ABSTRACT

WEBSTER, RAYMOND ANTHONY. Spatial Modeling of Detection and Abundance from Count Surveys of Animal Populations. (Under the direction of Kenneth H. Pollock.)

When analyzing data from surveys of animal populations, it has been common in the past to ignore important factors such as variation in animal detection probabilities across space, and spatial dependence in animal density. We present a unified framework for modeling animal survey data collected at spatially replicated survey sites in the form of repeated counts, "removal" counts, or "capture" history counts, that simultaneously models spatial variation in density and variation in detection probabilities due to changes in covariates across the landscape. The models have a complex hierarchical structure that makes them suited to Bayesian analysis using Markov chain Monte Carlo (MCMC) algorithms. To ensure that these algorithms are computationally efficient, we use conditional autogressive (CAR) models for modeling spatial dependence.

We apply our models to two examples of animal survey data. In the first, an intensive repeated count survey of juvenile Coho Salmon in McGarvey Creek, Northern California, we detected moderate spatial dependence in density, and models which account for spatial dependence produced more precise predictions at unsurveyed habitat units, and thus more precise estimates of total stream abundance, than models which assumed spatial independence. Through a small simulation study, we show that ignoring heterogeneity in detection probabilities can lead to significant underestimation of total abundance. However, inclusion of heterogeneity using a random effect in the detection component of the model can lead to problems in Bayesian MCMC modeling for typical survey designs, and for this reason we stress the importance of accounting for heterogeneity by incorporating covariates in modeling detection probability.

In our second example, we consider a large survey of birds in the Great Smoky Mountains National Park. We fit models to the three types of survey data, repeated counts, "removal" counts, and "capture" history counts. Our methods lead to maps of predicted relative density which are an improvement over those that would follow from ignoring spatial dependence. Modeling shows that variation in detection probability can also affect inference, particularly when a species is relatively difficult to detect. Our work also highlights the importance of good survey design for bird species modeling. We point out that these types of bird survey data, particularly removal and capture-recapture counts (which require individual birds to be identified), are prone to errors in bird identification. Although we obtain similar results for all three types of survey data, which

implies that the effect of identification errors may be small, the consequences of such errors in the data requires further investigation.

Finally, we present parametric models for combined distance and capture-recapture survey data from both line and point transect surveys that allow for two types of animal movement: permanent avoidance or attraction to a transect, or temporary displacement of animals in the vicinity of a transect. The models have a simple form, with parameters that quantify the impact transects and observers have on local density. We combine these density models with logistic-linear models for detection probability using the likelihood framework of Borchers et al. (1998) for combined distance and capture-recapture data. This allows us to separately estimate the parameters of both the density and detection components of the model, which is not possible using the standard methods of distance sampling. Through a simulation study, we show that, provided sufficient animals are detected, the model parameters have little bias, and lead to improved estimates of density over a simple uniform density model, particularly for line transect surveys. Model selection by AIC generally chooses the correct density model. We apply our models to the Great Smoky Mountains bird survey data, and find some evidence of observer effects on local bird density.

SPATIAL MODELING OF DETECTION AND ABUNDANCE FROM COUNT SURVEYS OF ANIMAL POPULATIONS

BY

RAYMOND A. WEBSTER

A DISSERTATION SUBMITTED TO THE GRADUATE FACULTY OF

NORTH CAROLINA STATE UNIVERSITY

IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

STATISTICS

RALEIGH,NORTH CAROLINA

DECEMBER 8, 2005

APPROVED BY:

KENNETH H. POLLOCK	SUJIT K. GHOSH
CHAIR OF ADVISORY COMMITTEE	
CAVELL BROWNIE	KEVIN GROSS

Biography

I was born and raised in the city of Dunedin, on the south-eastern coast of New Zealand's South Island, the son of a Kiwi father of Scottish descent (Henry) and a Dutch mother (Johanna). I have one younger brother, Frank. While undertaking my undergraduate degree in statistics at the University of Otago from 1988 to 1991, I developed at strong interest in environmental issues. I was able to combine this interest with my studies during my Masters thesis (1992-93), when I examined statistical methods applied to marine species populations. During my studies at Otago, I became involved with the Otago Student Green Movement, a small campus environmental activist group, and became a keen runner, regularly going for long runs around the beautiful hills overlooking Dunedin. In 1996, I traveled for seven months in Central America, Mexico and (briefly) the USA, and have since returned for three other trips in Latin America, including a six-month bike tour of South America in 1999/2000 and a shorter bike tour in Mexico in 2003/2004.

From 1997-2002 I worked as a biometrician for Landcare Research, a government-owned environmental research company, first in Pamerston North, in New Zealand's North Island, and from 2000-2002, in Christchurch. Seeking to expand my opportunities for statistical research while at the same time wanting to live in an exotic new environment, I began my PhD at North Carolina State University under the direction of Ken Pollock in 2002.

I am also a keen hiker/trekker (or tramper, as we say in New Zealand) and as well as many tramping trips in the wilderness of New Zealand, I have done multi-day hikes in Australia, Costa Rica, Honduras, Venezuela, Peru, Chile and Argentina. Most recently, I have joined the excellent trips organized by the NCSU Outdoor Adventure program to the mountains of North Carolina. My trips to Latin American have also given me a reasonable grasp of the Spanish language, while at home I am a keen gardener, best known for my cultivation of unusual chile peppers.

Acknowledgements

My thanks go to: Richard Barker, for motivating me to come to the USA to undertake this PhD; Ken Pollock, a wonderfully supportive adviser and all round good bloke; Sujit Ghosh, Montserrat Fuentes, Dave Hankin and Ted Simons, for their advice and input into this work; my mother, Johanna, for her support and encouragement; and all the wonderful friends I've made in Raleigh over the last three and a half years who have made my time here so enjoyable, especially Aarthi, Anindita, Jun, Karen, Kristy, Lavanya, Rosa, Surabhi and Tom.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Count Surveys of Animals when Detection is Imperfect

1.1.1 Background

Animal populations are often surveyed by sampling plots within a region of interest and counting the total number of animals detected within each sampled plot. Plots may range in size from small marked quadrats for some insects or pools in a stream for riverine fish (e.g., Coho Salmon, Hankin and Reeves, 1988), to large swaths of land or sea covered in aerial surveys of large mammals (e.g., Florida manatee, Craig et al., 1997), and are generally selected by simple, systematic or stratified random sampling. For certain animal species, it may be more convenient to survey the population using point or transect counts. When counting at a fixed point, some detection device such as an observer or set of traps is located at the point, while with line transects, observers traverse a fixed transect, or a series of detection devices such as traps are located along the length of the transect. In both cases, the animals that are detected within a fixed time interval are counted. Examples include the North American Breeding Bird Survey (Royle et al., 2002) and brushtail possum monitoring in New Zealand (Forsyth et al., 2006).

Although there is less information in data collected from count surveys compared to capture-recapture methods for monitoring animal populations (see Section 1.2), they can be more suitable for animals which are difficult to mark, such as juvenile fish, or to capture, as is the case for many bird species (e.g., Lichstein et al., 2002). Also, without the need to capture, mark and recapture individual animals, count surveys will be relatively inexpensive.

In some surveys the goal is to make a complete count of the total number of animals within

each sampled plot or within a specified area surrounding a point or transect, and such plot totals will lead directly to estimates of abundance (population size) or density (animals per unit area) for the entire monitoring region (see Seber, 1982). In almost all situations, however, detection will be imperfect. That is, the probability that an individual animal will be seen during a count survey is less than 1. Some animals may be hidden from the observer by vegetation, terrain, or even by other animals (e.g., when animals form herds). When animals are detected by sound or movement (e.g., counts of certain birds) then those animals which are silent or do not move during the survey period will not be detected. The result of this imperfect detection is that not all animals will be observed in a single count of a plot, and the counts themselves will only provide an index of animal abundance. Although such indices have been widely used without careful study, there is an implicit assumption that density is linearly related to the index. However, detection probability will be a function of many factors, including environmental variables (e.g., climate, vegetation density) and diurnal, seasonal and observer effects, and failure to account for such heterogeneity in detection will make valid comparisons of count indices over time or space difficult (Johnson, 1995; Barker and Sauer, 1995). Also, for some indices, the relationship with abundance is intrinsically nonlinear (e.g., Forsyth et al., 2006). It is, therefore, important to obtain information about detection probability as part of an animal population survey. By estimating detection probability, we will be able to directly estimate density or abundance without having to make the additional assumptions required to interpet count-based indices of abundance correctly.

1.1.2 Repeated Counts and Removal Sampling at a Single Site

One approach to estimating detection probability is to make independent, repeated counts of the animals at a site (i.e., a plot, point or transect). First, let us suppose that the counts are made in quick succession, allowing us to reasonably assume that the population is closed, i.e., that there is no immigration, emigration, births or deaths during the survey period. Two other important assumptions are that no errors are made in the counting, e.g., an individual animal's species is correctly identified and no animals are counted more than once, and that the counting process does not affect behavior and thereby bias future counts. If we further assume that detection probability does not vary across the repeated counts, the counts for an individual site can be thought of as being a simple random sample from a binomial distribution with parameters N, the true site abundance, and p, the detection probability at that site. If we observe T independent counts, $y_1, \ldots, y_T$, then

our model is:

$$y_j|N,p \sim \text{Bin}(N,p),$$

for $j = 1,\ldots,T$. In this simple case we can estimate the two parameters using either method of moments estimators (MMEs), or maximum likelihood (Johnson and Kotz, 1969). The former are given by

$$\tilde{N} = \frac{\bar{y}^2}{\bar{y} - s^2}, \tag{1.1}$$

$$\tilde{p} = \frac{\bar{y}}{\tilde{N}} = 1 - \frac{s^2}{\bar{y}}$$

where $\bar{y} = \sum_{j=1}^{T} y_j/T$ and $s^2 = \sum_{j=1}^{T}(y_j - \bar{y})^2/(T-1)$ are the usual estimators of the mean and variance of the sample of counts, while the maximum likelihood estimators (MLEs) are found by maximizing

$$\begin{aligned} \mathcal{L}(\mathbf{Y}|N,p) &= \prod_{j=1}^{T} \frac{N!}{(N-y_j)!y_j!} p^{y_j}(1-p)^{N-y_j} \\ &= \frac{N!^T}{\prod_{j=1}^{T}(N-y_j)!y_j!} p^{y_\cdot}(1-p)^{TN-y_\cdot}, \end{aligned} \tag{1.2}$$

where $\mathbf{Y} = [y_1,\ldots,y_T]$ is our data vector, and $y_\cdot = \sum_{j=1}^{N} y_j$. As given in Johnson and Kotz (1969), this maximization can be reduced to solving

$$\sum_{k=1}^{\mathcal{M}-1} (\widehat{N} - k) f_k = -T \log\left(1 - \frac{\bar{y}}{\widehat{N}}\right)$$

$$\widehat{p} = \frac{\bar{y}}{\widehat{N}},$$

where $f_k = \sum_j I(y_j > k)$, with $I(.)$ being the indicator function taking value 1 if the argument is true and zero otherwise, and $\mathcal{M} = \max(y_1,\ldots,y_T)$.

An approximate expression for the asymptotic variance of $\widehat{N}$ is also given in Johnson and Kotz (1969), and by substituting different values for N, p, and T into this expression it becomes clear that n will not be precisely estimated unless p or T are very large. Asymptotically, $\tilde{N}$ is slightly less efficient than $\widehat{N}$. In fact, both the MMEs and the MLEs are known to be very unstable, with small changes in the data producing large shifts in the estimates of N and p. For the MMEs, instability occurs when s^2 is close to $\bar{y}$, and so the denominator of (1.1) is close to zero and N is likely to be

overestimated. If $\bar{y}/s^2 < 1$, then $\tilde{N} < 0$ and $\tilde{p} < 0$ and so the estimates are invalid. Olkin et al. (1981) showed that for MLEs, instability also occurs when $\bar{y}/s^2 < 1$.

Some authors have approached the instability as a modeling problem, as the binomial model does not allow for hetergeneity in p across the repeated counts or between individual animals. When such heterogeneity exists, the variance will be greater than that expected under the binomial model, and the condition $\bar{y}/s^2 < 1$ is more likely to be true. Routledge (1981) and others have examined ways of allowing for such overdispersion in the model by using two sets of repeated count surveys with differing degrees of efficiency to estimate the dispersion parameter. However, they concluded that not only are such estimates dependent on the dispersion parameter being the same in both sets of surveys, but they are extremely imprecise even when this assumption is true. Olkin et al. (1981) and Carroll and Lombard (1985) instead proposed modifications to the MMEs and MLEs to make them more stable, and their estimators appear to perform quite well in many situations. However, when p is small, even these estimators have a tendency to overestimate p and underestimate N (Casella, 1986).

In undertaking a repeated count survey of a plot, point or transect we are sampling with replacement. An alternative counting method is removal sampling, in which the animals seen on each occasion are removed from the population to be counted on subsequent occasions. Removal sampling may be more appropriate for sampling animal pest species, which are killed upon capture on each sampling occasion (e.g., Forsyth et al., 2006). Also, some riverine fish populations are monitored using electrofishing, in which an electric current is run through a part of a stream temporarily stunning some of the fish, and the biologist temporarily removes these fish from the stream until completion of the survey. Removal methods have long been part of the animal population modeling literature, starting with Zippen (1956). For estimation of N, removal methods are statistically equivalent to the trap response model of the capture-recapture literature (see Pollock et al., 1990, and Section 1.2 below), although biologically they are distinctly different. If we also assume a binomial model for the removal data, again denoted by $\mathbf{Y} = [y_1, \ldots, y_T]'$, our count model will be:

$$y_i | N, p \sim \text{Bin}(N - M_j, p),$$

where $M_j = \sum_{j<j'} y_{j'}$ is the number of animals removed prior to occasion j, with $M_1 = 0$. Also, let $M_\cdot = M_{T+1} - y_T$, the number of unique animals detected prior to the final occasion. The likelihood

for this model is given by

$$\begin{aligned}\mathcal{L}(\mathbf{Y}|N,p) &= \prod_{j=1}^{T}\frac{(N-M_j)!}{(N-M_j-y_j)!y_j!}p^{y_j}(1-p)^{N-M_j-y_j} \\ &= \frac{N!}{(N-M_{T+1})!\prod_{j=1}^{T}y_j!}p^{M_{T+1}}(1-p)^{TN-M_\cdot-M_{T+1}} \qquad (1.3)\end{aligned}$$

which is the same as that given by Zippen (1956) using a multinomial specification for the removal data, and the same as the component of the likelihood involving N for the trap response model of the capture-recapture literature (Section 1.2). Zippen (1956) provides equations to solve for finding the MLEs, and discusses the properties of these estimators. Two things are of particular interest. The first is that when $N < 200$, a relatively large proportion of the population must be captured for precise estimation of N, implying that either p must be high (so most animals are captured quickly) or the number of occasions T must be large. Also, in some cases the removal method will 'fail' in the sense that we cannot obtain valid estimates of the parameters. The condition for failure is given in Seber (1982) as

$$\sum_{j=1}^{T}(T+1-2j)y_j < 0\,. \qquad (1.4)$$

For example, it follows from (1.4) that counts must decrease on average over the T occasions for estimation of N and p to be valid. Seber (1982) notes that the probability of failure decreases with increasing N and increasing p. As with a sample of counts, in removal sampling we must also assume that p is the same for all animals, and we cannot relax the assumption that p is constant across the sampling occasions without imposing other constraints on the parameters (Pollock et al., 1990).

Comparing removal and repeated count sampling of an individual plot or point, we note that while we will never fail to find parameter estimates from repeated counts, when it succeeds, the removal sampling estimator is stable and will almost always produce more precise estimates of N than repeated counts for the same p and T. Although this makes removal sampling seem much more appealing statistically, for many animal species count sampling will be more appropriate for practical reasons. For example, capturing or killing animals for removal may be more costly than undertaking a count, and as we have already noted, animals such as birds may be difficult to capture. Also, some species may be adversely affected by the removal process, which may be undesirable for animal welfare or conservation reasons. For example, with fish monitoring in

streams, removal sampling using electrofishing may harm or kill smaller fish of some species.

1.1.3 Extension of Models to Spatially Replicated Sites

Many of the problems discussed above for data from a single site can be overcome by combining information across the multiple sampling sites. As noted by Royle (2004), we are rarely interested in precise estimation of N at a single sampling site, as the animals at each site represent only a small fraction of the population being surveyed. Instead, the focus of a survey of an animal species is more frequently on estimation of the abundance of animals across a region. One approach could be to compute an estimate of total abundance from estimates of abundance at a random sample of individual sites in the region of interest. These could be calculated using a stable count estimator such as those of Olkin et al. (1981) or Carroll and Lombard (1985) for count data, or the removal method (Zippen, 1956). However, animal populations are often sparsely distributed and many sites will have few or no animals present. In such cases, even the stable count estimators will perform poorly for some sites (Casella, 1986) and the removal method is likely to fail, especially if p is low. Instead, by combining the data from the sampled sites we can extend the site-specific binomial models described above to estimate regional level abundance or density from our count data, without experiencing the same stability problems and with the scope to fit models with more flexible assumptions than those for data from a single site.

In our work we develop further the modeling framework of Royle (2004) for repeated count data and Wyatt (2002) and Forsyth et al. (2006) for removal data from independent sampling sites, and Wyatt (2003) for spatially correlated sites. We begin in this section by reviewing models for independent sites, while in Chapter 2 we discuss modifications to these models to incorporate site-level covariates in the modeling of animal density and detection probability, and to model spatial dependence in animal density across the monitoring region.

The most general model we can consider with closed population repeated count and removal data is one in which abundance varies across sites, and detectability varies across sites and between the independent counts within a site. Although with count data we cannot allow for heterogeneity in detection probability between individual animals, by allowing p to vary between the sampling occasions within a site, we are in part addressing the concerns outlined in Routledge (1981) regarding the effect of heterogenteity in p on estimation.

