $\widehat{var}(\hat{P}_r)/\hat{P}_r^2$ or $\widehat{var}(\widehat{1/P_r})/(\widehat{1/P_r})^2$ replace the equivalent term for $\hat{P}_a$ in eqn (B.2).

B.3   Log-based confidence intervals for abundance

To build 95% confidence intervals we use the log-based approach assuming that estimates of abundance are positively skewed and follow a log-normal distribution. The

limits of this interval are given by (Buckland et al., 2001):

$$\left(\hat{N}/C, \hat{N} \times C\right), \tag{B.3}$$

where

$$C = exp\left(z_\alpha \times \sqrt{v\hat{a}r(log_e\hat{N})}\right) \tag{B.4}$$

and

$$v\hat{a}r(log_e\hat{N}) = log_e\left(1 + \frac{v\hat{a}r(\hat{N})}{\hat{N}^2}\right). \tag{B.5}$$

This method is used for estimators with and without random effects in the detection function.

Appendix C

Obtaining approximations of derivatives via finite differences for chapter 6

C.1 Derivatives used for $\widehat{var}(\hat{P}_a)$

The likelihood given in eqn (6.3) is a function of the detection function parameters $\boldsymbol{\theta}$. Hence, we can obtain variance estimates for the elements in $\hat{\boldsymbol{\theta}}$ from the main diagonal of the inverse of the Hessian, i.e. $\boldsymbol{H}^{-1}$. The Hessian is calculated when using e.g. the *optim* or *nlm* function in R for maximising the likelihood.

The estimate of the average detection probability $\hat{P}_a$ is a function of the estimated detection function $\hat{g}(y)$ and hence depends on the model for $g(y)$ and the parameter estimates $\hat{\boldsymbol{\theta}}$ ($\hat{P}_a = \hat{\mu}/w = \int_0^w \hat{g}(y)dy/w$ from eqn (6.4)). Consequently, to obtain a variance estimate of $\hat{P}_a$, we require two components, the Hessian and the derivatives of the average detection probability $\hat{P}_a$ with respect to $\boldsymbol{\theta}$ evaluated at $\hat{\boldsymbol{\theta}}$ (see eqn (6.6)

page 109) (Borchers et al., 2002). To illustrate that these derivatives may be difficult to obtain analytically, we rewrote eqn (6.6) using $\hat{g}(y)$, as opposed to $\hat{\mu}$ from above.

$$\widehat{var}(\hat{P}_a) = \frac{1}{w^2} \left[\frac{\partial \int_0^w \hat{g}(y, \boldsymbol{\theta})dy}{\partial \boldsymbol{\theta}} \bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \right]^T \boldsymbol{H}^{-1} \left[\frac{\partial \int_0^w \hat{g}(y, \boldsymbol{\theta})dy}{\partial \boldsymbol{\theta}} \bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \right] \qquad (C.1)$$

However, a numerical approximation of the derivatives evaluated at the maximum likelihood estimates $\hat{\boldsymbol{\theta}}$ can be obtained via finite differences. For the ith element of $\boldsymbol{\theta}$ we use:

$$\frac{\partial \int_0^w \hat{g}(y, \theta_i, \hat{\boldsymbol{\theta}}_{-i})dy}{\partial \theta_i} \bigg|_{\theta_i=\hat{\theta}_i} = \frac{\int_0^w \hat{g}(y, \hat{\theta}_i + \delta\hat{\theta}_i, \hat{\boldsymbol{\theta}}_{-i})dy - \int_0^w \hat{g}(y, \hat{\theta}_i - \delta\hat{\theta}_i, \hat{\boldsymbol{\theta}}_{-i})dy}{2\delta\hat{\theta}_i}, \quad (C.2)$$

where $\delta\hat{\theta}_i$ represents a small fraction of $\hat{\theta}_i$ ($\delta = 0.0001$).