Suppose we draw a sample of L sites, and make T_i independent counts of animals at site i. We now denote the abundance at site i by N_i and the detection probability at site i on occasion j by

p_{ij}, $i = 1, \ldots, L$ and $j = 1, \ldots, T_i$. Let $T = \max_i T_i$. For both repeated count and removal data, we observe a data matrix $\mathbf{Y}$ with ith row $\mathbf{Y}_i = [y_{i1}, \ldots, y_{iT}]$, where y_{ij} is the number of animals seen on occasion j at site i, and y_{ij} is an empty cell ('missing data' for computational purposes) for $j > T_i$.

By first modeling y_{ij} as

$$y_{ij}|N_i, p_{ij} \sim \text{Bin}(N_i, p_{ij})$$

we can generalize Equation (1.2) to write the full likelihood for repeated count data as

$$\mathcal{L}(\mathbf{Y}|\mathbf{N}, \boldsymbol{p}) = \prod_{i=1}^{L} \prod_{j=1}^{T_i} \frac{N_i!}{(N_i - y_{ij})! y_{ij}!} p_{ij}^{y_{ij}} (1 - p_{ij})^{N_i - y_{ij}}. \tag{1.5}$$

For removal data, our binomial model is

$$y_{ij}|N_i, p_{ij} \sim \text{Bin}(N_i - M_{ij}, p_{ij}),$$

and the generalization of the single-site removal likelihood, Equation (1.3), is given by

$$\mathcal{L}(\mathbf{Y}|\mathbf{N}, \boldsymbol{p}) = \prod_{i=1}^{L} \frac{N_i!}{(N_i - M_{i,T_i+1})! \prod_j y_{ij}!} \prod_{j=1}^{T_i} p_{ij}^{y_{ij}} (1 - p_{ij})^{N_i - M_{ij} - y_{ij}},$$

where $M_{ij} = \sum_{j<j'} y_{ij'}$, with $M_{i1} = 0$, $i = 1, \ldots, L$. Here $\mathbf{N} = [N_1, \ldots, N_L]'$ and $\boldsymbol{p}$ is a matrix of the same form as $\mathbf{Y}$ with rows $\boldsymbol{p}_i = [p_{i1}, \ldots, p_{iT_i}]$. In both cases we have $L + \sum_i T_i$ parameters for only $\sum_i T_i$ counts and clearly estimation is impossible. For repeated count data, Royle (2004) considered a simpler version of our model with p equal for all sites on all occasions, but he notes that even with this reduction to $L + 1$ parameters, past research suggests that estimation is likely to be highly unstable. This is particularly true since we are counting animals at multiple sites, and with the natural tendency of many species to cluster, data may be sparse, with many sites with few or no animals either present or detected.

The approach taken by a number of authors to reducing the parameter space is to consider the N_i and p_{ij} as being samples from some parametric distributions. When we have data from independent sites and we are not attempting to model covariates or spatial dependence, the N_i are typically given a Poisson distribution, or to account for likely overdispersion in the N_i due to the clustering of animal populations, a Poisson-gamma mixture. Thus in the latter case we have

$$N_i|\lambda_i \sim \text{Poisson}(\lambda_i), \tag{1.6}$$

$$\lambda_i | a_\lambda, b_\lambda \sim \text{Gamma}(a_\lambda, b_\lambda).$$

Note that unconditionally, N_i has a negative binomial distribution, with probability mass function

$$\wp(N_i | a_\lambda, b_\lambda) = \left(\begin{array}{c} N_i + a_\lambda - 1 \\ a_\lambda - 1 \end{array} \right) \left(\frac{1}{b_\lambda + 1} \right)^{N_i} \left(\frac{b_\lambda}{b_\lambda + 1} \right)^{a_\lambda}, N_i = 0, 1, \ldots \tag{1.7}$$

In our work, the symbol $\wp$ is used to denote the density or probability mass function of a random variable, in order to clearly distinguish it from detection probability, p. A Poisson-gamma or negative binomial model was considered by Wyatt (2002) and Dorazio et al. (2005) for removal data and by Royle (2004) for repeated counts, and it has also been proposed for count data with perfect detection in modeling disease rates across spatially replicated regions (see Banerjee et al., 2004).

In modeling detectability, most authors make the simplifying assumptions that either $p_{ij} = p$, that is, detectability is constant for all sites on all occasions (Royle, 2004), or that $p_{ij} = p_i$ and so only varies between sites (Wyatt, 2002; Dorazio et al., 2005). In the constant detectability case, maximum likelihood estimation can proceed by first integrating out the 'nuisance' parameters N_i. With the negative binomial model (1.7) for N_i, this gives a likelihood function for repeated count data of:

$$\mathcal{L}(\mathbf{Y} | a_\lambda, b_\lambda, p) = \prod_{i=1}^{L} \sum_{N_i = \mathcal{M}_i}^{\infty} \left\{ \frac{N_i!^{T_i} \wp(N_i | a_\lambda, b_\lambda)}{\prod_{j=1}^{T_i} (N_i - y_{ij})! y_{ij}!} p^{y_{\cdot i}} (1-p)^{T_i N_i - y_{i\cdot}} \right\},$$

where $y_{i\cdot} = \sum_j y_{ij}$ and $\mathcal{M}_i = \max(y_{i1}, \ldots, y_{iT_i})$.

Defining $M_{i\cdot} = M_{i,T_i+1} - y_{iT_i}$, for removal sampling we have

$$\mathcal{L}(\mathbf{Y} | a_\lambda, b_\lambda, p) = \prod_{i=1}^{L} \sum_{N_i = \mathcal{M}_i}^{\infty} \left\{ \frac{N_i! \wp(N_i | a_\lambda, b_\lambda)}{(N_i - M_{i,T_i+1})! \prod_j y_{ij}!} p^{M_{i,T_i+1}} (1-p)^{T_i N_i - M_{i\cdot} - M_{i,T_i+1}} \right\}.$$

It is generally not difficult to maximize this likelihood numerically, and Royle (2004) found that, for his examples, it was possible to restrict the integration of N_i to a summation with a large, finite upper bound without affecting the MLEs.

In the models where detectability is not constant, we can assume the p_{ij} are also sampled from a distribution. A natural choice is the beta distribution, and Wyatt (2002) used this to model $p_{ij} = p_i$. More generally, we can build a model for p_{ij} by using a hierarchy of beta distributions:

$$p_{ij} | \nu_i, b_{p1} \sim \text{Beta}\left(\nu_i b_{p1}, (1-\nu_i) b_{p1}\right),$$

$$\nu_i | a_p, b_p \sim \text{Beta}(a_p, b_p).$$

Note that we have parameterized the distribution of p_{ij} in terms of its mean, ν_i, and then modeled these site means as a sample from a beta distribution with different parameters. While it seems natural to develop models in this hierarchical framework, maximum likelihood estimation becomes increasingly unappealing due to the need to numerically evaluate the integrals that arise from this approach. Thus for this model, the likelihood for count data becomes

$$\mathcal{L}(\mathbf{Y}|a_\lambda, b_\lambda, b_{p1}, a_p, b_p) = \prod_{i=1}^{L} \sum_{N_i=M_i}^{\infty} \left\{ \prod_{j=1}^{T_i} \right. \tag{1.8}$$
$$\left. \int_0^1 \frac{N_i! \wp(N_i|a_\lambda, b_\lambda)}{(N_i - y_{ij})! y_{ij}!} p_{ij}^{y_{ij}} (1 - p_{ij})^{N_i - y_{ij}} \left\{ \int_0^1 \wp(p_{ij}|\nu_i, b_{p1}) \wp(\nu_i|a_p, b_p) d\nu_i \right\} dp_{ij} \right\},$$

while that for removal data has an equally complex form. For models of such hierarchical complexity, maximum likelihood is no longer a feasible method of parameter estimation, and we instead turn to Bayesian hierarchical modeling which provides a more natural, albeit computationally intensive, framework for these models. A brief introduction to Bayesian hierarchical modeling is provided in Section 1.4.

We now return to the focus of the modeling, the distribution of abundance. We note that if the N_i follow a negative binomial distribution of form (1.7), then the expectation of N_i is given by a_λ / b_λ. If the sites are a random sample of plots each of known area, A_i say, then this suggests the following possible estimator for overall abundance, which we denote by N:

$$\widehat{N} = \frac{A}{\sum_i A_i} \frac{\widehat{a}_\lambda}{\widehat{b}_\lambda},$$

where A is the total area of the region from which the plots are sampled. If the sampled sites are points or transects, then to estimate total abundance we must be able to define an area around each site beyond which we do not sample, for example a circle around a point or a strip surrounding a transect. Where there is no clear sampling region around a site, we are back in the situation where we are estimating an index of abundance or density, but unlike the indices based on the raw counts, indices estimated in this way do allow for varying detectability and so would be more reliable for monitoring population trends or for other comparisons.

Although the individual site abundances are often not of interest themselves, their distribution and its moments could be estimated using an empirical Bayes procedure as proposed by Royle

(2004). This involves plugging the estimates of the parameters into $P(N_i = k|\boldsymbol{\theta})$ expanded using Bayes' theorem, where $\boldsymbol{\theta}$ is the parameter vector for a given model. Because of the complexity of our models, we prefer a fully Bayesian approach, from which posterior distributions of the individual N_i and p_{ij} follow automatically (see Section 1.4).

In this section we have described methods based on a binomial model for estimating abundance at a site from a sequence of independent counts and from removal sampling data when detection is imperfect, and we have reviewed generalizations of the models to spatially replicated sites. For the single site case, estimation from a sample of counts is imprecise and is affected by the assumptions we need to make on detection probability. By shifting focus to estimation of total abundance or mean density across a monitoring region, the imprecision of estimates at the site level is no longer of such importance, and by combining information across multiple sites, we can model detectability in a much more flexible and realistic way.

Further improvements to the estimation can be made by modeling the relationship between density and covariates measured at each sampling site, and between detectability and covariates at sites and on sampling occasions within sites. Covariates can be incorporated into the density component by replacing the gamma component of the Poisson-gamma mixture with a lognormal model, for example,

$$\log(\lambda_i) = \mu_i + \phi_i$$

Here ϕ_i is a zero-mean Gaussian error term and μ_i represents the covariate model, for example, $\mu_i = \mathbf{X}_i'\boldsymbol{\alpha}$, where $\mathbf{X}_i$ is the vector of covariates measured on unit i and $\boldsymbol{\alpha}$ is the corresponding parameter vector. Similarly in the detection model we can employ a logistic-normal model in place of the beta,

$$\text{logit}(p_{ij}) = \nu_{ij} + \epsilon_i + \eta_{ij},$$

where ν_{ij} is a covariate function, and ϵ_i and η_{ij} are respectively between and within unit Gaussian errors. Such covariate models have been considered by Wyatt (2003), Forsyth et al. (2006) and others, and we discuss this approach further in Chapter 2. Further, we can model spatial correlation in animal density through the error term ϕ_i. We introduce possible spatial models in Section 1.5 and discuss their application in subsequent chapters. By making use of covariate and spatial information, we not only improve estimation at the sites which we have monitored, but we can also predict density at unmonitered sites. This can improve estimates of total abundance or mean density, and lead to the creation of maps showing how animal density varies spatially across the monitoring

region.

Although the modeling framework we have presented is quite flexible, count and removal data nevertheless generally yield less information than the individual capture histories generated in a capture-recapture monitoring study. For example, we are unable to model heterogeneity in detection (or capture) probabilities of individual animals or incorporate covariates measured on individual animals into our models. We are, therefore, particularly interested in how closed population abundance estimation from repeated count and removal data compares to estimates from capture-recapture studies of the same populations. Based on our understanding of the single-site models, we would also anticipate removal sampling to give more precise estimates than repeated count sampling. However, when sampling sparse populations we can expect data from many sites to individually meet the failure criterion for removal studies, Equation (1.4), and it is unclear what effect this will have on parameter estimates which are essentially averages across all sites. An introduction to closed population capture-recapture models is given in Section 1.2, and a comparison of repeated count, removal and capture-recapture methods for multiple site data is presented in Chapter 3.

1.2 Closed Population Capture-Recapture Models

1.2.1 Background

Although the repeated count and removal modeling of Section 1.1 provides a quite flexible framework for abundance estimation for closed animals populations surveyed at multiple sites, there are important limitations on this type of modeling. In particular, we cannot model heterogeneity in detection probabilities between individual animals since counts do not distinguish between individuals. Also, in practice it may be difficult to separate variability in detection from that due to abundance, particularly for very sparse data or when detection probabilities are low. For example, if we observe many site counts similar to $[0, 1, 1]$, in estimation it is difficult to distinguish whether such counts are due to high abundance with very low detectability, or vice versa, and for single site models at least, there is a tendency to overestimate p and underestimate N (Casella, 1986).

Capture-recapture studies, like repeated count and removal surveys of animal populations, involve collecting data at the same sites on several occasions. However, in these studies individual animals are tracked across the sampling occasions, thus providing more information on detectability than count surveys and overcoming many of the limitations of count data. Typically, on the first

sampling occasion a sample of animals at a site is 'captured' (although for some species, this may require accurate tracking of animals rather than physical capture), the animals are marked and the marks are recorded, and then the sample is released at the capture site. On the second occasion, another sample is drawn, newly captured animals are marked and their marks are recorded, previously captured animals are recorded as being recaptured, and the sample is then released. This process is repeated for each sampling occasion, and over the course of the study a capture-history is constructed for each animal caught at least once. For example, for a study with four occasions, an animal only captured on the final occasion will have history denoted by 0001, while one captured on the first and second occasions has history 1100.

There are important general assumptions we must make in capture-recapture modeling that are related to the assumptions of repeated count and removal data. First, although open population models exist and are widely used (see Pollock et al., 1990), for comparison with count models we only consider those models that satisfy the closure assumption, that is, that there is no immigration, emigration, births or deaths during the survey period. We must also assume that animals do not lose their marks and that all marks are correctly recorded. For a detailed discussion of these assumptions, see Otis et al. (1978). Further, we assume that all animals are captured independently on each occasion, which is also an implicit assumption of the binomial count models. The capture-recapture literature covers a set of closed population models, each with their own assumptions regarding whether detection (or capture) probability varies with time or occasion, behavioral response to capture, between individual animals, or combinations of these three factors. What follows is a brief review of these models, largely based on Otis et al. (1978) and the later references that build upon it.

1.2.2 Modeling Capture-Recapture Data at a Single Survey Site

We assume that observations are recorded on T sampling occasions as before. If the population size is N, the most general model we consider is that with detection probability p_{jk} for the kth individual observed on occasion j, $j = 1, \ldots, T$ and $k = 1, \ldots, N$. Our observed data consist of values x_{jk}, where $x_{jk} = 1$ if animal k is captured on occasion j, and 0 otherwise. Most generally we can think of these as coming from a Bernoulli distribution,

$$x_{jk}|p_{jk} \sim \text{Bern}(p_{jk}), \tag{1.9}$$

although as we shall see, for particular capture-recapture models, this general form for the data model can be simplified. Together with the assumption of independence of capture, we can use (1.9) to construct the likelihood for a particular model. Note that animals for which $x_{jk} = 0$ for all j are not observed during the course of the survey.

The simplest closed population capture-recapture model, denoted by M_0, assumes that detection probability, p, does not vary with time, in response to capture, nor with the individual animal, that is $p_{jk} = p$ for $j = 1, \ldots, T$ and $k = 1, \ldots, N$. As with other models for which there is no heterogeneity in detection between individuals, we can aggregate the data into counts of animals with each unique capture history. We denote the total number of animals with history ω by X_ω. For example, for a survey with four sampling occasions, X_{1100} is the number of animals that were seen on the first two occasions but not thereafter. It follows from (1.9) that these counts will have a multinomial distribution, and for model M_0, the likelihood is given by

$$\mathcal{L}(\mathbf{X}_\omega | N, p) = \frac{N!}{\prod_\omega X_\omega! (N - M_{T+1})!} p^{n.} (1-p)^{TN - n.}.$$

M_{T+1} denotes the number of unique animals detected during the entire survey, and $n. = \sum_j n_j$, where n_j is the count of animals captured on occasion j. Note that $n_j = y_j$ in the notation of repeated count modeling for a single site, and the component of this likelihood involving p is identical to the single site repeated count likelihood, Equation (1.2).

When detection probability varies across sampling occasions (model M_t), the likelihood has a similar form. Here we wish to estimate N and a vector of probabilities $\boldsymbol{p} = [p_1, \ldots, p_T]$.

$$\mathcal{L}(\mathbf{X}_\omega | N, \boldsymbol{p}) = \frac{N!}{\prod_\omega X_\omega! (N - M_{T+1})!} \prod_{j=1}^{T} p_j^{n_j} (1-p_j)^{N - n_j}.$$

A particularly relevant case for our work is the behavioral response model, M_b. In this model, the detection probability takes the value p for animals that have never been captured, and a different value, c say, for those previously captured. The likelihood is given by

$$\begin{aligned} \mathcal{L}(\mathbf{X}_\omega | N, p, c) \quad = \quad & \frac{N!}{\prod_\omega X_\omega! (N - M_{T+1})!} p^{M_{T+1}} (1-p)^{TN - M_{T+1} - M.} \\ & \times c^{m.} (1-c)^{M. - m.}. \end{aligned} \tag{1.10}$$

Here $m. = \sum_j^T m_j$, where m_j is the number of previously marked animals caught on occasion j, and

$M_\cdot = \sum_{j=1}^{T-1} M_j$, where M_j is the number of unique animals caught prior to occasion j, with $M_0 = 0$. It is clear from (1.10) that when there is a behavioral response to capture, the recaptures give no information about N, although we note that for some generalizations of this model, recaptures do contribute to estimation of N (Norris and Pollock, 1995). This model is, therefore, statistically equivalent to the single site removal model for estimation of abundance, and comparing (1.10) with (1.3), we can see that the component of the two likelihoods involving N are identical. As mentioned when we examined the removal model, it is difficult to extend this model to allow p and c to vary across occasions, model M_{tb}, without placing further constraints on the probability parameters (Pollock et al., 1990).

The remaining closed population models allow for heterogeneity in detection probabilities between individual animals. A model allowing for such heterogeneity is denoted by M_h, while indexes b and t are added to denote generalizations to account for behavior and occasions. Clearly a model which attempts to estimate a different p_{jk} for every animal on every occasion will be over-parameterized without additional assumptions, and the capture-recapture literature contains many attempts to deal with this problem. For model M_h, the jackknife estimators of Burnham and Overton (1978) and the moment estimator of Chao (1987) are robust to heterogeneity and have been quite widely used. Chao et al. (1992) also proposed a moment estimator for model M_{th}, while Otis et al. (1978) developed a 'generalized removal' estimator for model M_{bh}.

Other authors have tackled this problem by assuming that the p_{jk} form a random sample from some distribution. For example, for model M_h, which allows heterogeneity between individuals only, one possibility is to suppose that the p_k are drawn from a beta distribution, which has been used with some success by Dorazio and Royle (2003). Coull and Agresti (1999) instead suggested logistic-normal models of the kind we discussed for modeling p for count and removal data. Similar approaches, both parametric and non-parametric, have been suggested for models M_{bh} and M_{th}. As an alternative to continuous distributions, finite mixture models have been proposed for the distribution of probabilities for models M_h and its extensions, M_{bh}, M_{th}, M_{tbh}, (Norris and Pollock, 1995, 1996; Pledger, 2000). Using the notation of Pledger (2000), animals are assumed to come independently from group a with probability π_a, $a = 1, \ldots, A$. The logit of the capture probability for an animal from group a on occasion j with behavioral response b (newly captured or previously captured) is then expressed as a linear function of parameters for group, occasion, behavior and their interactions. A multinomial likelihood can be constructed as for the simpler models discussed above, with an integration over the group allocations.

Link (2003) notes that estimation of N for a closed population is essentially the same as estimating the probability that an animal is unobserved during the survey period. The models for heterogeneity in detection probability that we have reviewed attempt to do this by estimating the distribution of the p_{jk} from a model fitted to data which come from observed animals only. Link (2003) shows that different modeling approaches can fit the observed data equally well yet produce quite different estimates of the probability that an animal is unobserved, and hence quite different estimates of N. This is a particularly worrisome feature of these models, and for our work we seek alternative approaches to modeling heterogeneity. One such approach is to model heterogeneity in detection probability by making p_{jk} a function of covariates measured on the individual animals. Examples of such covariates include animal characteristics (weight, sex, age) or the distance of the animal from the observer. If we assume that covariates are the only source of capture heterogeneity in our model, we can apply the logistic regression models of Huggins (1989) and Alho (1990). These authors condition on the observed values of the covariates, but the conditional likelihood that results does not include N and it must be estimated using other methods (see also Pollock, 2002). We do not consider this conditional approach in our work as it does not fit within our hierarchical framework for modeling data from multiple sites. Instead we can make use of the full multinomial likelihood constructed by Borchers et al. (1998). Consider the case of a single covariate, x, which (following Borchers et al., 1998) we assume has a distribution with density $\wp(x|\boldsymbol{\theta}_x)$ dependent on parameters $\boldsymbol{\theta}_x$. For animal k, x has measurement x_k, and we write the capture probability as a parametric function of x, $p_{jk} = p_j(x_k|\boldsymbol{\theta}_p)$. Usually this function is also of linear logistic form, e.g., $\text{logit}(p_{jk}) = \alpha + \beta x_k$. Now denote the probability that an animal is seen at least once by $g(x|\theta_p)$. It follows from the original Bernoulli model, (1.9), that

$$g(x|\boldsymbol{\theta}_p) = 1 - \prod_{j=1}^{T} \left\{1 - p_j(x|\boldsymbol{\theta}_p)\right\}.$$

The full likelihood can now be written as

$$\begin{aligned}\mathcal{L}(\mathbf{Z}, \boldsymbol{x}|\boldsymbol{\theta}_x, \boldsymbol{\theta}_p) &= \frac{N!}{M_{T+1}!(N - M_{T+1})!}(1 - p_\cdot)^{N - M_{T+1}} \\ &\quad \times \prod_{k=1}^{M_{T+1}} \left\{ \prod_{j=1}^{T} p_{jk}^{z_{jk}} (1 - p_{jk})^{1 - z_{jk}} \right\} \wp(x_k),\end{aligned}$$

where for clarity we have suppressed the conditioning on the parameters on the right hand side.

Here $p.$ is the expected probability that an animal is observed at least once during the survey, and is given by

$$\begin{aligned} p. &= \int P(\text{animal detected}|x)\wp(x)dx \\ &= \int g(x)\wp(x)dx. \end{aligned}$$

This integral must be evaluated numerically which has the potential to cause computational difficulties in maximizing the likelihood. Pollock (2002) also points out that, depending on the parameterization of $\wp(x)$ and $g(x)$, there may be many parameters and consequent problems with identifiability and estimation. However, when it does work, this can be an extremely useful method for estimating N while allowing for heterogeneity in capture due to covariates.

In this section we have reviewed capture-recapture models for closed populations sampled at a single survey site. Our focus has been on the models and methods of which we will make use of in later chapters when we develop extensions of these models to account for capture-recapture data from spatially replicated sites.

1.2.3 Marking and Identification Errors in Animal Surveys

The assumption of no misidentification or loss of marks in capture-recapture is an important one. For example, if marks are lost, then recaptured animals will be incorrectly identified as new captures and population estimates will be positively biased (Pollock et al., 1990). Animal misidentification is particularly problematic with natural marks, which may be similar for different individuals or may change with time. Some methods of marking are more likely to lead to errors than others. In surveys of bird species, one approach is to divide an observation period of 10 minutes at each site into three smaller intervals of time, and a three-period capture history is recorded for each bird observed during the 10 minutes (Alldredge, 2004), e.g., a capture history of 101 if the bird was seen in the first and third interval but not the second. The 'marking' here is simply the observer attempting to track each bird over the course of the observation period, something which is particularly prone to error when birds move quickly or are identified by sound alone.

We may however, expect repeated count and removal surveys to be less affected by these types of errors as they do not require individuals to be distinguished across the multiple sampling occasions. It is possible, therefore, that when such marking errors are serious, count and removal survey data will lead to less biased parameter estimates than capture-recapture data. However,

repeated count surveys are themselves subject to other errors in the counting process – animals can be counted more than once on a single survey occasion, two individuals seen separately can be identified as the same animal (both similar to marking errors), or, particularly when density is high at a site, it may be difficult to make an accurate count of the observed animals.

Perhaps least prone to error will be presence-absence surveys (Section 1.3 below), as they simply require at least one member of the species to be correctly identified at a site. Thus data from the surveys which in theory yield the least information (presence-absence), will be the most reliable, while data from surveys which yield the most information (capture-recapture), may be the least reliable. It will be interesting, therefore, to examine the effect of errors on parameter estimates for the three types of animal surveys. In particular, we can compare site occupancy estimates for all four survey types, and individual detection and abundance estimation for repeated count, removal, and capture-recapture surveys when there is error in the marking or counting processes. While we note that measurement errors are likely to affect our modeling in Chapters 3 and 4, detailed analysis of their effects is left for future work.

1.3 Presence-Absence Surveys with Imperfect Detection

1.3.1 Background

The principle advantage that repeated count and removal surveys of animals have over closed population capture-recapture methods is their relatively low cost. Nevertheless, in large-scale species surveys or in surveys of animals that are expensive to find, even surveys based on counts can be very costly to undertake. Presence-absence surveys, on the other hand, can be much less expensive as they require the identification of no more than one animal of the species of interest to establish presence. The goal of a repeated count, removal or capture-recapture study is usually to estimate abundance or density, or map spatial variation in density across a region. With a presence-absence survey, we are instead trying to establish the spatial range of a species. For rare and endangered species, establishing presence at a location can be important for species management and protection, while for pest species, control operations may be more effective when they target areas of known species presence.

The result of a presence-absence survey is often a map showing sites at which the species is present. However, as with count data, detection of individual animals is almost always imperfect – an individual may be present, but remain undetected during the survey period. The consequence of

this for presence-absence surveys is that while detection implies presence of the species (assuming no misidentification), failure to detect a species at a site does not necessarily mean that the species is absent. Therefore, any method that aims to estimate occupancy probabilities or predict presence, must also account for the probability of detecting the species at a site.

Some authors have tried to deal with detection by attempting to ensure, through study design, that species detection probabilities are as close to 1 as possible. For example, in analysing point observations of bird species, Shriner (2001) excluded data beyond a certain distance of the observer and assumed that within that distance, detection of individual birds was perfect. Others such as Stauffer et al. (2002) recommend making sufficient repeated observations of a site so that species detection is almost certain if the species is present. Such approaches risk wasting information or resources with no guarantee that the issue of detection has been adequately dealt with.

Bayley and Peterson (2001) instead used an empirical Bayesian approach to estimate the probability that a site is occupied by a species when no animals were observed at that site. They began by estimating detection probabilities and modeling species abundance as a function of habitat covariates using independent field data. These models were then used to predict abundance at each site sampled in the current survey, and these predictions became empirical priors for estimating site occupancy probabilities using Bayes' Theorem. Their methods are appealing when the field data already exists, since less sampling effort will be required if we do not need to estimate detection probabilities during the current survey. However, when useful data do not exist, we must turn to survey designs and models which permit estimation of detectability from the current presence-absence survey data.

Of particular interest in the present context are presence-absence surveys in which the sampled sites are repeatedly surveyed for the presence of a species, since such survey designs can be thought of as a special case of the repeated count designs examined in Section 1.1. Geissler and Fuller (1986) proposed using these surveys to estimate the proportion of sites which are occupied as a useful statistical measure of the range of a species. Their ideas were developed into a cohesive presence-absence modeling framework by MacKenzie et al. (2002). In the remainder of this section we review this framework and examine its connections to repeated count modeling.

1.3.2 Estimating Site Occupancy

Previously we showed that with repeated count and removal data, we are able to estimate the probability of detecting an individual by making multiple independent observations at a site and

assuming a binomial model for the observed counts. Although estimators exist for data from a single site, more flexible and stable modeling can be undertaken using data from spatially replicated sites. For presence-absence sampling, we can again obtain information about detection probability (this time of the species) by repeatedly observing a site, but although it is possible to estimate abundance from single-site count data, in order to estimate site occupancy probabilities, we now require data from multiple sites.

In presence-absence surveys, rather than animal abundance or density, we are interested in estimating the proportion of sites occupied, or the mean site occupancy probability, as a measure of the status of a species. Using the notation of MacKenzie et al. (2002), we denote the probability that site i is occupied by $\psi_i = P(N_i > 0)$, where N_i is again the animal abundance at site i. As before, suppose we have a sample of L sites from a region, and site i is observed on T_i occasions. In presence-absence surveys, instead of recording a count y_{ij} of animals observed at site i on occasion j, we record $z_{ij} = I(y_{ij} > 0)$, with $I(.)$ being the indicator function taking value 1 if the argument is true and 0 otherwise. That is, z_{ij} takes the value 1 if the species was detected, and 0 otherwise. Now let ρ_{ij} be the probability of detecting the species at site i on occasion j. Then given that a site is occupied, that is, one or more individuals are present ($N_i > 0$), our observations follow a Bernoulli distribution,

$$z_{ij}|\rho_{ij}, (N_i > 0) \sim \text{Bern}(\rho_{ij}). \tag{1.11}$$

Let $\mathbf{Z}$ be the full matrix of observations with rows of unequal length $\mathbf{Z}_i = [z_{i1}, \ldots, z_{iT_i}]$. Based on the parameters $\boldsymbol{\rho}_i = [\rho_{i1}, \ldots, \rho_{iT_i}]$ and ψ_i we can construct the likelihood component for site i under the Bernoulli model (1.11) for occupied sites. As recognised by MacKenzie et al. (2002), who used an equivalent capture-recapture framework to derive the likelihood, this depends on whether the species has been detected or not over the entire survey. For a site at which at least one animal was seen, we have

$$\mathcal{L}_1(\mathbf{Z}_i|\boldsymbol{\rho}_i, \psi_i) = \psi_i \prod_{j=1}^{T_i} \rho_{ij}^{z_{ij}} (1 - \rho_{ij})^{1-z_{ij}}.$$

For sites with no animals observed, the species was either present but undetected or the species was absent, leading to a likelihood with two components,

$$\mathcal{L}_2(\mathbf{Z}_i|\boldsymbol{\rho}_i, \psi_i) = \psi_i \prod_{j=1}^{T_i} (1 - \rho_{ij}) + (1 - \psi_i).$$

If animals are detected at $L.$ sites, the full likelihood is given by

$$\mathcal{L}(\mathbf{Z}|\boldsymbol{\rho},\boldsymbol{\psi}) = \prod_{i=1}^{L.} \mathcal{L}_1(\mathbf{Z}_i|\boldsymbol{\rho}_i,\psi_i) \times \prod_{i=L.+1}^{L} \mathcal{L}_2(\mathbf{Z}_i|\boldsymbol{\rho}_i,\psi_i)$$

where $\boldsymbol{\rho}$ is the matrix with ith row given by $\boldsymbol{\rho}_i$. As it stands, this model is overparameterized and additional information or assumptions are required for parameter estimation. MacKenzie et al. (2002) put restrictions on the parameters. In particular, they assumed that occupancy probabilities were constant, $\psi_i = \psi$, across the sites, and species detection varies only across the sampling occasions, that is $\rho_{ij} = \rho_j$. This reduces the full likelihood to

$$\begin{aligned} \mathcal{L}(\mathbf{Z}|\boldsymbol{\rho},\psi) &= \left\{ \psi^{L.} \prod_{j=1}^{T_i} \rho_j^{z_{\cdot j}} (1-\rho_j)^{L.-z_{\cdot j}} \right\} \\ &\quad \times \left\{ \psi \prod_{j=1}^{T_i} (1-\rho_j) + (1-\psi) \right\}^{L-L.}, \end{aligned} \tag{1.12}$$

where in this case $\boldsymbol{\rho} = [\rho_1, \ldots, \rho_{T_{\max}}]$, with $T_{\max} = \max(T_1, \ldots, T_L)$, and $z_{\cdot j} = \sum_i z_{ij}$. Thus for this model the statistics $(z_{\cdot 1}, \ldots, z_{\cdot T_{\max}}, L.)$ are sufficient for the parameters, and maximization of (1.12) is generally straightforward. MacKenzie et al. (2002) suggest using a nonparametric bootstrap procedure for calculating standard errors for this model.

This model is very restrictive, however. In particular, there will always be heterogeneity in detection probability between sites due to spatial variation in abundance. One approach to allowing for heterogeneity is to model ψ_i and ρ_{ij} as functions of covariates, using a logistic link function as discussed in MacKenzie et al. (2002). A summary statistic of site occupancy could then be the average of the $\widehat{\psi}_i$. Another possibility is to assume the ψ_i and ρ_{ij} are samples from some parametric distribution, for example the beta distribution as we did for p_{ij} for count and removal models in Section 1.1, or to apply the finite mixture models proposed for capture-recapture modeling of heterogeneity in detection (Section 1.2, Pledger, 2000). We can also allow for both covariates and random effects using a logistic-normal. For example, for ψ_i we may have

$$\text{logit}(\psi_i) = \mu_i + \phi_i$$

where μ_i is the covariate model, e.g., $\mu_i = \mathbf{X}_i'\boldsymbol{\alpha}$, and ϕ_i is a zero-mean Gaussian error term. We can also model spatial dependence in occupancy through the ϕ_i term as we did for count data

(described in Section 1.5). Such models allow us to predict the occupancy probabilities of sites at which no animals were observed and at unsurveyed sites. The complexity of these models means a likelihood approach is again no longer feasible, but by specifying the models hierarchically, we can instead use Bayesian hierarchical modeling for parameter estimation (Section 1.4). However, because there is less information in presence-absence data than count data, we will almost certainly need to sample more sites and use more sampling occasions in order to get good parameter estimates from such complex models.

1.3.3 Count Modeling and Site Occupancy Estimation

Previously we considered the case examined by MacKenzie et al. (2002) where ψ is constant and $\rho_{ij} = \rho_j$, that is, species detection varies only across the sampling occasions. Another interesting model is where ψ is also constant but $\rho_{ij} = \rho_i$. Setting aside for now covariate models or assumptions about the distribution of parameters, with a ρ_i associated with each site, this model is more difficult to fit, particularly for sparse data. However, it is informative to examine this model as it provides a link to the repeated count models of Section 1.1. In that section we discussed the model for count data examined by Royle (2004) in which individual animal detection probability p is constant across all sites on all occasions. That author's approach was to place a distribution (such as the negative binomial) on the site abundances N_i and integrate N_i out of the likelihood. With count data, we are also able to estimate occupancy probabilities, and for the negative binomial model the proportion of sites occupied follows from Equation (1.7) as

$$\begin{aligned} \psi = P(N_i > 0) &= 1 - P(N_i = 0) \\ &= 1 - \left(\frac{b_\lambda}{b_\lambda + 1}\right)^{a_\lambda}. \end{aligned}$$

An estimate, $\widehat{\psi}$, will follow automatically from the MLEs for a_λ and b_λ. If a key goal of species monitoring is the estimation of ψ, the question arises, what do we gain by having count data rather than presence-absence data? We approach this by reconsidering the likelihood for the p constant model for count data. For illustration purposes, we do not place a distribution on N_i, but treat these as parameters to be estimated. Now, rather than writing the full likelihood as a version of Equation (1.5), we split it into two components, one for sites at which at least one animal was seen, and another for sites where no animals were seen (thus $y_{ij} = 0$, $j = 1, \ldots, T_i$). For each of the L.

sites at which the species was detected, the likelihood component is

$$\mathcal{L}_1(\mathbf{Y}_i|N_i, p, \psi) = \psi \frac{N_i!^{T_i}}{\prod_j (N_i - y_{ij})! y_{ij}!} p^{y_{i\cdot}} (1-p)^{T_i N_i - y_{i\cdot}},$$

while for sites with no animals detected, the likelihood component simplifies to

$$\mathcal{L}_2(\mathbf{Y}_i|N_i, p, \psi) = \psi(1-p)^{T_i N_i} + (1-\psi). \tag{1.13}$$

This leads to a full likelihood of

$$\begin{aligned} \mathcal{L}(\mathbf{Y}|\mathbf{N}, p, \psi) &= \prod_{i=1}^{L.} \mathcal{L}_1(\mathbf{Y}_i|N_i, p, \psi) \times \prod_{i=L.+1}^{L} \mathcal{L}_2(\mathbf{Y}_i|N_i, p, \psi) \\ &= \psi^{L.} \prod_{i=1}^{L.} \frac{N_i!^{T_i}}{\prod_j (N_i - y_{ij})! y_{ij}!} p^{y_{i\cdot}} (1-p)^{T_i N_i - y_{i\cdot}} \\ &\quad \times \prod_{i=L.+1}^{L} \left\{ \psi(1-p)^{T_i N_i} + (1-\psi) \right\}. \end{aligned} \tag{1.14}$$

Now let us return to modeling presence-absence data only. As pointed out by Royle and Nichols (2003), the model with constant species detection probability ρ across sites is unrealistic as it ignores the fact that the species is more likely to be detected at a site with more animals present. In fact, for constant individual detection probability, p, species detection is related to site abundance, N_i, by

$$\rho_i = 1 - (1-p)^{N_i}. \tag{1.15}$$

Therefore, the presence-absence model that corresponds to the count model with p constant and abundance varying across sites is the model with species detection, ρ_i, varying across sites. The likelihood for this presence-absence model is

$$\begin{aligned} \mathcal{L}(\mathbf{Z}|\boldsymbol{\rho}, \psi) &= \psi^{L.} \prod_{i=1}^{L.} \rho_i^{z_{i\cdot}} (1-\rho_i)^{T_i - z_{i\cdot}} \\ &\quad \times \prod_{i=L.+1}^{L} \left\{ \psi(1-\rho_i)^{T_i} + (1-\psi) \right\}, \end{aligned} \tag{1.16}$$

where $z_{i\cdot} = \sum_j z_{ij}$. We note that in both cases, information about detection probabilities and abundance only comes through the first component of the likelihoods, since data from the $L - L.$ sites at which nothing was detected tell us nothing about these parameters. We can see that for both

count and presence-absence data, the only information about ψ from the first component of the likelihoods is $L.$, the number of sites at which the species was observed. But from Equation (1.15), the second components of both (1.14) and (1.16) are mathematically identical, since

$$(1-p)^{T_i N_i} = (1-\rho_i)^{T_i}.$$

The only additional information about ψ from count data in the second component, therefore, comes through p, and since there is more information about p in the count data, we would expect ψ to be more precisely estimated when we use the count data rather than the presence-absence data. In general, when estimating p from binomial data, precision will be greater for larger N. This implies that for our estimation problem, we can expect the gain in precision from using count data over presence-absence data to estimate ψ to be greater when the N_i, and, therefore, the observed counts, are larger. We plant to revisit this problem in more detail in future work.