C.2   Derivatives used for $\widehat{var}(\hat{P}_r)$

To obtain an approximation for the derivative in eqn (6.10) in section 6.3.1 is similar to eqn (C.2). However, we need to account for the fact that the P_r estimator involves a simulation, where $E = 10000$ samples are drawn for the random effects coefficients b_e from $N(0, \hat{\sigma}_b^2)$. These b_es are incorporated in the model for the scale parameter of the detection function (eqn (6.7)) and a derivative calculated for each of E iterations using finite differences. The derivative for the ith element of $\boldsymbol{\theta}$ is the average of these derivatives evaluated at $\hat{\theta}_i$ of the E samples:

$$\frac{\partial \int_0^w \hat{g}(y, \theta_i, \hat{\boldsymbol{\theta}}_{-i})dy}{\partial \theta_i} \bigg|_{\theta_i=\hat{\theta}_i} = \frac{1}{E} \sum_{e=1}^{E} \frac{\int_0^w \hat{g}(y, b_e, \hat{\theta}_i + \delta\hat{\theta}_i, \hat{\boldsymbol{\theta}}_{-i})dy - \int_0^w \hat{g}(y, b_e, \hat{\theta}_i - \delta\hat{\theta}_i, \hat{\boldsymbol{\theta}}_{-i})dy}{2\delta\hat{\theta}_i}.$$
$$(C.3)$$

Using a half-normal detection function with a random effect, $\boldsymbol{\theta}$ contains two parameters, β_0 and σ_b. For the derivative with respect to β_0, the detection function for the eth sample is calculated using the scale parameter which is obtained using $\hat{\sigma}_e = (\hat{\beta}_0 + \delta\hat{\beta}_0) \times \exp(b_e)$ and $\hat{\sigma}_e = (\hat{\beta}_0 - \delta\hat{\beta}_0) \times \exp(b_e)$ for $\hat{\theta}_i + \delta\hat{\theta}_i$ and $\hat{\theta}_i - \delta\hat{\theta}_i$, respectively. As before, we set $\delta = 0.0001$.

For the derivative with respect to the random effects standard deviation, the samples for the random effects coefficients b_e are drawn from $N(0, (\hat{\sigma}_b + \delta\hat{\sigma}_b)^2)$ and $N(0, (\hat{\sigma}_b - \delta\hat{\sigma}_b)^2)$. To avoid extra variability due to random sampling, we set the seed of the random number generator to the same constant using the *set.seed* function in R before drawing the E samples from each of the respective distributions. Using the seed, the value for the variance estimate is approximately the same regardless of the seed. Without the seed, we found that the value for the variance estimate remains highly variable despite the large number of samples.

C.3   Derivatives for $\widehat{var}(1/\widehat{P_r})$

Using the estimator for the reciprocal of the average detection probability with a random effect in the detection function, $(\widehat{1/P_r})$ from section 6.3.2 involves simulating E samples from the random effects distribution as in the previous section. Again, the derivative of the ith element of $\boldsymbol{\theta}$ is given by the average of the derivatives evaluated at $\hat{\theta}_i$ over the E samples:

$$\frac{\partial \frac{1}{\int_0^w \hat{g}(y,\theta_i,\hat{\boldsymbol{\theta}}_{-i})dy}}{\partial\theta_i}\bigg|_{\theta_i=\hat{\theta}_i} = \frac{1}{E}\sum_{e=1}^{E} \frac{\frac{1}{\int_0^w \hat{g}(y,b_e,\hat{\theta}_i+\delta\hat{\theta}_i,\hat{\boldsymbol{\theta}}_{-i})dy} - \frac{1}{\int_0^w \hat{g}(y,b_e,\hat{\theta}_i-\delta\hat{\theta}_i,\hat{\boldsymbol{\theta}}_{-i})dy}}{2\delta\hat{\theta}_i}. \qquad \text{(C.4)}$$

The remainder is equivalent to section C.2.

Bibliography

Akaike, H. (1979). A Bayesian extension of the minimum AIC procedure of autoregressive model fitting. *Biometrika* **66 (2),** 237–242.

Al-Awadhi, F., Hurnb, M., and Jennison, C. (2004). Improving the acceptance rate of reversible jump MCMC proposals. *Statistics & Probability Letters* **69,** 189–198.

Archer, F. I., Henry, A. E., and Ballance, L. T. (2008). Stenella abundance research line transect and ecosystem (STAR-LITE) 2007 cruise report. Technical report, U.S. Department of Commerce. NOAA Technical Memorandum NMFS, NOAA-TM-NMFS-SWFSC-433, 19 p.

Bates, D. (2009a). Adaptive Gauss-hermite quadrature for generalized linear or nonlinear mixed models. R package version 0.999375-31. Technical report, http//lme4.r-forge.r-project.org/.

Bates, D. (2009b). Computational methods for mixed models. R package version 0.999375-31. Technical report, http//lme4.r-forge.r-project.org/.