The relationship (1.15) is the link between repeated count models and presence-absence models. Royle and Nichols (2003) exploited this relationship in an attempt to estimate abundance from presence-absence data alone by taking the same approach that Royle (2004) took for count data: construct a likelihood from the data parameterized in terms of p and N_i, give the N_i a parametric distribution such as the Poisson or negative binomial, and integrate out the nuisance N_i. Except for the model for the data (Bernoulli instead of binomial) the procedure is exactly as for repeated count data. As with the likelihoods for count data from the previous section, their likelihood did not explicitly include occupancy parameters as estimates of ψ_i follow directly from $P(N_i > 0)$, for example Equation (1.13) for the negative binomial distribution. Given detection probabilities of $p > 0.3$, they found that fitting models with a Poisson distribution for abundance was possible for sufficiently large data sets (e.g., $L = 100$, $T_i = T = 5$), but modeling with the more realistic negative binomial distribution was extremely unstable. For smaller p, much larger samples were required for estimation, even with the Poisson model. Our view is that although it may be possible in some cases to get useful estimates of abundance from presence-absence data, it would be inadvisable to design a presence-absence survey for the purpose of estimating abundance.

1.4 Bayesian Hierarchical Modeling

1.4.1 Background

The complexity of many of the models presented in the previous sections means that standard maximum likelihood estimation is not feasible, as we illustrated with Equation (1.8) in Section 1.1. However, we have constructed these models with a hierarchical structure that makes them particularly suitable for Bayesian modeling using Markov Chain Monte Carlo (MCMC) algorithms. In this section we present a short review of Bayesian modeling as it relates to our work. For a detailed overview we refer the reader to Gelman et al. (2004), while Link et al. (2002) has a nice introduction to Bayesian MCMC modeling aimed at wildlife scientists.

In Bayesian modeling, the goal is to compute the form of the distribution of the parameters, given the data and some prior information. The distribution thus determined is known as the posterior distribution, and for a parameter vector $\boldsymbol{\theta}$ and data $\mathbf{Y}$, this is found by solving the equation

$$\wp(\boldsymbol{\theta}|\mathbf{Y}) = \frac{\wp(\mathbf{Y}|\boldsymbol{\theta})\wp(\boldsymbol{\theta})}{\int \wp(\mathbf{Y}|\boldsymbol{\theta})\wp(\boldsymbol{\theta})d\boldsymbol{\theta}}, \tag{1.17}$$

where $\wp(\boldsymbol{\theta}|\mathbf{Y})$ is the posterior density, $\wp(\mathbf{Y}|\boldsymbol{\theta})$ is the likelihood of the data and $\wp(\boldsymbol{\theta})$ is a prior density through which previous knowledge about $\boldsymbol{\theta}$ can be incorporated.

The main philosophical difference between Bayesian and traditional frequentist modeling is our ability to incorporate prior information about the parameters into Bayesian modeling. Instead of ignoring relevant previous research and expert opinion, making use of such information in the analysis of present data seems very appealing, particularly when the amount of data is not large, which is almost always the case in wildlife population research. There is a large element of subjectivity in specifying informative priors for model parameters, as we must decide exactly how to weight past knowledge relative to the present data, that is, determine how important or relevant this information is in the context of the present study. It is important, therefore, that any informative priors are clearly justified by the analyst. Many authors instead avoid the issue of subjectivity by choosing non-informative or vague priors for the parameters in what Link et al. (2002) calls an objective Bayesian analysis. In this case the analysis focuses on the information contained in the present data, with the form of the posterior being approximately determined by the likelihood alone. However, even choosing a non-informative prior requires some care. There are usually several alternatives and in some situations the choice may affect model inference, e.g., variance

parameters for hierarchical models (Gelman, 2006). Therefore, it is advisable to test the robustness of any analysis to different choices of prior.

Regardless of whether priors are informative or not, for most useful models, and certainly the ones of interest in our work, the solution to Equation (1.17) does not have a closed form and instead we rely on MCMC algorithms to generate samples from the posterior distributions of the parameters. For many of the models we consider, we can use the WinBUGS software for Bayesian hierarchical modeling using MCMC methods (Spiegelhalter et al., 2003). Where use of this software is not be possible for a particular model, we will write our own algorithms based on Gibbs sampling (Geman and Geman, 1984) and the Metropolis-Hastings algorithm (Hastings, 1970). We now review these methods.

1.4.2 Markov Chain Monte Carlo Methods

Bayesian analysis of complex hierarchical modeling relies on the computationally intensive MCMC methods which aim to generate samples from the posterior distribution of the parameters. Once these algorithms have converged, with sufficiently large samples we may compute any statistic from the posterior distribution such as posterior means, standard deviations, quantiles, correlations between the parameters, or the form of posterior densities themselves (for example, using kernel density estimation).

Suppose we have a parameter vector of length d, $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_d]$, and data $\mathbf{Y}$. Gibbs sampling depends on being able to sample from the full conditional distributions, $\wp(\theta_i|\boldsymbol{\theta}_{-i}, \mathbf{Y})$, $i = 1, \ldots, d$, where

$$\boldsymbol{\theta}_{-i} = [\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_d].$$

When this is possible, the algorithm will generally proceed very efficiently compared to some alternatives. Gibbs sampling begins with the specification of a vector of starting values, $\boldsymbol{\theta} = [\theta_1^{(0)}, \ldots, \theta_d^{(0)}]$ and then proceeds as follows for iterations $b = 1, \ldots, B$:

Sample $\theta_1^{(b)}$ from $\wp(\theta_1|\theta_2^{(b-1)}, \theta_3^{(b-1)}, \ldots, \theta_d^{(b-1)}, \mathbf{Y})$

Sample $\theta_2^{(b)}$ from $\wp(\theta_2|\theta_1^{(b)}, \theta_3^{(b-1)}, \ldots, \theta_d^{(b-1)}, \mathbf{Y})$

...

Sample $\theta_d^{(b)}$ from $\wp(\theta_d|\theta_1^{(b)}, \theta_2^{(b)}, \ldots, \theta_{d-1}^{(b)}, \mathbf{Y})$.

Under mild conditions, $\boldsymbol{\theta} = [\theta_1^{(b)}, \ldots, \theta_d^{(b)}]$ will converge in distribution to a sample from the joint posterior distribution, $\wp(\boldsymbol{\theta}|\mathbf{Y})$ (Geman and Geman, 1984).

For many models, including ours, it will not be possible to write down the form of the full conditional distributions for all d parameters. Nevertheless, we will be able to write down the full conditional up to a constant of proportionality, since from (1.17), $\wp(\boldsymbol{\theta}|\mathbf{Y}) \propto \wp(\mathbf{Y}|\boldsymbol{\theta})\wp(\boldsymbol{\theta})$. This fact is exploited by the Metropolis-Hastings algorithm to generate samples from the posterior distribution. Consider again the case of a d dimension parameter vector $\boldsymbol{\theta}$. In the Metropolis-Hastings algorthim we begin by specifying a candidate density $q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(b-1)})$ which is valid for the full support of $\boldsymbol{\theta}^{(b-1)}$. At each iteration, a sample is drawn from the candidate distribution and examined to see how plausible it is as a draw from the posterior compared to the parameter value from the previous iteration, and then accepted or rejected as the value for the current iteration. Again for iterations $b = 1, \ldots, B$, the algorithm has the following form:

Sample $\boldsymbol{\theta}^*$ from $q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(b-1)})$

Compute

$$r = \frac{\wp(\mathbf{Y}|\boldsymbol{\theta}^*)\wp(\boldsymbol{\theta}^*)q(\boldsymbol{\theta}^{(b-1)}|\boldsymbol{\theta}^*)}{\wp(\mathbf{Y}|\boldsymbol{\theta}^{(b-1)})\wp(\boldsymbol{\theta}^{(b-1)})q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(b-1)})}$$

If $r \geq 1$, let $\boldsymbol{\theta}^{(b)} = \boldsymbol{\theta}^*$.

If $r < 1$, let $\boldsymbol{\theta}^{(b)} = \boldsymbol{\theta}^*$ with probability r and $\boldsymbol{\theta}^{(b)} = \boldsymbol{\theta}^{(b-1)}$ with probability $(1 - r)$.

Again, under mild conditions, a sample generated using Metropolis-Hastings will converge to a sample from the joint posterior distributions (Hastings, 1970). When the candidate density is symmetric, that is $q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(b-1)}) = q(\boldsymbol{\theta}^{(b-1)}|\boldsymbol{\theta}^*)$, the ratio r simplifies and we have the Metropolis algorithm. An example would be if $\boldsymbol{\theta}^* \sim \mathrm{N}(\boldsymbol{\theta}^{(b-1)}, \boldsymbol{\Sigma})$. For this and any other candidate density, we must specify additional parameters (here it is the matrix $\boldsymbol{\Sigma}$) and the choice of these will determine how frequently candidate values are accepted for both the Metropolis and Metropolis-Hastings algorithms. If new values are accepted too frequently, the samples will be highly correlated and the sample space will be traversed very slowly. Conversely, if most values are rejected, the samples will hardly change at all. In practice, we may need to try different values of the parameters of $q(.)$ to achieve an optimal rejection rate.

In many instances, full conditionals will be available for some parameters but not for others. When this is the case, the Metropolis-Hastings algorithm can be inserted as a single step within the Gibbs sampler for each of the parameters for which direct sampling from the full conditional is not possible. See Gelman et al. (2004) for a more detailed discussion.

Whichever MCMC algorithm or combination of algorithms is used, we must take care to ensure

that the algorithms converge and that we only analyze samples generated after convergence. Convergence itself can be determined by generating samples from multiple Markov chains, begun with widely dispersed initial values, and comparing the behavior of the chains. We can do this visually by making simultaneous plots of the values of the multiple chains for each parameter, with good mixing of the chains implying convergence. In addition we can also compute the Gelman-Rubin scale-reduction factor (Gelman and Rubin, 1992), which compares variation within the chains to variation across the chains. Convergence will lead to values of this factor close to 1.

We have already mentioned that we can use the samples that are generated following convergence of the algorithms to calculate posterior summary statistics. In particular, the posterior mean or median is often used as a point estimate of a parameter, with uncertainty in the value measured by the posterior standard deviation or a pair of quantiles (e.g., 2.5% and 97.5%). The precision of the posterior means can be estimated using the methods of time series analysis which allow for correlation in the series of generated values, for example by estimating the spectral density at zero (see Monahan, 2001).

1.4.3 Model Selection

One of the goals of model fitting will be to select a "best" model from a set of competing models. When fitting models using maximum likelihood estimation, the Akaike Information Criterion (AIC) is often used for comparing models the fit of a selection of models (Akaike, 1973). It is calculated as

$$AIC = -2\log\left\{\max_\theta \mathcal{L}(\mathbf{Y}|\boldsymbol{\theta})\right\} + 2p_\theta$$

where $\max_\theta \mathcal{L}(\mathbf{Y}|\boldsymbol{\theta})$ is the likelihood of the data maximized with respect to the parameter vector $\boldsymbol{\theta}$ and p_θ here is the number of parameters in the model. The first term in the AIC is a measure of the goodness of fit of the model, while the second term penalizes more complex models. For Bayesian modeling, a Bayesian version of the AIC statistic has also been proposed. Let $D(\boldsymbol{\theta}) = -2\log\mathcal{L}(\mathbf{Y}|\boldsymbol{\theta})$ be the deviance statistic. We compute this for every value of the parameter vector $\boldsymbol{\theta}$ generated by the MCMC algorithms following convergence, and from these samples, we compute $\bar{D}$, the posterior mean. Gosky (2004) suggested using $AIC = \bar{D} + 2p_\theta$ as a measure of model fit. He applied it to standard closed-population capture-recapture models and found to to work very well. However, we note that such models lack the complex mixed-effects structure of those we consider, and neither version of the AIC has formal theoretical justification for such models.

Alternatively, Spiegelhalter et al. (2002) proposed a generalization of the AIC for Bayesian modeling called the Deviance Information Criterion (DIC). Like the AIC, the DIC is the sum of two components, one measuring model fit and the other, model complexity. Let $D(\boldsymbol{\theta}) = -2\log\mathcal{L}(\mathbf{Y}|\boldsymbol{\theta})$ be the deviance statistic. In the DIC, the model fit is measured by $\bar{D}$, the posterior mean of the deviance statistic. The complexity is measured by the effective number of parameters p_D, calculated as $p_D = \bar{D} - D(\bar{\boldsymbol{\theta}})$, the difference between $\bar{D}$ and $D(\bar{\boldsymbol{\theta}})$, the deviance evaluated at the posterior mean of the parameter vector. This gives

$$DIC = \bar{D} + p_d = D(\bar{\boldsymbol{\theta}}) + 2p_d$$

$D(\boldsymbol{\theta})$ can be computed at each iteration of the MCMC algorithm and its posterior expectation estimated, as for other parameters, by averaging the generated values after convergence of the algorithm. $D(\bar{\boldsymbol{\theta}})$ is easily computed from the estimated posterior means of the model parameters.

Spiegelhalter et al. (2002) discuss how the DIC can be affected by choice of which part of the likelihood to focus on. A practical approach, and the one taken by WinBUGS, is to calculate the likelihood as a function of the "parent nodes" of the data, the parameters of which the data are directly a function. For example, for repeated count data where $y_{ij} \sim \text{Bin}(N_i, p_i)$, the parent nodes are N_i and p_i, and at each iteration of the MCMC algorithm we compute

$$\log\mathcal{L}(\mathbf{Y}|\mathbf{N}, \boldsymbol{p}) = \log\left\{\prod_{i=1}^{L}\prod_{j=1}^{T_i}\binom{N_i}{y_{ij}}p_i^{y_{ij}}(1-p_i)^{N_i-y_{ij}}\right\}.$$

Computing a more complete version of the likelihood is impractical for the same reasons maximum likelihood estimation is difficult, since we must numerically integrate out the parameters N_i and p_i from the likelihood.

Although the DIC appears to work well in many situations, problems can arise for some mixed effects models, including those we consider here. For these models it is possible for the effective number of parameters to be negative (Spiegelhalter et al., 2002) and more complex models are rewarded rather than penalized. In such cases the DIC criterion breaks down as a model selection tool. Other versions of the DIC have been proposed for mixed-effects models by Celeux et al. (2006), but the authors were unable to make clear recommendations based on the performance of their proposals in a simulation study. Their versions of the DIC require further study.

Alternatively, some authors have considered cross-validation for model fit checking. Each observation is removed in turn and predicted from a model fitted to the remaining data, and then these predicted values are compared to the observed data. Stern and Cressie (2000) discuss methods for predictive model checking for count data models similar to those we have examined. However, in our models detection is imperfect and we do not have any direct observations of density or abundance to compare with prediction. Wyatt (2003) instead compared the cross-validation predictions with estimates from the model fitted to the full data, but the closeness of predictions from reduced data to that from the full data may not be very informative if the full data estimation is poor. Further, with large data sets, fitting Bayesian hierarchical models can be time consuming, and refitting a model again for each observation will be impractical.

Gelfand and Ghosh (1998) proposed a criterion based on the minimizing a posterior predictive loss function. Their criterion is more general than the DIC of Spiegelhalter et al. (2002), and with an appropriate choice of loss function, will not suffer the problem of a negative penalty term for complex hierarchical models. Further, unlike cross-validation methods, it is computationally efficient to compute within an MCMC algorithm for Bayesian modeling. Suppose we observe data vector $\mathbf{Y} = [y_1, \ldots, y_L]$ and that $\mathbf{Y}^{rep}$ is a replicate vector generated under the same underlying model as $\mathbf{Y}$. Gelfand and Ghosh (1998) proposed constructing a loss function based on the weighted sum of one penalty term for departures from the observed value, and another penalty term for departures from the replicate. They give weights k and 1 to these terms respectively, and consider different values of k. A common choice of loss function is the square predicted error loss, and Gelfand and Ghosh (1998) show when this is minimized their criterion becomes

$$\begin{aligned} D_{GG} &= \frac{k}{k+1}\sum_{i=1}^{L}(\mu_i - y_i)^2 + \sum_{i=1}^{L}\sigma_i^2 \\ &= \frac{k}{k+1}G + P, \end{aligned}$$

where $\mu_i = \mathrm{E}(y_i^{rep}|\mathbf{Y})$ and $\sigma_i^2 = \mathrm{var}(y_i^{rep}|\mathbf{Y})$. These predictive means and variance are easily computed from the posterior samples of y_i^{rep}, which themselves are easily generated at each MCMC iteration using the values of the model parameters for that iteration. The terms G and P are respectively a measure of model fit and a penalty for model complexity, and so D_{GG} is of the same form as the AIC and DIC. Gelfand and Ghosh (1998) show that their criterion is little affected by the

choice of weighting parameter, k, and by choosing $k = \infty$, the criterion simplifies to

$$D_{GG} = \mathrm{E}\left[\sum_{i=1}^{L}(y_i^{rep} - y_i)^2|\mathbf{Y}\right].$$

Alternatively, Ghosh and Norris (2005) use the mean square predicted error loss function ($MSPE$) on the log-scale for selecting models for heterogeneity in detectibility for capture-recapture data. We use this loss function as the basis for a criterion for model selection in Chapter 2, when we compare models fitted to data from a repeated count survey of the fish population in a stream.

1.5 Models for Spatially Correlated Animal Species Data

1.5.1 Background

Until recently, modeling of the spatial distribution of animal species was done using by assuming independent observation and applying regression-type models relating species abundance or presence to ecological covariates, e.g., habitat and climate variables (see Guisan et al., 2002, for a review). As discussed in Gelfand et al. (2006), this modeling assumes that environmental variables are the principle factors affecting species distribution, and that species have reached equilibrium with respect to these variables, something which is rarely true. Factors such as species dispersal, competition with other species, and population dynamics also influence the distribution of animals, and can to lead to spatial clustering in the distribution of animal populations which is unexplained by environmental variables alone, something ignored by regression models. Failure to account for this spatial dependence in modeling animal survey data can exaggerate the effects of covariates and lead to misleading estimates of the precision of predicted values at unsurveyed locations. In this section we begin by describing how spatial dependence can be included in the modeling framework we have outlined previously for surveys of animals population, and go on to review some of the more important spatial models which have been proposed for animal survey data.

As before, we assume that survey data are collected at L sites within a monitoring region. Recall that the first part of modeling site abundances for repeated count data was to assume that they follow a Poisson distribution with mean λ_i (Equation (1.6)) for site i. The next stage, which allowed for likely overdispersion in the site abundances, was to model the λ_i as coming from a gamma distribution. The standard approach for introducing both covariates and a spatial component into

the model is from Diggle et al. (1998), and replaces the gamma model with a lognormal model,

$$\log(\lambda_i) = \mu_i + \phi_i. \tag{1.18}$$

The term μ_i represents the covariate model, e.g., a linear function of environmental variables, $\mu_i = \mathbf{X}_i'\boldsymbol{\alpha}$. ϕ_i is a zero-mean Gaussian random error, and spatial dependence in the site abundances is modeled only through specifying a covariance model for the multivariate normal vector of the errors, $\boldsymbol{\phi} = [\phi_1, \ldots, \phi_L]'$.

Similarly, we can incorporate spatial structure into models for presence-absence data by specifying a model for the site occupancy probabilities, ψ_i. In this case we use the logistic link function, as suggested by Diggle et al. (1998) for binary data in general,

$$\text{logit}(\psi_i) = \mu_i + \phi_i, \tag{1.19}$$

where the terms on the right hand side have the same interpretation as before.

Modeling the spatial dependence through the random error term in a linear model is not the only approach available. For example, Gelfand et al. (2003) note that in some applications, fitting regression-type covariate models with fixed coefficients may not be appropriate. Instead they propose models in which the parameters of the covariate model are themselves a function of spatial location, while the ϕ_i are independent. While noting the potential usefulness of such models, we do not consider them further in our work.

In the remainder of this section we review two approaches to modeling spatial dependence which have been used for animal survey data, and discuss their advantages and disadvantages. A third method, spectral modeling, has recently been proposed by Wikle (2002) as an efficient way of modeling spatial dependence in animal count data, but at least for now, we do not consider this method in detail. There is a vast and ever expanding body of literature on spatial models, and our goal here is simply to provide an elementary introduction to some important models which are relevant to our work, largely based on Cressie (1993) and Banerjee et al. (2004). For geostatistical models in particular, much research has been done on developing generalizations which relax in some way the assumptions of the models we present, such as assumptions of stationarity and isotropy. While noting its existence and importance, it is beyond our scope to examine this research here.

1.5.2 Geostatistical Models

In geostatistical modeling, the data we gather are thought of as a realization of a random process, $\{Z(\boldsymbol{s}) : \boldsymbol{s} \in D\}$ observed at some fixed set of locations $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_L\}$, where D is a fixed subset of d-dimension Euclidian space. In animal population modeling, d is typically 2, although we also consider the case of $d = 1$ when modeling the distribution of fish in streams. If $Z(\boldsymbol{s})$ has constant mean for all s in D, and the covariance of the process at two locations $\text{cov}(Z(\boldsymbol{s}_1), Z(\boldsymbol{s}_2))$ is a function only of the separation vector $\boldsymbol{h} = \boldsymbol{s}_1 - \boldsymbol{s}_2$, that is,

$$\text{cov}(Z(\boldsymbol{s}_1), Z(\boldsymbol{s}_2)) = C(\boldsymbol{h})$$

for some function $C(.)$, then as defined by Cressie (1993), the process $Z(\boldsymbol{s})$ is called second-order (or weakly) stationary. Further, if $C(\boldsymbol{h})$ is a function only of the euclidian distance between two locations, $r = \|\boldsymbol{h}\|$, then we say the process is isotropic. Isotropic models have been widely used largely because they are relatively straightforward to fit and are easy to iterpret, and for this reason we focus on this class of geostatistical model for modeling the spatial dependence in the component ϕ of our animal population models. Nevertheless, isotropy is not always a reasonable assumption. It may be that due to factors such unmeasured ecological covariates or patterns of animal dispersal, spatial dependence is stronger over the same distance in some directions than others and it is, therefore, important to check the validity of the isotropy assumption (see Cressie, 1993, for examples).

Models for spatial processes are usually defined in terms of the semivariagram, $\gamma(\boldsymbol{s}_1 - \boldsymbol{s}_2) = 0.5\text{var}(Z(\boldsymbol{s}_1), Z(\boldsymbol{s}_2))$, which for isotropic models can be written as $\gamma(\|\boldsymbol{h}\|)$, or for greater simplicity, $\gamma(r)$. We only consider models with valid semivariograms (see Cressie, 1993, for conditions for validity) and corresponding covariance functions, $C(r)$. Although models are usually defined in terms of the semivariogram, for ease of interpretation, we present the covariance functions for a selection of useful models.

Spherical

$$C(r) = \begin{cases} \sigma_\tau^2 + \sigma_\phi^2 & r = 0 \\ \sigma_\phi^2 \left\{ 1 - \frac{3}{2}\frac{r}{a} + \frac{1}{2}\left(\frac{r}{a}\right)^3 \right\} & 0 < r \leq a \\ 0 & r > a \end{cases}$$

The three parameters of this function are commonly known as the nugget, partial sill, and range. The nugget, σ_τ^2 represents microscale variation or variation due to measurement error in the spatial process. The parameter σ_ϕ^2 is the partial sill, while $\sigma_\tau^2 + \sigma_\phi^2$ is known as the sill. It is the range parameter, a, that determines the rate of decay of spatial dependence with distance, and for the spherical model, there is assumed to be no spatial dependence between sites when they are separated by a distance greater than a. We note that the spherical model is valid for $d = 1, 2$ or 3 dimenstions, and so is applicable to modeling data from surveys of animal populations.

Exponential

$$C(r) = \begin{cases} \sigma_\tau^2 + \sigma_\phi^2 & r = 0 \\ \sigma_\phi^2 \exp(-r/a) & r > 0 \end{cases}$$

Unlike the spherical model, the range parameter a does not represent the absolute range of non-zero spatial dependence since the spatial covariance only approaches 0 as distance r goes to infinity. Therefore, a represents the rate of exponential decay in the strength of the dependence with increasing distance between two points. In particular, we note that $C(r)$ goes to zero as a goes to zero, leading to the case of spatial independence. Royle et al. (2002) applied the exponential model to point count data from bird surveys, while a spectral approximation to this model was fitted by Wikle (2002) to similar data. This model is included in the GeoBUGS add-on to WinBUGS (Thomas et al., 2002).

Gaussian

$$C(r) = \begin{cases} \sigma_\tau^2 + \sigma_\phi^2 & r = 0 \\ \sigma_\phi^2 \exp(-r^2/a^2) & r > 0 \end{cases}$$

The Gaussian model is very similar in form to the exponential, and in practice it can be difficult to distinguish differences in fit when these models are applied to the same data. The Gaussian model, however, yields a 'smoother' spatial process. Smoothness refers to the continuity of the process, and the underlying nature of a process may mean a smoother or less smooth spatial model is more meaningful in a physical or biological sense.

Matérn

$$C(r) = \begin{cases} \sigma_\tau^2 + \sigma_\phi^2 & r = 0 \\ \sigma_\phi^2 \frac{(2\sqrt{\nu} r/a)^\nu}{2^{\nu-1}\Gamma(\nu)} K_\nu(2\sqrt{\nu} r/a) & r > 0 \end{cases}$$

K_ν is the modified Bessel function of order ν, where the additional parameter ν specifies how smooth the spatial process is. The Matérn is really a class of models, with the exponential ($\nu = 1/2$) and Gaussian ($\nu = \infty$, the smoothest model) as special cases.

We now describe how these spatial models can be incorporated into the hierarchical framework for modeling animal populations. Recall that we introduced the spatial modeling through the log of density of animals at a site. Using the notation introduced above for spatial process Z, we now write the model for density, λ, as a function of spatial location $\boldsymbol{s}$

$$\log(\lambda(\boldsymbol{s})) = \mu(\boldsymbol{s}) + Z(\boldsymbol{s}).$$

As before $\mu(\boldsymbol{s})$ is a covariate (or trend) model, for example, $\mu(\boldsymbol{s}) = \mathbf{X}(\boldsymbol{s})'\boldsymbol{\alpha}$. $Z(\boldsymbol{s})$ represents the spatial process, which we decompose into two independent components, the first analagous to the measurement error described above, $\tau(\boldsymbol{s})$, and the other, a spatially dependent component, $\phi(\boldsymbol{s})$,

$$Z(\boldsymbol{s}) = \tau(\boldsymbol{s}) + \phi(\boldsymbol{s}).$$

Suppose we collect data at L locations, $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_L$. We denote the realizations of $Z(\boldsymbol{s})$ at these locations by $\mathbf{Z} = [Z_1, \ldots, Z_L]'$, where $Z_i = \tau_i + \phi_i$. Now assume that $Z(\boldsymbol{s})$ is a zero mean, Gaussian process with isotropic covariance structure following one of the models described above. The vector $\mathbf{Z}$ will then be multivariate normal,

$$\mathbf{Z}|\sigma_\tau^2, \sigma_\phi^2, a, \nu \sim \text{MVN}(\mathbf{0}, \sigma_\tau^2\mathbf{I} + \sigma_\phi^2\mathbf{V})$$

where $\mathbf{V}$ is an $L \times L$ covariance matrix with the ii' element defined by $C(r_{ii'}) = C(\|\boldsymbol{s}_i - \boldsymbol{s}_{i'}\|)$, $i, i' = 1, \ldots, L$. We prefer to work with the components of the spatial process, that is,

$$\boldsymbol{\tau}|\sigma_\tau^2 \sim \text{MVN}(\mathbf{0}, \sigma_\tau^2\mathbf{I}),$$

$$\boldsymbol{\phi}|\sigma_\phi^2, a, \nu \sim \text{MVN}(\mathbf{0}, \sigma_\phi^2\mathbf{V}),$$

where $\boldsymbol{\tau} = [\tau_1, \ldots, \tau_L]'$ and $\boldsymbol{\phi} = [\phi_1, \ldots, \phi_L]'$. These models define the final level in the hierarchy of our models for animal population data. In fact, we will assume in our work that $\tau = 0$ since in practice it can be very difficult to distinguish this term from ϕ. These terms occur far from the data in the model hierarchy, and it is our experience that there is unlikely to be sufficient information to separate the parameters of the two processes, particularly since animal population data is typically sparse and we are simultaneously attempting to model random effects in the detection component of the models.

There is an important practical limitation in using geostatistical models within Baysesian MCMC modeling. At each step of the algorithm we must sample from the full conditional for $\boldsymbol{\phi}$, which requires computing the inverse of an $L \times L$ covariance matrix. In practice, we can actually solve a system of L equations, but this still becomes an enormous drain on computer time when L is large and the operation must be repeated for thousands of MCMC iterations. Things look even worse when our goal is spatial prediction ("kriging", Cressie, 1993; Diggle et al., 1998) at a large number of unsurveyed locations. (We will return to the problem of spatial prediction in Chapter 3.) One possibility is to consider models that lead to covariance matrices with sparse structure, that is, those that have many zero elements. Such matrices require less computer memory to store and efficient algorithms exist for operating upon them (see for example Monahan, 2001). The statistical software R (Ihaka and Gentleman, 1996) includes the SparseM library for storing and performing operations on sparse matrices (Koenker and Ng, 2004). The only one of the isotropic models we have described which naturally leads to such sparse structure is the spherical model, since when the distance between two locations is greater than the range, a, the corresponding element in the covariance matrix will be 0. The degree of sparseness will be a function of how quickly spatial dependence declines with distance and of the geographical spacing of the survey locations. We considered using the spherical covariance function for Bayesian MCMC modeling of large data sets, in particular a bird survey data set with observations at over 3000 locations in the Great Smoky Mountains National Park (Shriner, 2001; Lichstein et al., 2002) in Chapter 3, although we found the models of the following section to be more suitable.

1.5.3 Conditional Autoregressive Models

Much recent interest in spatial modeling has focused on conditional autoregressive (CAR) models (Besag, 1974), largely due to their ease of implementation within a Bayesian hierarchical modeling framework. Unlike geostatistical models, which required a full unconditional multivariate normal

specification for the distribution of $\boldsymbol{\phi}$, in CAR models the distribution of each ϕ_i is defined conditional on $\boldsymbol{\phi}_{-i} = [\phi_1, \ldots, \phi_{i-1}, \phi_{i+1}, \ldots, \phi_L]'$. Specifically, in a CAR model ϕ_i is assumed to be normally distributed with the mean defined to be the weighted average of the elements of $\boldsymbol{\phi}_{-i}$. Following Banerjee et al. (2004) we begin by defining a symmetric $L \times L$ matrix of weights, $\mathbf{W}$, with elements $w_{ii'}$. The simplest case would be to let $w_{ii'} = 1$ if sites i and i' are 'neighbors', and zero otherwise. Alternatively, we may wish the weights to reflect the different distances between sites (or the site centers), and choose $w_{ii'} = 1/r_{ii'}$, where $r_{ii'}$ is the Euclidean distance of site i from site i'. Definition of the neighborhood weight structure is the key factor in specifying a CAR model, and care must be taken depending on the nature of the sampling units. We return to this point a little later.

Now define $w_{i.} = \sum_{i'=1}^{L} w_{ii'}$, the row totals of $\mathbf{W}$. A 'proper' Gaussian CAR model then follows by defining the distribution of ϕ_i conditionally as

$$\phi_i | \boldsymbol{\phi}_{-i}, \gamma, \sigma_\phi^2 \sim \mathrm{N}\left(\gamma \sum_{i' \neq i} \frac{w_{ii'}}{w_{i.}} \phi_{i'}, \frac{\sigma_\phi^2}{w_{i.}} \right). \tag{1.20}$$

The parameter γ in (1.20) is interpreted as a measure of the strength of the spatial correlation, with $\gamma = 0$ implying independence and higher values of $|\gamma|$ leading to greater positive or negative correlation. We note that the weights not only define the mean of the conditional distribution, but also that when $w_{i.}$, the sums of the neighbors' weights for a site, is large (perhaps because that site is far from other sites), its conditional variance will be larger than if $w_{i.}$ is small. Thus the neighborhood structure also ultimately affects the precision with which site abundance is estimated and the precision of predictions at unsampled sites. Unconditionally, $\wp(\boldsymbol{\phi})$, the probability density of $\boldsymbol{\phi}$, is given by

$$\wp(\boldsymbol{\phi} | \gamma, \sigma_\phi^2) \propto \exp\left\{ -\frac{1}{2\sigma_\phi^2} \boldsymbol{\phi}' \mathbf{M}^{-1} (\mathbf{I} - \gamma \mathbf{C}) \boldsymbol{\phi} \right\}, \tag{1.21}$$

where $\mathbf{M}$ is a diagonal matrix with ith diagonal element $1/w_{i.}$, and $\mathbf{C} = \mathbf{MW}$, the scaled weight matrix (Banerjee et al., 2004). Equation (1.21) specifies the kernel of a multivariate normal distribution with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}_\phi = \sigma_\phi^2 (\mathbf{I} - \gamma \mathbf{C})^{-1} \mathbf{M}$. With weights defined above, $\boldsymbol{\Sigma}_\phi^{-1} = \mathbf{M}^{-1} (\mathbf{I} - \gamma \mathbf{C})$ will be symmetric, and provided that $|\gamma| < 1$, it will also be non-singular and $\boldsymbol{\Sigma}_\phi$ will be a valid covariance matrix. In this case, (1.20) defines a proper prior distribution in Bayesian hierarchical modeling (Banerjee et al., 2004), and hence this model specification is known as a proper Gaussian CAR model.