Besbeas, P., Freeman, S. N., Morgan, B. J. T., and Catchpole, E. A. (2002). Integrating mark-recapture-recovery and census data to estimate animal abundance and demographic parameters. *Biometrics* **58,** 540–547.

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., and White, J.-S. S. (2008). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology and Evolution* **24 (3),** 127–135.

Borchers, D. L., Buckland, S. T., Goedhart, P. W., Clarke, E. D., and Hedley, S. L. (1998). Horvitz-Thompson estimators for line transect surveys. *Biometrics* **54,** 1221–1237.

Borchers, D. L., Buckland, S. T., and Zucchini, W. (2002). *Estimating Animal Abundance: Closed Populations.* Springer-Verlag.

Borchers, D. L., Marques, T. A., Gunlaugsson, T., and Jupp, P. (2010). Estimating distance sampling detection functions when distances are measured with errors. *Journal of Agricultural, Biological and Environmental Statistics* **15,** 346–361.

Borchers, D. L., Pike, D., Gunlaugsson, T., and Vikingson, G. (2009). Minke whale abundance estimation from the NASS 1987 and 2001 cue counting surveys taking account of distance estimation errors. *North Atlantic Marine Mammal Commission Publications* **7: North Atlantic Sightings Surveys (1987-2001),** 95–110.

Borchers, D. L., Zucchini, W., and Fewster, R. M. (1998). Mark-recapture models for line transect surveys. *Biometrics* **54,** 1207–1220.

Borchers, D. L. Laake, J. L., Southwell, C., and Paxton, C. G. M. (2006). Accommodating unmodeled heterogeneity in double-observer distance sampling surveys. *Biometrics* **62(2),** 372–378.

Brooks, S. P. and Gelman, A. (1998). Alternative methods for monitoring convergence

of iterative simulations. *Journal of Computational and Graphical Statistics* **7,** 434–455.

Brooks, S. P., Giudici, P., and Roberts, G. O. (2003). Efficient construction of reversible jump Markov chain Monte Carlo proposal distributions. *Journal of the Royal Statistical Society B* **65(1),** 3–55.

Browne, W. J. (2004). An illustration of the use of reparameterisation methods for improving MCMC efficiency in crossed random effect models. *Multilevel Modelling Newsletter* **16,** 13–25.

Browne, W. J., Steele, F., Golalizadeh, M., and Green, M. J. (2009). The use of simple reparameterizations to improve the efficiency of Markov chain Monte Carlo estimation for multilevel models with applications to discrete time survival models. *Journal of the Royal Statistical Society A* **172 (3),** 579–598.

Buckland, S. T. (2006). Point transect surveys for songbirds: robust methodologies. *The Auk* **123(2),** 345–345.

Buckland, S. T., Anderson, D. R., Burnham, K. P., Laake, J. L., Borchers, D. L., and Thomas, L. (2001). *Introduction to Distance Sampling.* Oxford University Press.

Buckland, S. T., Anderson, D. R., Burnham, K. P., Laake, J. L., Borchers, D. L., and Thomas, L. (2004). *Advanced Distance Sampling.* Oxford University Press.

Buckland, S. T., Borchers, D. L., Johnston, A., Henrys, P. A., and Marques, T. A. (2007). Line transect methods for plant surveys. *Biometrics* **63,** 989–998.

Buckland, S. T., Burnham, K. P., and Augustin, N. H. (1997). Model selection: an integral part of inference. *Biometrics* **52(2),** 603–618.

Buckland, S. T., Goudie, I. B. J., and Borchers, D. L. (2000). Wildlife population assessment: past developments and future directions. *Biometrics* **56,** 1–12.

Buckland, S. T., Laake, J. L., and Borchers, D. L. (2010). Double-observer line transect methods: levels of independence. *Biometrics* **66(1),** 169–177.

Buckland, S. T., Russell, R. E., Dickson, B. G., Saab, V. A., Gorman, D. G., and Block, W. M. (2009). Analysing designed experiments in distance sampling. *Journal of Agricultural, Biological and Environmental Statistics* **14,** 432–442.

Buckland, S. T., Summers, R. W., Borchers, D. L., and Thomas, L. (2006). Point transect sampling with traps or lures. *Journal of Applied Ecology* **43,** 377–384.

Burnham, K. P., Anderson, D. R., and Laake, J. L. (1980). Estimation of density from line transect sampling of biological populations. *Wildlife Monographs* **72,** 3–202.