There are some problems in intepretation of this model. Banerjee et al. (2004) illustrate that values of γ close to 1 do not necessarily imply high correlation, and so γ cannot be interpreted in the same way as a correlation coefficient. Further, the proper CAR model, (1.20), gives the conditional mean of ϕ_i as a *proportion* of a weighted average of its neighbors, which itself makes precise interpretation unclear. For these reasons, an alternative model, known as the intrinsic CAR model, omits γ, leading to the mean of ϕ_i being expressed directly as a weighted sum of its neighbors,

$$\phi_i|\boldsymbol{\phi}_{-i} \sim \mathrm{N}\left(\sum_{i' \neq i} \frac{w_{ii'}}{w_{i.}}\phi_{i'}, \frac{\sigma_\phi^2}{w_{i.}}\right). \tag{1.22}$$

However, examination of Equation (1.22) shows that it is not a proper prior for Bayesian modeling. Here $\wp(\boldsymbol{\phi})$ has kernel

$$\wp(\boldsymbol{\phi}|\sigma_\phi^2) \propto \exp\left\{-\frac{1}{2\sigma_\phi^2}\boldsymbol{\phi}'\mathbf{M}^{-1}\left(\mathbf{I}-\mathbf{C}\right)\boldsymbol{\phi}\right\}.$$

This again appears to be the kernel of a multivariate normal distribution with mean vector $\mathbf{0}$, but this time with covariance matrix $\boldsymbol{\Sigma}_\phi = \sigma_\phi^2(\mathbf{I}-\mathbf{C})^{-1}\mathbf{M}$. However, $\mathbf{M}^{-1}(\mathbf{I}-\mathbf{C})$ is now singular and so $\boldsymbol{\Sigma}_\phi$ does not exist. Nevertheless, as the posterior distribution of $\boldsymbol{\phi}$ is still proper in many situations (Sun et al., 1999), this simpler CAR model is also worth consideration.

CAR models are particularly well suited to survey regions which are divided into large contiguous units and we collect data from every or almost every unit in the region. Thus these models have proved useful in disease modeling when the data consists of counts of disease cases within each county of a state, or region of a country (e.g. Pascutto et al., 2000; Gelfand and Vounatsou, 2003). Here it is relatively clear how to define the neighborhood structure, with a unit's neighbors typically defined to be adjacent units only. When a count is obtained at a point or a transect, or when the sampled plots comprise just a small fraction of the entire survey region, much more care is required when defining the neighborhood weights. A possible approach is that of Lichstein et al. (2002). These authors began by fitting a model without spatial dependence to point count data and examining the empirical semivariogram (Cressie, 1993) of the residuals to estimate the distance at which spatial correlation between two points is approximately zero. A survey point's 'neighbors' were then defined to be those points closer than the estimated range, and these were given weights inversely proportional to the distance or the square of the distance from the survey point. The choice of weighting function is intuitive, but somewhat arbitrary, although this may not be so important in a proper CAR model, where the strength of spatial dependence is also determined

by γ. Ideally one would check the robustness of estimates to the choice of weights, or if possible use a model selection procedure to choose the weighting function from a set of candidates, as per Lichstein et al. (2002).

As with geostatistical models, we are able to perform prediction at unmonitored sites. For CAR models this proceeds in a quite different way than kriging for geostatistical models, but is computationally straightforward when part of a Bayesian MCMC modeling. In the case where the survey region is divided into contiguous subregions, all units, whether surveyed or not, must be included in the CAR modeling in order that the neighborhood structure be correctly defined. When survey sites are points or transects, we can define a set of additional sites (e.g., on a regular grid) at which we wish to make predictions. Again the neighborhood structure will be defined for the full set of sampled and unsampled points.

In the Chapter 2 we apply a proper CAR model to allow for spatial dependence in the population distribution of stream-dwelling fish sampled using a repeated count survey. In Chapter 3 we examine a variation of the standard CAR models due to Hrafnkelsson and Cressie (2003) for modeling spatial dependence in a large bird survey data set.

Bibliography

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, *in* B. N. Petrov and F. Csaki (eds), *Proceedings of the Second International Symposium on Information Theory*, pp. 267–281.

Alho, J. M. (1990). Logistic regression in capture-recapture models, *Biometrics* **46**: 639–649.

Alldredge, M. W. (2004). *Avian Point Count Surveys: Estimating Components of the Detection Process*, PhD thesis, North Carolina State University.

Banerjee, S., Carlin, B. P. and Gelfand, A. E. (2004). *Hierarchical Modeling and Analysis for Spatial Data*, Boca Raton: Chapman and Hall.

Barker, R. J. and Sauer, J. R. (1995). Statistical aspects of point count sampling, *in* C. J. Ralph, J. R. Sauer and S. Droege (eds), *Monitoring Bird Populations by Point Counts*, pp. 125–130. U.S. Department of Agriculture, Forest Service General Technical Report PSW-GTR-149.

Bayley, P. and Peterson, J. (2001). An approach to estimate probability of presence and richness of fish species, *Transactions of the American Fisheries Society* **130**: 620–633.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems, *Journal of the Royal Statistical Society, Series B* **36**: 192–236.

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

Burnham, K. P. and Overton, W. S. (1978). Estimation of the size of a closed population when capture probabilities vary among animals, *Biometrika* **65**: 625–633.

Carroll, R. J. and Lombard, F. (1985). A note on n estimators for the binomial distribution, *Journal of the American Statistical Association* **80**: 423–426.

Casella, G. (1986). Stabilizing binomial n estimators, *Journal of the American Statistical Association* **81**: 172–176.

Celeux, G., Forbes, F., Robert, C. P. and Titterington, D. M. (2006). Deviance information critera for missing data models, *Bayesian Analysis* **(to appear)**.

Chao, A. (1987). Estimating the population size for capture-recapture data with unequal catchability, *Biometrics* **43**: 783–791.

Chao, A., Lee, S.-M. and Jeng, S.-L. (1992). Estimating the population size for capture-recapture data when capture probabilities vary by time and individual animal, *Biometrics* **48**: 201–216.

Coull, B. A. and Agresti, A. (1999). The use of mixed-logit models to reflect heterogeneity in capture-recapture studies, *Biometrics* **55**: 294–301.

Craig, B. A., Newton, M. A., Robert, A. G., Reynolds, J. E. and Wilcox, J. R. (1997). Analysis of aerial survey data on Florida manatee using Markov Chain Monte Carlo., *Biometrics* **53**: 524–541.

Cressie, N. (1993). *Statistics for Spatial Data*, 2nd edn, New York: Wiley.

Diggle, P. J., Tawn, J. A. and Moyeed, R. A. (1998). Model-based geostatistics (with discussion), *Applied Statistics* **47**: 299–350.

Dorazio, R. M., Jelks, H. L. and Jordan, F. (2005). Improving removal-based estimates of abundance by sampling a population of spatially distinct subpopulations, *Biometrics* **61**: 1093–1101.

Dorazio, R. M. and Royle, J. A. (2003). Mixture models for estimating the size of a closed population when capture rates vary among individuals, *Biometrics* **59**: 351–364.

Forsyth, D. M., Link, W. A., Webster, R., Nugent, G. and Warburton, B. (2006). Nonlinearity and seasonal bias in an index of brushtail possum abundance, *Journal of Wildlife Management* **(to appear)**.

Geissler, P. H. and Fuller, M. R. (1986). Estimation of the proportion of area occupied by an animal species, *Proceedings of the Section on Survey Research Methods of the American Statistical Association* **1**: 533–538.

Gelfand, A. E. and Ghosh, S. (1998). A minimum predictive posterior loss approach, *Biometrika* **85**: 1–11.

Gelfand, A. E., Kim, H.-J., Sirmans, C. F. and Banerjee, S. (2003). Spatial modeling with variable coefficients, *Journal of the American Statistical Association* **98**: 387–396.

Gelfand, A. E., Valenti, J., Wu., S. and Sirmans, C. F. (2006). Building statistical models to analyze species distributions, *Ecological Applications* **(to appear)**.

Gelfand, A. E. and Vounatsou, P. (2003). Proper multivariate conditional autoregressive models for spatial data analysis, *Biostatistics* **4**: 11–25.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models, *Bayesian Analysis* **(to appear)**.

Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (2004). *Bayesian data analysis*, 2nd edn, Boca Raton: Chapman and Hall.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences, *Statistical science* **7**: 457–472.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian resolution of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**: 721–741.

Ghosh, S. K. and Norris, J. L. (2005). Bayesian capture-recapture analysis and model selection allowing for heterogeneity and behavioral effects, *Journal of Agricultural, Biological and Environmental Statistics* **10**: 35–49.

Gosky, R. M. (2004). *Bayesian Analysis and Matching Errors in Closed Population Capture Recapture Models*, PhD thesis, North Carolina State University.

Guisan, A., Edwards, T. C. and Hastie, T. (2002). Generalized linear and generalized additive models in studies of species distributions: setting the scene, *Ecological Modelling* **157**: 89–100.

Hankin, D. G. and Reeves, G. H. (1988). Estimating total fish abundance and total habita area in small streams based on visual estimation methods, *Canadian Journal of Fisheries and Aquatic Science* **45**: 834–844.

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

Hrafnkelsson, B. and Cressie, N. (2003). Hierarchical modeling of count data with application to nuclear fall-out, *Environmental and Ecological Statistics* **10**: 179–200.

Huggins, R. M. (1989). On the statistical analysis of capture experiments, *Biometrika* **76**: 133–140.

Ihaka, R. and Gentleman, R. (1996). R: A language for data analysis and graphics, *Journal of Computational and Graphical Statistics* **5**: 299–314.

Johnson, D. H. (1995). Point counts of birds: what are we estimating?, *in* C. J. Ralph, J. R. Sauer and S. Droege (eds), *Monitoring Bird Populations by Point Counts*, pp. 117–123. U.S. Department of Agriculture, Forest Service General Technical Report PSW-GTR-149.

Johnson, N. L. and Kotz, S. (1969). *Discrete Distributions*, 1st edn, Boston: Haughton Mifflin Company.

Koenker, R. and Ng, P. (2004). *The SparseM package, Version 0.56*. http://www.econ.uiuc.edu/-∼roger/research/sparse/sparse.html.

Lichstein, J. W., Simons, T. R., Shriner, S. A. and Franzreb, K. E. (2002). Spatial autocorrelation and autoregressive models in ecology, *Ecological Monographs* **72**: 445–463.

Link, W. A. (2003). Nonidentifiability of population size from capture-recapture data with heterogeneous detection probabilities, *Biometrics* **59**: 1123–1130.

Link, W. A., Cam, E., Nichols, J. D. and Cooch, E. G. (2002). Of BUGS and birds: Markov Chain Monte Carlo for hierarchical Bayesian modeling in wildlife research, *Journal of Wildlife Management* **66**: 277–291.

MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Royle, J. A., and Langtimm, C. A. (2002). Estimating site occupancy when detection probabilities are less than one, *Ecology* **83**: 2248–2255.

Monahan, J. F. (2001). *Numerical Methods of Statistics*, Cambridge: Cambridge University Press.

Norris, J. L. and Pollock, K. H. (1995). A capture-recapture model with heterogeneity and behavioural response, *Environmental and Ecological Statistics* **2**: 305–313.

Norris, J. L. and Pollock, K. H. (1996). Nonparametric MLE under two closed capture-recapture models with heterogeneity, *Biometrics* **52**: 639–649.

Olkin, I., Petkau, A. J. and Zidek, J. V. (1981). A comparison of n estimators for the binomial distribution, *Journal of the American Statistical Association* **76**: 637–642.

Otis, D. L., Burnham, K. P., White, G. C. and Anderson, D. R. (1978). Statistical inference from capture data on closed animal experiments, *Wildlife Monographs* **62**: 1–135.

Pascutto, C., Wakefield, J. C., Best, N. G., Richardson, S., Bernardinelli, A., Staines, A. and Elliot, P. (2000). Statistical issues in the analysis of disease mapping data, *Statistics in medicine* **19**: 2493–2519.

Pledger, S. (2000). Unified maximum likelihood estimates for closed capture-recapture models using mixtures, *Biometrics* **56**: 434–442.

Pollock, K. H. (2002). The use of auxiliary variables in capture-recapture modelling: an overview, *Journal of Applied Statistics* **29**: 85–102.

Pollock, K. H., Nichols, J. D., Brownie, C. and Hines, J. E. (1990). Statistical inference for capture-recapture experiments, *Wildlife Monographs* **107**: 1–97.

Routledge, R. D. (1981). The unreliability of population estimates from repeated incomplete aerial surveys, *Journal of Wildlife Management* **45**: 997–1000.

Royle, J. A. (2004). N-mixture models for estimating population size from spatially replicated counts, *Biometrics* **60**: 108–115.

Royle, J. A., Link, W. A. and Sauer, J. R. (2002). Statistical mapping of count survey data, *in* J. M. Scott, P. J. Heglund, M. L. Morrison, J. B. Haufler, M. G. Raphael, W. A. Wall and F. B. Samson (eds), *Predicting Species Occurences: Issues of Accuracy and Scale*, Washington: Island Press.

Royle, J. A. and Nichols, J. D. (2003). Estimating abundance from repeated presence-absence data or point counts, *Ecology* **84**: 777–790.

Seber, G. A. F. (1982). *The estimation of animal abundance and related parameters*, 2nd edn, New York: Macmillan.

Shriner, S. A. (2001). *Distribution of breeding birds in Great Smoky Mountains National Park*, PhD thesis, North Carolina State University.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002). Bayesian measures of model complexity and fit, *Journal of the Royal Statistical Society, B* **64**: 583–639.

Spiegelhalter, D., Thomas, A., Best, N. and Lunn, D. (2003). *WinBUGS User Manual, Version 1.4*.

Stauffer, H. B., Ralph, C. J. and Miller, S. L. (2002). Incorporating detection uncertainty into presence-absence surveys for marbled mullet, *in* J. M. Scott, P. J. Heglund, M. L. Morrison, J. B. Haufler, M. G. Raphael, W. A. Wall and F. B. Samson (eds), *Predicting Species Occurences: Issues of Accuracy and Scale*, Washington: Island Press.

Stern, H. S. and Cressie, N. (2000). Posterior predictive model checks for disease mapping models, *Statistics in Medicine* **19**: 2377–2397.

Sun, D., Tsutakaw, R. K. and Speckman, P. L. (1999). Posterior distribution of hierarchical models using car(1) distributions, *Biometrika* **86**: 341–350.

Thomas, A., Best, N., Arnold, R. and Spiegelhalter, D. (2002). *GeoBUGS User Manual, Version 1.1 Beta June 2002*.

Wikle, C. (2002). Spatial modelling of count cata: a case study in modelling breeding bird survey data on large spatial domains, *in* A. B. Lawson and G. T. Denison (eds), *Spatial Cluster Modelling*, Boca Raton: Chapman and Hall/CRC.

Wyatt, R. J. (2002). Estimating riverine fish population size from single- and multiple-pass removal sampling using a hierarchical model, *Canadian Journal of Fisheries and Aquatic Science* **59**: 695–706.

Wyatt, R. J. (2003). Mapping the abundance of riverine fish populations: integrating hierarchical bayesian models with a geographic information system (GIS), *Canadian Journal of Fisheries and Aquatic Science* **60**: 997–1006.

Zippen, C. (1956). An evaluation of the removal method of estimating animal populations, *Biometrics* **12**: 163–189.

Chapter 2

Bayesian Spatial Modeling of Data from Unit-Count Surveys of Fish in Streams

Raymond A. Webster, Kenneth H. Pollock and Sujit K. Ghosh

ABSTRACT

We describe a framework for spatial modeling of data from surveys of stream dwelling fish species in which repeated counts are made of animals within a sample of habitat units. Using Bayesian modeling with Markov chain Monte Carlo (MCMC) algorithms, we are able to estimate fish population size from repeated count survey data while allowing fish detection probabilities to vary across the stream. We propose the use of conditional autoregressive (CAR) models for modeling the spatial dependence of density across the habitat units of the stream. The spatial dependence model can be used along with covariate models for density and detection to predict density at unsampled units, and thereby estimate total fish abundance across the stream. We apply these models to data sampled from an intensive repeated count survey of juvenile Coho Salmon in McGarvey Creek, Northern California. Moderate spatial dependence in fish density was detected, and models which account for spatial dependence produced more precise predictions at unsurveyed units, and thus more precise estimates of total stream abundance, than models which assumed spatial independence. Through a small simulation study, we show that ignoring heterogeneity in detection probabilities can lead to significant underestimation of total abundance. However, inclusion of heterogeneity using a random effect in the detection component of the model can lead to problems in Bayesian MCMC modeling for typical survey designs, and for this reason we stress the importance of accounting for heterogeneity by incorporating covariates in modeling detection probability.

2.1 Introduction

Population surveys of stream-dwelling fish often involve dividing the stream into discrete units, drawing a sample of these units, and counting the number of individual fish in each unit of the sample. These units can be physical habitat units such as pools and runs (e.g., juvenile Coho Salmon, Brakensiek, 2002) or sampling units defined only for the purposes of the survey, for example sections of the stream of a specified length (e.g., sea trout, Wyatt, 2003). Stream surveys are typically conducted using either multiple snorkel dives or multiple-pass electrofishing. In snorkel dives, a diver or a team of divers makes a pass through the unit and a count of the total fish seen on the pass is recorded. Electrofishing involves running a low level electrical current through the unit which temporarily stuns the fish, and the number of stunned fish is recorded.

The goals of stream fish surveys include assessment of the distribution of fish, determining the factors that influence fish density, and estimation of population size. In general the whole population of a unit is not observed on a single pass and therefore to achieve these goals through statistical modeling, we also require information about the probability of detection of individual fish (Thompson, 2002). This is provided by using a two-stage sampling design, in which the first stage is the selection of the units from the stream, and the second stage involves making independent, repeated passes of the survey units. When the survey is conducted using snorkel dives, this requires a sequence of independent dive counts to be made of the population of fish in each of the units selected in the first stage. A possible approach to estimating abundance and detection probability for each unit is to assume a binomial model for the unit counts, and use either method of moments or maximum likelihood estimators (Johnson and Kotz, 1969). This method requires us to assume that the fish population is closed to births and deaths, immigration and emigration over the course of the survey, but this is reasonable when the repeated passes are made in quick succession as is generally the case in these types of surveys. However, estimation of both binomial parameters (population size and detection probability) from a single sample of counts can be highly unstable, with small changes in the data producing large shifts in the estimates of abundance and detection probability. Although more stable estimators exist, e.g., Carroll and Lombard (1985) and Olkin et al. (1981), they are not without problems of their own, particularly for the sparse data that results when population density is low in a unit (Casella, 1986).