Burt, M. L., Hedley, S. L., and Paxton, C. G. M. (2003). Spatial modelling of humpback whales using data from the 1995 and 2001 North Atlantic Sightings Surveys. Paper SC/11/AE/7 presented at the NAMMCO Scientific Committee Working Group on Abundance Estimates, March 2003. Available from: www.nammco.no.

Cañadas, A. and Hammond, P. S. (2006). Model-based abundance estimates for bottlenose dolphins off southern Spain: implications for conservation and management. *Journal of Cetacean Research and Management* **8(1),** 13–27.

Conover, R. R., Burger, L. W., and Lindner, E. T. (2011). Grassland bird nest ecology and survival in upland habitat buffers near wooded edges. *Wildlife Society Bulletin* **35(4),** 353–361.

Davison, A. C. (2003). *Statistical Models.* Cambridge University Press.

De Segura, A. G., Hammond, P. S., Cańadas, A., and Raga, J. A. (2007). Comparing cetacean abundance estimates derived from spatial models and design-based line transect methods. *Marine Ecology Progress Series* **329,** 289–299.

DiTraglia, F. J. (2007). Models of random wildlife movement with an application to distance sampling. Master's thesis, University of St Andrews.

Durban, J. and Elston, D. (2005). Mark-recapture with occasion and individual effects: abundance estimation through Bayesian model selection in a fixed dimensional parameter space. *Journal of Agricultural, Biological, and Environmental Statistics* **10,** 291–305.

Edwards, E. G. and Kleiber, P. M. (1989). Effects of nonrandomness on line transect estimates of dolphin school abundance. *Fishery Bulletin* **87(4),** 859–876.

Eguchi, T. and Gerrodette, T. (2009). A Bayesian approach to line-transect analysis for estimating abundance. *Ecological Modelling* **220,** 1620–1630.

Evans, K. O., Burger, L. W., Oedekoven, C. S., Smith, M. D., Riffell, S. K., Martin, J. A., and Buckland, S. T. (2013). Multi-region response to conservation buffers targeted for northern bobwhite. *The Journal of Wildlife Management.* In press.

Fewster, R., Southwell, C., Borchers, D., Buckland, S., and Pople, A. (2008). The influence of animal mobility on the assumption of uniform distances in aerial line-transect surveys. *Wildlife Research* **35,** 275–288.

Fewster, R. M., Buckland, Stephen T. Burnham, K. P., Borchers, David, L., Peter E. Jupp, P. E., Laake, J. L., and Thomas, L. (2009). Estimating the encounter rate variance in distance sampling. *Biometrics* **65,** 225–236.

Forster, J. J., Gill, R. C., and Overstall, A. M. (2012). Reversible jump methods for generalised linear models and generalised linear mixed models. *Statistics and Computing* **22 (1),** 107–120.

Gelfand, A. E., Sahu, S. K., and Carlin, B. P. (1995). Efficient parametrisations for normal linear mixed models. *Biometrika* **82 (3),** 479–488.

Gelman, A., Roberts, G. O., and Gilks, W. R. (1996). *Bayesian statistics*, chapter Efficient Metropolis jumping rules, pages 599–608. Oxford University Press, Oxford.

Gimenez, O., Bonner, S. J., King, R., Parker, R. A., Brooks, S. P., Jamieson, L. E., Grosbois, V., Morgan, B. J., and Thomas, L. (2009). WinBUGS for population ecologists: Bayesian modeling using Markov Chain Monte Carlo Methods. In Thomson, D. L., Cooch, E. G., and Conroy, M. J., editors, *Modeling Demographic Processes In Marked Populations*, volume 3 of *Environmental and Ecological Statistics*, pages 883–915. Springer US.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82(4),** 711–732.

Green, P. J. and Mira, A. (2001). Delayed rejection in reversible jump Metropolis-Hastings. *Biometrika* **88(4),** 1035–1053.

Halekoh, U., Højsgaard, S., and Yan, J. (2006). The R Package geepack for Generalized Estimating Equations. *Journal of Statistical Software* **15 (2),** 1–11.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57(1),** 97–109.

Hedley, S. L. and Buckland, S. T. (2004). Spatial models for line transect sampling. *Journal of Agricultural, Biological and Environmental Statistics* **9,** 181–199.

Johnson, D. S., Laake, J. L., and Ver Hoef, J. M. (2010). A model-based approach for making ecological inference from distance sampling data. *Biometrics* **66,** 310–318.