Alternatively, we could use multiple-pass electrofishing, which is a removal method of sampling as the fish stunned on a pass are effectively removed from the population for subsequent

passes before being returned to the stream. In this case, maximum likelihood-based removal estimators can be used to estimate population size for an individual unit (Zippen, 1956; Seber, 1982), and this estimation will generally be much more stable and precise than estimation from a sequence of counts made with replacement. However, this type of removal estimation will sometimes fail to produce sensible estimates for particular units, especially when the population is sparsely distributed or when detection probabilities are low (Seber, 1982). Also, both removal and multiple count estimators generally require quite restrictive assumptions on the detection probability, for example, that it does not vary across the multiple passes (Routledge, 1981; Pollock et al., 1990).

Recently a number of authors have proposed hierarchical models for multiple-pass surveys of animal populations that, instead of separately estimating parameters for individual units, combine the information from all surveyed units, leading to more stable estimation and the potential for more flexible modeling. As with a sample from a single unit, these models typically assume a binomial distribution for the observed counts for a unit, with parameters being the unit abundance and detection probability. For removal sampling from independent units, Wyatt (2002) modeled the removals as having a binomial distribution, while Royle (2004) did the same for multiple count data. Dorazio et al. (2005) described an equivalent multinomial specification for removal data. In the next level of the model hierarchy, parametric models are formulated for the binomial parameters themselves. Thus instead of estimating abundance and detection separately for each unit, the parameters of their respective distributions are estimated using the information from the entire data set. The most common model for unit abundances is the negative binomial or, equivalently, the Poisson-gamma mixture, which has the advantage over the simple Poisson model of allowing for the overdispersion, where the variance is greater than the mean. Due to the clustered nature of wildlife populations, overdispersion is almost always present in counts of animals.

Hierarchical models can be easily adapted to include covariates in both the abundance and the detection models and to allow for spatial dependence of density across units. For the abundance component of the model, Wyatt (2003), Royle et al. (2002), and Forsyth et al. (2006) all use a log-linear model to incorporate covariates, while logistic models are routinely used to relate detection or capture probability to covariates when sampling from animal populations (Pollock, 2002; Wyatt, 2003; Royle, 2004; Forsyth et al., 2006). Additonal heterogeneity in the parameters can be allowed for by including additive Gaussian error terms in the models. Although we show in our work that allowing for heterogeneity in the detection model is important, it can be difficult to model when there is also a random effect in the animal density component of the model. Therefore, our

preference is to allow for variation in detection probability through a covariate model only.

Spatial models can give us insight into the patterns in the distribution of fish within a stream or across a region and are important for prediction and mapping the distribution of animals. Further, when modeling the relationship between density and covariates, ignoring spatial correlation will tend to exaggerate the effects of covariates on density. A number of authors discuss models that account for spatial correlation in count data, although few simulateously allow for imperfect detectability of animals. Royle et al. (2002) used an exponential covariance function (see Cressie, 1993) for bird counts, while Wikle (2002) discusses a more computationally efficient way of fitting similar geostatistical models. For other types of spatial count data drawn from discrete units or areas, for example disease modeling, conditional autoregressive (CAR) models have been used for some time (see Banerjee et al., 2004); Lichstein et al. (2002) apply a CAR model to bird point counts. For stream fish data, Wyatt (2003) incorporates estimation of detectability into a model with a simple autoregressive structure for the spatial dependence for modeling density at stream sections of 50m length.

Parameter estimation for the complex hierarchical models that we explore in this paper can be carried out within a Bayesian framework using Markov chain Monte Carlo (MCMC) algorithms. Only for the simplest models can we estimate parameters using maximum likelihood estimation (e.g., Royle, 2004; Dorazio et al., 2005). In other cases computing the likelihood involves numerically integrating out 'nuisance' parameters, which makes finding maximum likelihood estimates impractical. On the other hand, within a Bayesian framework, MCMC algorithms based on Gibbs sampling (Geman and Geman, 1984) and Metropolis-Hastings (Hastings, 1970) are relatively straightforward to apply for the hierarchical models considered in this paper. A nice introduction to Bayesian MCMC modeling for ecologists is provided by Link et al. (2002). The software package WinBUGS (Spiegelhalter et al., 2003), with GeoBUGS (Thomas et al., 2002) for spatial modeling, can be used for most models, avoiding the need for programming. However, care must be taken in using Bayesian MCMC. Bayesian modeling requires the specification of prior distributions for the parameters, through which we may choose to incorporate past knowledge. Even if we largely avoid using prior information by choosing vague or non-informative priors, in certain cases the results may not be robust to varying the particular form of prior (Gelman, 2006). At present, model selection can be much more difficult in a Bayesian framework. Often we may wish to choose the 'best' model from a set of plausible, competing models for a data set, and although Bayesian model selection criteria exist (e.g., Spiegelhalter et al., 2002), most are not well suited to the complex

mixed-effects models that we examine. However, versions of the posterior predictive loss criterion of Gelfand and Ghosh (1998) have proved useful for mixed models (e.g., Ghosh and Norris, 2005), and we apply such a criterion in this work. The application of Bayesian MCMC algorithms is perhaps the only practical approach available and a number of authors have successfully fitted models in this way to count and removal data, e.g., Royle et al. (2002) for avian points counts, Wyatt (2003) for fish removal data, and Forsyth et al. (2006) for brushtail possum removal data.

We give an overview of the hierarchical modeling framework for multiple pass repeated count and removal survey data from a sample of habitat units. Although the framework is the same for both types of sampling, we focus on repeated count surveys. These surveys have received less attention in the literature, yet the single-site research implies that analysis may be more difficult. We propose the application of CAR models for modeling spatial dependence in the unit abundances for surveys of stream fish populations. We also discuss Bayesian MCMC methods for estimation and prediction along with approaches to Bayesian model selection and their limitations for our models. As an example, we apply the models to an intensive multiple count survey of juvenile Coho Salmon in McGarvey Creek, Northern California. This example illustrates some of the potential problems in parameter estimation. In particular, when few passes are made of each unit, it can be difficult to distinguish heterogeneity in density from heterogeneity in detection probability by means of including random effects in both components of the model. Nevertheless, through a simulation study we demonstate the importance of allowing for heterogeneity in detection probability, particularly with respect to estimating total stream abundance. We recommend that this be done by including covariates in the detection model.

2.2 Hierarchical Models for Multiple Count Data

2.2.1 Modeling Abundance and Detection

Suppose we sample L habitat units from a stream, and for unit i, we make T_i independent counts of the fish population and let $T = \max_i T_i$. From our survey, we observe a data matrix $\mathbf{Y}$ with rows $\mathbf{Y}_i = [y_{i1}, \ldots, y_{iT}]$, where y_{ij} is the number of animals seen on occasion j at site i, where $j = 1, \ldots, T_i$ and $i = 1, \ldots, L$, and y_{ij} is an empty cell ('missing data' for computational purposes) for $j > T_i$. In addition to the independence of the multiple counts, we assume that the population in each unit is closed to immigration, emigration, births and deaths. The counts for each unit can then be assumed to be conditionally independent binomial random variables, with the exact form of the

model depending on whether we have repeated count or removal data. If the population in unit i prior to any removals is N_i and the probability of detecting an individual in this unit on pass j is p_{ij}, for count data we have

$$y_{ij}|N_i, p_{ij} \sim \text{Binomial}\left(N_i, p_{ij}\right). \tag{2.1}$$

For removal data, we need only replace N_i in (2.1) with $N_i - M_{ij}$, where $M_{ij} = \sum_{j'=1}^{j-1} y_{ij'}$ is the number of removals prior to occasion j, with $M_{i1} = 0$. The removal version has been used by Wyatt (2002) and Forsyth et al. (2006), while a less general formulation of Equation (2.1) was considered by Royle (2004) for repeated count data.

Turning to models for population size, we model the unit abundances, N_i, as

$$N_i|\lambda_i \sim \text{Poisson}(\lambda_i),$$

and we allow for different sized units by writing $\lambda_i = A_i d_i$, where A_i is the surface area of the unit and d_i the mean density of fish per unit of area. To incorporate covariates and spatial correlation, we follow Diggle et al. (1998) and subsequent authors by modeling the natural logarithm of density as the sum of a trend function μ_i and a zero-mean Gaussian process ϕ_i

$$\log(d_i) = \mu_i + \phi_i. \tag{2.2}$$

μ_i can either be modeled as a linear function of covariates, $\mu_i = \mathbf{X}_i'\boldsymbol{\alpha}$ say, or a non-linear function of covariates, $\mu_i = g(\mathbf{X}_i, \boldsymbol{\alpha})$, where $\mathbf{X}_i$ is the vector of covariates measured on unit i and $\boldsymbol{\alpha}$ is the corresponding parameter vector.

In some cases the log link function used in Equation (2.2) is not the most realistic choice. For example, there may exist a maximum possible density due to constraints on available food in a habitat unit. The log-normal model for mean abundance is not appropriate in this case as density d_i is unconstrained. In order that it not exceed some maximum, K say, a logistic model can be used:

$$\begin{aligned} & \log\left(\frac{d_i}{K - d_i}\right) = \mu_i + \phi_i \\ \Rightarrow \quad & d_i = \frac{K}{1 + \exp(-\mu_i - \phi_i)} \end{aligned}$$

In practice, it may only be possible to fit the logistic model in high density situations when at least some of the units are near carrying capacity. In those cases it may also be preferable to also model

K as a function of resource variables.

The second component of our modeling involves constructing a model for detection probability, p_{ij}. Detection probability will vary between units, as did abundance, and we wish to allow for such heterogeneity. While some authors have made use of random effects models for detection probability, e.g., beta models (Wyatt, 2002) and logistic-normal models (Wyatt, 2003; Forsyth et al., 2006), in our experience it can be difficult to distinguish random variation in detection from that due to abundance unless there are many passes or a very large number of survey units, which is often not possible for surveys of riverine fish. Instead, our preference is to model heterogeneity in detection through a covariate model alone. Employing the logit link function, we can express the detection probability for unit i on pass j, p_{ij} in terms of a covariate function,

$$\text{logit}(p_{ij}) = \nu_{ij}, \tag{2.3}$$

where ν_{ij} is some linear or non-linear function of covariates measured on pass j of unit i. In many situations, covariate values will vary between but not within the units, and we can write $p_{ij} = p_i$ and thus $\nu_{ij} = \nu_i$. In this case we may employ a linear regression model, $\nu_i = \mathbf{Z}_i'\boldsymbol{\beta}$, with a covariate vector for unit i of $\mathbf{Z}_i$ and parameter vector $\boldsymbol{\beta}$. The effect that ignoring heterogeneity in detection has on estimates of abundance and other model parameters is examined in a simulation study in Section 2.6.

2.2.2 Spatial Modeling of Fish Density

Spatial dependence in fish densities may be present due to unmeasured covariates or because of some underlying biological process. For example, when some species colonize a stream, the population may intially cluster around a small number of units of high quality habitat. As the population grows, fish will disperse throughout the stream, but may be more likely to settle in suitable habitat closer to the original units than those further away. Accounting for the spatial dependence in our model, we not only gain a better understanding of how fish populations are distributed in a stream, but we also improve parameter estimates when modeling the relationship between density and measured covariates, and improve the accuracy and precision of predictions at unmonitored locations. The latter is particularly important if we wish to map fish density, at both local and regional scales (see Discussion, Section 2.7).

Following Diggle et al. (1998), we introduce the spatial dependence through modeling the

Gaussian error term ϕ_i of the abundance component of the model. For point count data in two-dimensional continuous space, the most common approach is to assume an isotropic geostatistical model, such as the exponential model or the Matérn class of models (Cressie, 1993), which define a parametric form for the covariance between values of ϕ_i at different locations (e.g., Royle et al., 2002; Wikle, 2002). If we write $\boldsymbol{\phi} = [\phi_1, \ldots, \phi_L]'$, then in a Gaussian geostatistical model, $\boldsymbol{\phi}$ has a multivariate normal distribution,

$$\boldsymbol{\phi} \sim \text{MVN}(\mathbf{0}, \boldsymbol{\Sigma}_\phi).$$

The elements of $\boldsymbol{\Sigma}_\phi$, the covariances between values at different sites, are expressed as a decreasing function of the distance between two sites, $\text{cov}(\phi_i, \phi_i') = C(r_{ii'})$ say, where $r_{ii'}$ is the Euclidean distance between units i and i'. Stream data is essentially in one-dimensional space, but this does not pose any problems as instead of Euclidean distance we can work with the total linear stream distance between two units, that is, the distance between the centers of two units measured by following the line of the stream. Further, we will not concern ourselves here with anisotropty, i.e., strength of spatial dependence varying with direction.

An alternative and more suitable approach for data from contiguous units or areas is to fit Gaussian conditionally autoregressive (CAR) models (Besag, 1974). These have been used for some time for spatial modeling of regional disease rates and other medical data (e.g., Pascutto et al., 2000; Gelfand and Vounatsou, 2003), but have seen only limited use in wildlife applications. The essence of these models is that conditionally each ϕ_i is a weighted average of the values in neighboring units. Following Banerjee et al. (2004) we begin by defining a symmetric $L \times L$ matrix of weights, $\mathbf{W}$, with elements $w_{ii'}$. The simplest case would be defining $w_{ii'} = 1$ if units i and i' are adjacent, and zero otherwise. When the units are pools or runs in a stream, it may be that one neighboring habitable unit is closer than another, with sections of uninhabitable stream between them, and in this case we could choose $w_{ii'} = 1/r_{ii'}$ for neighboring units and zero otherwise, giving the closer neighbor greater weight. Now define $w_{i.} = \sum_{i'=1}^{L} w_{ii'}$, the row totals of $\mathbf{W}$, and let the vector $\boldsymbol{\phi}$ that omits ϕ_i be given by

$$\boldsymbol{\phi}_{-i} = [\phi_1, \ldots, \phi_{i-1}, \phi_{i+1}, \ldots, \phi_L]'.$$

A proper Gaussian CAR model then follows by defining the distribution of ϕ_i conditionally as

$$\phi_i | \boldsymbol{\phi}_{-i} \sim \text{N}\left(\gamma \sum_{i' \neq i} \frac{w_{ii'}}{w_{i.}} \phi_{i'}, \frac{\sigma_\phi^2}{w_{i.}}\right). \tag{2.4}$$

One of the most attractive features of this model is its ease of implementation as part of an MCMC algorithm for Bayesian hierarchical modeling. Unlike the parametric geostatistical models discussed above, using Equation (2.4) as a prior distribution leads to a full conditional distribution for ϕ_i that allows updates in Gibbs sampling to be done without requiring the inversion of large matrices, making the algorithm much more efficient for large data sets.

The parameter γ appears to have an appealing interpretation as a measure of the strength of the spatial dependence. However, as Banerjee et al. (2004) illustrate, even values of γ close to 1 do not necessarily imply high autocorrelation, and so meaningful intepretation of γ is difficult. Further, the proper CAR model (2.4) gives the conditional mean of ϕ_i as a *proportion* of a weighted average of its neighbors, which itself makes interpretation difficult. For these reasons, an alternative model, known as the intrinsic CAR model, fixes $\gamma = 1$, leading to each ϕ_i being expressed simply as a weighted sum of its neighbors. Nevertheless, $\gamma = 0$ does imply spatial independence, and inclusion of this parameter does allow for varying degrees of spatial correlation and we include γ in our models.

We note briefly here that some authors (e.g., Royle et al., 2002; Wikle, 2002) include an additional spatially independent error term, τ_i, in the density model

$$\log(d_i) = \mu_i + \tau_i + \phi_i.$$

where τ_i are independent, identically distributed Gaussian random variables with mean 0 and variance σ_τ^2. This term can be thought of as accounting for between unit hetereogeneity (Banerjee et al., 2004) or for such factors as observer differences (Wikle, 2002). However, distinguishing between ϕ_i and τ_i in practice can be difficult and we do not consider this model any further, although in Section 2.5 we compare models with spatial structure, $\log(d_i) = \mu_i + \phi_i$, with those with spatial independence, $\log(d_i) = \mu_i + \tau_i$.

2.3 Bayesian MCMC Modeling and Model Selection

In Bayesian modeling, the goal is to compute the form of the distribution of the model parameters given the data and some prior information. The distribution thus determined is known as the

posterior distribution and for a parameter vector $\boldsymbol{\theta}$, this is found by solving the equation

$$\wp(\boldsymbol{\theta}|\mathbf{Y}) = \frac{\wp(\mathbf{Y}|\boldsymbol{\theta})\wp(\boldsymbol{\theta})}{\int \wp(\mathbf{Y}|\boldsymbol{\theta})\wp(\boldsymbol{\theta})d\boldsymbol{\theta}}, \tag{2.5}$$

where $\wp(\theta|\mathbf{Y})$ is the joint posterior density, $\wp(\mathbf{Y}|\boldsymbol{\theta})$ is the likelihood of the data and $\wp(\boldsymbol{\theta})$ is a prior density through which prior knowledge about $\boldsymbol{\theta}$ can be incorporated. The idea of allowing previous research and expert opinion to inform our understanding of the present data seems very appealing, particularly when the present information is poor as is often the case in wildlife population research. However, deciding how much weight to give to these two sources of information is an inherently subjective process, and clear justification must be given for any informative priors. In cases where there is no useful prior information about a parameter, or if we wish the focus of our analysis to be on the information contained in the present data alone, we can select non-informative or vague priors so that the form of the posterior is approximately determined by the likelihood alone. This is the approach we take in the present work. See Gelman et al. (2004) for a discussion on the practicalities of prior specification.