Karunamuni, R. J. and Quinn, T. J. (1995). Bayesian estimation of animal abundance for line transect sampling. *Biometrics* **51,** 1325–1337.

Kéry, M., Royle, J. A., and Schmid, H. (2005). Modeling avian abundance from replicated counts using binomial mixture models. *Ecological Applications* **15(4),** 1450–1461.

King, R., Morgan, J., Gimenez, O., and Brooks, S. (2010). *Bayesian Analysis for Population Ecology.* Chapman & Hall/CRC.

Komárek, A. and Lesaffre, E. (2008). Generalized linear mixed model with a penalized Gaussian mixture as a random effects distribution. *Computational Statistics and Data Analysis* **52,** 3441–3458.

Laake, J. L., Collier, B. A., Morrison, M. L., and Wilkins, R. N. (2011). Point-based mark-recapture distance sampling. *Journal of Agricultural, Biological, and Environmental Statistics* **16(3),** 389–408.

Laake, J. L. and Skaug, H. Distance sampling with random scale detection function.

Lesaffre, E. and Spiessens, B. (2001). On the effect of the number of quadrature points in a logistic random-effects model: an example. *Applied Statistics* **50(3),** 325–335.

MacKay, D. J. C. (2003). *Information Theory, Inference, and Learning Algorithms.* Cambridge University Press.

Marcot, B. G., Holthausen, R. S., Raphael, M. G., Rowland, M. M., and Wisdom, M. J. (2001). Using Bayesian belief networks to evaluate fish and wildlife population viability under land management alternatives from an environmental impact statement. *Forest Ecology and Management* **153,** 29–42.

Marques, F. F. C. and Buckland, S. T. (2003). Incorporating covariates into standard line transect analyses. *Biometrics* **53,** 924–935.

Marques, F. F. C. and Buckland, S. T. (2004). *Advanced Distance Sampling*, chapter Covariate models for the detection function, pages 31–47. Oxford University Press.

Marques, T. A. (2004). Predicting and correcting bias caused by measurement error in line transect sampling using multiplicative error models. *Biometrics* **60,** 757–763.

Marques, T. A., Buckland, S. T., Bispo, R., and Howland, B. Accounting for animal density gradients using independent information in distance sampling surveys. *Statistical Methods & Applications* pages 1–14.

Marques, T. A., Buckland, S. T., Borchers, D. L., Tosh, D., and McDonald, R. A. (2010). Point transect sampling along linear features. *Biometrics* **66,** 1247–1255.

Marques, T. A., Thomas, L., Fancy, S. G., and Buckland, S. T. (2007). Improving estimates of bird densities using multiple covariate distance sampling. *The Auk* **124,** 1229–1243.

McCrea, R. S., Morgan, B. J. T., Giminez, O., Besbeas, P., Lebreton, J.-D., and

Bregnballe, T. (2010). Multi-site integrated population modelling. *Journal of Agricultural, Biological and Environmental Statistics* **15 (4),** 539–561.

McCulloch, C. E. and Searle, S. R. (2001). *Generalized, Linear, and Mixed Models.* John Wiley & Sons, Inc.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics* **21,** 1087–1091.

Miller, D. L. and Thomas, L. Mixture models for distance sampling detection functions. Unpublished manuscript.

Moore, J. E. and Barlow, J. (2011). Bayesian state-space model of fin whale abundance trends from a 1991-2008 time series of line-transect surveys in the California Current. *Journal of Applied Ecology* **48,** 1195–1205.

Newson, S. E., Rexstad, E. A., Baillie, S. R., Buckland, S. T., and Aebischer, N. J. (2010). Population change of avian predators and grey squirrels in England: is there evidence for an impact on avian prey populations? *Journal of Applied Ecology* **47,** 244–252.

Niemi, A. and Fernández, C. (2010). Bayesian spatial point process modeling of line transect data. *Journal of Agricultural, Biological, and Environmental Statistics* **15 (3),** 327–345.

Palacios, D. M., Herrera, J. C., Gerrodette, T., Garcia, C., Soler, G. A., Avila, I. C., Bessudo, Sandra Hernandez, E., Trujillo, F., Florez-Gonzalez, L., and Kerr, I. (2012). Cetacean distribution and relative abundance in Colombia's Pacific EEZ

from survey cruises and platforms of opportunity. *Journal of Cetacean Research and Managemnt* **12(1),** 45–60.

Papaspiliopoulos, O., Roberts, G. O., and Sköld, M. (2007). A general framework for the parametrization of hierarchical models. *Statistical Science* **22 (1),** 59–73.

Papathomas, M., Dellaportas, P., and Vasdekis, V. G. S. (2011). A novel reversible jump algorithm for generalized linear models. *Biometrika* **98 (1),** 231–236.

Peterson, I. K., Mackenzie, M. L., Rexstad, E., Wisz, M. S., and Fox, A. D. (2011). Comparing pre- and post-construction distributions of long-tailed ducks *Clangula hyemalis* in and around the Nysted offshore wind farm, Denmark: a quasi-designed experiment accounting for imperfect detection, local surface features and autocorrelation. Technical report, Centre for Research into Ecological and Environmental Modelling 2011-01.

Potts, J. M. (2011). *Estimating abundance of rare, small mammals: a case study of the key largo woodrat (Neotoma floridana smalli).* PhD thesis, University of St Andrews.

Potts, J. M., Buckland, S. T., Thomas, L., and Savage, A. (2012). Estimating abundance of cryptic but trappable animals using trapping point transects: a case study for Key Largo woodrats. *Methods in Ecology and Evolution* **3,** 695–703.

Rabe-Hesketh, S., Skrondal, A., and Pickels, A. (2002). Estimation of generalized linear mixed models. *The Stata Journal* **2(1),** 1–21.

Riddle, J. D. and Moorman, C. E. (2010). The importance of agriculture-dominated

landscapes and lack of field border effect for early-succession songbird nest success. *Avian Conservation and Ecology* **5(2),** 9.

Royle, J. A., Dawson, D. K., and Bates, S. (2004). Modelling abundance effects in distance sampling. *Ecology* **85(6),** 1591–1597.

Schmidt, J. H., Lindberg, M. S., Johnson, D. S., Conant, B., and King, J. (2009). Evidence of Alaskan trumpeter swan population growth using Bayesian hierarchical models. *The Journal of Wildlife Management* **73(5),** 720–727.

Spear, L., Nur, N., and Ainley, D. G. (1992). Estimating absolute densities of flying seabirds using analyses of relative movement. *The Auk* **109,** 385–389.

Spear, L. B. and Ainley, D. G. (1997a). Flight behaviour of seabirds in relation to wind direction and wing morphology. *Ibis* **139,** 221–233.

Spear, L. B. and Ainley, D. G. (1997b). Flight speed of seabirds in relation to wind speed and direction. *Ibis* **139,** 234–251.

Spiegelhalter, D., Thomas, A., Best, N., and Lunn, D. (2003). WinBUGS user manual. Technical report, MRC Biostatistics Unit, Institute of Public Health, Cambridge.

Strindberg, S. and Buckland, S. T. (2004). Zigzag survey designs in line transect sampling. *Journal of Agricultural, Biological, and Environmental Statistics* **9(4),** 443–461.

Thomas, L., Borchers, D. L., Buckland, S. T., and Hammond, I. E. Estimation of population size from distance sampling data with heterogeneous detection probabilities. Unpublished manuscript.

Thomas, L., Buckland, S. T., Rexstad, E. A., Laake, J. L., Strindberg, S., Hedley, S. L., Bishop, J. R. B., Marques, T. A., and Burnham, K. P. (2010). Distance software: design and analysis of distance sampling surveys for estimating population size. *Journal of Applied Ecology* **47,** 5–14.

Vines, S. K., Gilks, W. R., and Wild, P. (1995). Fitting multiple random effects models. Technical report, MRC Biostatistics Unit, Cambridge.

Yuan, Y. (2012). *Estimating anglerfish abundance from trawl surveys, and related problems.* PhD thesis, University of St Andrews.

Yuan, Y., Borchers, D. L., Buckland, S. T., Potts, J., and Thomas, L. Horvitz-Thompson-like estimators with random inclusion probabilities. Unpublished manuscript.

Zhang, S. (2011). On parametric estimation of population abundance for line transect sampling. *Environmental and Ecological Statistics* **18,** 79–92.

Zucchini, W., Borchers, D., Erdelmeier, M., Rexstad, E., and Bishop, J. (2007). WiSP 1.2.4. Technical report, Institut für Statistik und Ökonometrie, Geror-August-Universität Göttingen, Platz der Göttinger Seiben 5, Göttingen, Germany.