For most useful models, the solution to Equation (2.5) does not have a closed form. Instead, we can use MCMC algorithms such as Gibbs sampling (Geman and Geman, 1984), Metropolis-Hastings (Hastings, 1970), and slice sampling (Neal, 2003) to generate samples from the posterior distributions of the parameters. As an example, a common approach is to combine Gibbs sampling with Metropolis steps for certain parameters. Suppose we have parameter vector $\boldsymbol{\theta} = [\theta_1, \theta_2]'$ and data $\mathbf{Y}$. From Bayes' theorem, Equation (2.5) the posterior distribution of $\boldsymbol{\theta}$ will be given by

$$\wp(\theta_1, \theta_2|\mathbf{Y}) \propto \wp(\mathbf{Y}|\theta_1, \theta_2)\wp(\theta_1, \theta_2)$$

where again $\wp(\mathbf{Y}|\theta_1, \theta_2)$ is the likelihood and $\wp(\theta_1, \theta_2)$ the prior density. Assuming independent priors, $\wp(\theta_1, \theta_2) = \wp(\theta_1)\wp(\theta_2)$ the *full conditional* for θ_1 is given by

$$\wp(\theta_1|\mathbf{Y}, \theta_2) \propto \wp(\mathbf{Y}|\theta_1, \theta_2)\wp(\theta_1)$$

and that for θ_2 has the same form. If the right hand side of this equation is the kernel of a density with closed form, then $\wp(\theta_1|\mathbf{Y}, \theta_2)$ is known and at step b of the MCMC algorithm, we generate $\theta_1^{(b)}$ as a random sample of size 1 from the distribution with density $\wp(\theta_1|\mathbf{Y}, \theta_2^{(b-1)})$. If the full conditional for θ_2 also has closed form, then the bth sample is generated randomly from $\wp(\theta_2|\mathbf{Y}, \theta_1^{(b)})$.

Now suppose the density for θ_2 does not have closed form. Instead of sampling from the full conditional, we can insert a Metropolis step into the algorithm. Let $q(x|\theta_2^{(b-1)})$ be a density valid for all possible values of $\theta_2^{(b-1)}$ such that $q(x|\theta_2^{(b-1)}) = q(\theta_2^{(b-1)}|x)$. Then the Metropolis step proceeds as follows: (1) Sample θ_2^* from $q(x|\theta_2^{(b-1)})$; (2) Compute the ratio

$$r = \frac{\wp(\mathbf{Y}|\theta_1^{(b)}, \theta_2^*)\wp(\theta_2^*)}{\wp(\mathbf{Y}|\theta_1^{(b)}, \theta_2^{(b-1)})\wp(\theta_2^{(b-1)})};$$

(3) Let $\theta_2^{(b)} = \theta_2^*$ with probability $\min(r, 1)$ and $\theta_2^{(b)} = \theta_2^{(b-1)}$ otherwise. In theory, samples generated from such an MCMC algorithm will converge to the posterior distributions of the parameters, although choosing the proposal density q and its parameters in Metropolis steps requires some care (see Gelman et al., 2004)

After discarding data generated prior to convergence of the algorithms (the "burn-in" period), the MCMC samples can be used to estimate posterior summary statistics such as means, standard deviations, quantiles or the form of densities themselves (for example, using kernel density estimation). The precision of the posterior means can be estimated using the methods of time series analysis which allow for correlation in the series of generated values, for example by estimating the spectral density at zero (see Monahan, 2001). Convergence itself can be determined by generating samples from multiple Markov chains begun with widely dispersed initial values and comparing the behavior of the chains. This can be done by simultaneous plots of the values of chains for each parameter, with good mixing of the chains implying convergence. In addition we can also compute the Gelman-Rubin scale-reduction factor (Gelman and Rubin, 1992), which compares variation within the chains to variation across the chains. Convergence will lead to values of this factor close to 1.

We would also like to be able to select a model from a set of competing models, but although Bayesian model selection criteria exist, they can behave poorly for the types of models we are examining. The Akaike Information Criterion (AIC) is typically used for comparing models fitted by maximum likelihood estimation, and is computed by adding a penalty term for model complexity to the deviance evaluated at the MLEs (Akaike, 1973). A Bayesian version of the AIC statistic has been proposed by replacing the MLEs with the posterior means of the parameters (Gosky, 2004), but neither criterion has a formal theoretical justification for complex hierarchical models. Alternatively, Spiegelhalter et al. (2002) proposed a generalization of the AIC for Bayesian modeling called

the Deviance Information Criterion (DIC). Although the DIC appears to work well in many situations, it has undesirable properties for mixed effects models, including those we consider here). In particular, the DIC penalty term can become negative (Spiegelhalter et al., 2002) and thus more complex models are rewarded.

Gelfand and Ghosh (1998) proposed a criterion based on the minimizing a posterior predictive loss function. Their criterion is more general than the DIC of Spiegelhalter et al. (2002), and with an appropriate choice of loss function, will not suffer the problem of a negative penalty term for complex hierarchical models. It is also computationally efficient to calculate within an MCMC algorithm for Bayesian modeling. For our model comparisons, we adopt the mean square predicted error loss function ($MSPE$) on the log-scale used by Ghosh and Norris (2005) for selecting competing models with heterogeneity in detectability for capture-recapture data. Recall that in modeling repeated count data from multiple sites, we assume a binomial model for the counts, $y_{ij} \sim \text{Bin}(N_i, p_i)$, $i = 1, \ldots, L$, $j = 1, \ldots, T_i$. Again denoting the full observed data matrix by $\mathbf{Y}$, we define $\mathbf{Y}^{rep}$ to be a replicate of $\mathbf{Y}$ predicted under the same model. These replicates are easily generated within our MCMC algorithm by sampling from the posterior distribution $\wp(\mathbf{Y}^{rep}|\mathbf{Y})$, that is, at the bth MCMC iteration, values y_{ij}^{rep} are sampled from a binomial distribution with parameters $N_i^{(b)}$ and $p_i^{(b)}$, the current values of those parameters. We use the $MSPE$ loss function to measure departures of $\mathbf{Y}^{rep}$ from $\mathbf{Y}$. For our models we compute the $MSPE$ on the log-scale as

$$MSPE = \frac{1}{\sum_{i=1}^{L} T_i} \sum_{i=1}^{L} \sum_{j=1}^{T_i} \left[\log(y_{ij}^{rep} + 0.5) - \log(y_{ij} + 0.5)\right]^2. \tag{2.6}$$

The factor of 0.5 added to each value was suggested by Ghosh and Norris (2005) to avoid computing the logarithm of 0 when no animals are observed. Finally, the criterion itself is given by the posterior predictive expectation of the $MSPE$,

$$D_{GG} = \text{E}[MSPE|\mathbf{Y}], \tag{2.7}$$

which we estimate as for any other model parameter, by computing the posterior mean of the MCMC samples of $MSPE$ following convergence of the algorithm.

2.4 Prediction of Abundance at Unsurveyed Units

One of the goals of this type of monitoring is prediction of mean density or probability of fish presence for unmonitored sites within streams or, if modeling at a regional level, at unmonitored streams. With CAR models applied to discrete units, provided that covariate measurements are also available at unmonitored locations, estimation of density at these locations can be undertaken as a natural part of the MCMC algorithm. Suppose that we are using Gibbs sampling, with Metropolis steps when the full conditional distributions do not have closed form. Consider the full conditional distribution of N_i, the abundance in unit i of a stream. If data $\mathbf{Y}_i$ are available for site i, and A_i is the surface area of the unit i and d_i the fish density, this is given by

$$\wp(N_i|\mathbf{Y}_i, p_i, d_i) \propto \wp(\mathbf{Y}_i|N_i, p_i)\wp(N_i|d_i)$$

where $\wp(\mathbf{Y}_i|N_i, p_i)$ is the likelihood component for site i and $\wp(N_i|d_i)$ is the Poisson$(A_i d_i)$ density. We note that the full conditional of the unit densities is a function of the density covariate matrix, $\mathbf{X}$, and the prediction of N_i will also depend on covariate values, and thus we also require these to be measured at the unmonitored units. Here the full conditional for N_i does not have the form of a familiar distribution and we would update N_i at each iteration using a Metropolis-Hastings step. However, if no data are available for a unit, then the full conditional is just $\wp(N_i|d_i)$ and we update N_i by drawing a sample from a Poisson distribution with mean $A_i d_i$, using the value of d_i from the previous iteration. Thus the posterior distribution of abundances at units with no data follow from the MCMC algorithm in the same way as that for other unit abundances and all model parameters. In fact, sites for which no animal survey data were recorded should not be excluded from the modeling. Even though models with gaps in the neighborhood structure can be fitted, estimation of abundance at points adjacent to an unsurveyed unit may be improved by making use of its correlation with units on the other side of the gap, this correlation being induced through its direct dependence on the unsurveyed neighbor.

2.5 Example: Juvenile Coho Salmon in McGarvey Creek

In the summer of 2003, a team from Humboldt State University undertook an intensive repeated count survey of juvenile Coho Salmon in McGarvey Creek, a tributary of the Klamath River, Northern California (Hankin, 2004). The surveyed part of McGarvey Creek comprised 206 habitable units

(167 pools and 39 runs) spaced over a 3.6 km section of the stream. Sixty-five units were monitored with 3 passes, and 133 with a single pass. For the remaining 8 units, only incomplete counts were obtained for logistical reasons, and these units were considered unsurveyed in the modeling. Each unit had its surface area, A_i, estimated by multiplying its length by the mean of three measures of its width. Note that, although unavailable for this example, precise measurements of surface area for some units could be used to account for the measurement error in the area estimates in our models. In our work, we assume that all covariates are measured without error. As well as surface area, the Humboldt crew recorded two other habitat variables, unit cover class and maximum unit depth. The cover score is an ordinal variable with values 1 (high cover), 2 (moderate), and 3 (low). Units with higher cover are thought to be more suitable habitat for juvenile coho and so we included cover as a covariate for density model, leading to the μ_i component of Equation (2.2), the model for $\log(d_i)$, having the following form:

$$\mu_i = \alpha_0 + \alpha_1 C_{2i} + \alpha_2 C_{3i}.$$

Here C_{2i} and C_{3i} are dummy variables taking the value 1 for units with cover class 2 or 3 respectively and zeros otherwise. Because it is plausible that juvenile Coho are harder to detect at greater depths where visibility may be poorer, we examined a model which allowed unit detection probability p_i to vary with maximum unit depth. Thus $\nu_{ij} = \nu_i$ in Equation (2.3), and we write

$$\nu_i = \beta_0 + \beta_1 D_i,$$

where D_i is the measured maximum depth of unit i.

For comparison, we fit a sequence of models of increasing complexity to the McGarvey Creek data: (1) no covariates ($\alpha_1 = \alpha_2 = \beta_1 = 0$); (2) cover as density covariate, no covariate for detectability ($\beta_1 = 0$); (3) cover as density covariate, depth as detectability covariate. For each of these three models, we fitted both spatial independence and spatial dependence versions. To model spatial dependence, we used a proper CAR model with neighbor weights equal to 1. We also looked at models with weights inversely proportional to the distance between unit centers, but found no apparent difference in parameter estimates nor model fit, and do not present these results here. We attempted to fit models with random effects in detection, but we encountered some problems which we discuss below.

As well as the model parameters, we also estimated the unit occupancy rate (MacKenzie et al.,

2002), $\psi = \sum_{i=1}^{L} I(N_i > 0)/L$ at each iteration of the Bayesian modeling, where $I(x)$ takes the value 1 when x is true and zero otherwise, and total abundance as $N = \sum_{i=1}^{L} N_i$. This estimate includes predictions at the eight unsurveyed sites, which for our example, where detection probability is high, are the greatest source of uncertainty in total abundance. As a summary statistic for detection, we computed the $\bar{p}$, the average of the p_i, although we note that for models with $\beta_1 = 0$, detection does not vary across units.

Bayesian MCMC model fitting was done using WinBUGS 1.4.1 (Spiegelhalter et al., 2003). We chose independent, non-informative proper priors for the parameters. In particular, we used a Uniform(0,20) prior for σ_ϕ in the CAR model and σ_τ in the spatial independence model, the proper CAR correlation parameter γ had a Uniform(-1,1) prior, and all other parameters had N(0,0.001) priors. As a check, alternative priors were tried for some models, such as a beta prior for p in Models 1 and 2, and a vague gamma prior for $1/\sigma_\phi^2$, but these had little effect on the posterior distributions of the parameters. For model selection, we computed the criterion of Gelfand and Ghosh (1998) based on the mean mean square predicted error loss, Equations (2.6) and (2.7).

Three chains with different starting values were run for each model and convergence was assessed using the Gelman-Rubin statistic and by examining the behavior of the trace plots of the chains. The number of iterations required for convergence (the burn-in period) varied from model to model, as did the total number of iterations. Our aim was to generate sufficiently large samples so that posterior means of parameters were precisely estimated. Due to auto-correlation in the chains, particularly for the more complex models, up to one million iterations were required following burn-in for some models. We monitored the precision of the MCMC estimates of the parameters using estimates of the spectral density of the chains at 0. Gelman-Rubin statistics and MCMC precision estimates (not presented here) were computed using the functions in the coda package version 0.9-1 for R (Plummer et al., 2004).

Posterior means of parameters with 95% credible intervals (the 2.5% and 97.5% quantiles of the posterior distribution) are presented in Table 2.1. In order to see more clearly the effect of cover class on Coho density, we computed estimates of the percentage difference in density between classes from α_0, α_1 and α_2 at each iteration of the MCMC algorithm. The comparisons of cover classes 1 and 2, and of 1 and 3 are given by $100(e^{\alpha_1} - 1)$ and $100(e^{\alpha_2} - 1)$ respectively, while $100(e^{\alpha_2 - \alpha_1} - 1)$ compares classes 2 and 3. Posterior means and 95% credible intervals are given in Table 2.2. For all models, while the evidence for a difference between cover classes 1 and 2 is not great, mean density in cover classes 1 and 2 is significantly higher than cover class 3. This is true for both

Table 2.1: Posterior means and 95% credible intervals of parameters for modeling of McGarvey Creek data for both spatial independence and spatial dependence models, along with D_{GG} model selection statistics for each of the fitted models.

	Spatial independence			Proper CAR		
parameter	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
α_0	–1.49 (–1.66,–1.31)	–0.98 (–1.38, –0.58)	–0.92 (–1.33,–0.51)	–1.45 (–1.71,–1.20)	–1.13 (–1.51, –0.76)	–1.09 (–1.48,–0.70)
α_1	–	–0.35 (–0.80, –0.10)	–0.41 (–0.87,0.04)	–	–0.21 (–0.56, 0.14)	–0.26 (–0.62,0.11)
α_2	–	–0.79 (–1.24,–0.35)	–0.88 (–1.34,–0.43)	–	–0.54 (–0.92,–0.17)	–0.61 (–1.01,–0.21)
γ	–	–	–	0.82 (0.64,0.93)	0.79 (0.59, 0.92)	0.77 (0.54,0.91)
σ^2_ϕ	–	–	–	0.91 (0.60,1.33)	0.92 (0.60, 1.34)	0.96 (0.62,1.41)
σ^2_τ	0.89 (0.67,1.17)	0.82 (0.61,1.08)	0.82 (0.61,1.09)	–	–	–
β_0	1.10 (0.74,1.44)	1.10 (0.72, 1.43)	2.02 (1.24,2.73)	1.12 (0.78,1.43)	1.11 (0.78, 1.42)	1.63 (0.72,2.43)
β_1	–	–	–1.20 (–2.06,–0.26)	–	–	–0.63 (–1.59,0.44)
$\bar{p}$	0.75 (0.68,0.81)	0.75 (0.67,0.81)	0.77 (0.70,0.82)	0.75 (0.69,0.81)	0.75 (0.68,0.80)	0.77 (0.70,0.82)
ψ	0.91 (0.89,93)	0.91 (0.90,0.93)	0.91 (0.89,0.93)	0.92 (0.90,0.94)	0.92 (0.90,0.94)	0.91 (0.89,0.93)
$\sum N_i$	2490 (2260,2820)	2600 (2310, 3050)	2700 (2380,3200)	2450 (2260,2710)	2520 (2300,2810)	2530 (2310,2850)
D_{GG}	0.123	0.123	0.108	0.124	0.124	0.113

Table 2.2: Posterior means and 95% credible intervals for percentage differences in Coho density between the three cover classes.

	Spatial independence		Proper CAR	
parameter	Model 2	Model 3	Model 2	Model 3
1-2	46 (–9,123)	55 (–3,139)	25 (–13,75)	32 (–11,87)
1-3	127 (42,247)	148 (54,282)	75 (18,151)	88 (24,176)
2-3	57 (15,112)	61 (17,117)	41 (8,81)	44 (10,86)

independence and CAR versions of the models, but it is clear from Table 2.2 that the magnitude of the effect appears stronger when we do not account for spatial dependence. For example, in Model 3, ignoring spatial dependence would lead us to believe that density is 148% (95% CI: 54-282) greater in class 1 units than class 3 units, while this is reduced to only 88% (24-176) when we account for spatial dependence.

Based on the posterior distribution of β_1 under the CAR model, there appears to be no evidence that the logit of detection is linearly related to the maximum depth of a unit (Table 2.1), with the credible intervals for this parameter including zero for all Model 3. As with the density component of the model, the estimated coefficient of depth is attenuated when we allow for spatial dependence, from –1.20 to –0.63. Not surprisingly for a stream with good abundance and high detectability, the estimates of occupancy rates, ψ were both high (0.91-0.92) and quite precise. The mean individual detection probability, $\bar{p}$, was consistently estimated to be around 0.75 (Table 2.1), although it was slightly higher for Model 3. When detection is estimated to be lower, the estimate of abundance generally becomes higher.

Given these results, it is interesting that the model selection criterion based on the mean square predicted loss, D_{GG}, selects Model 3 as the best model. The parameter β_1 does not appear significantly different from zero under a CAR model, yet it appears that there is some type of heterogeneity in detection and at least in terms of prediction, Model 3 is the best of those we consider. It is also somewhat surprising that D_{GG} barely distinguishes Models 1 and 2 when the cover class otherwise appears to be an important predictor of fish density (Table 2.2). The posterior means and credible intervals for the spatial dependence parameter, γ, show that there is significant spatial dependence

Table 2.3: Posterior means and 95% credible intervals for abundance estimated at the eight unsurveyed sites for spatial independence and CAR versions of Model 3.

unit number	Spatial independence	Proper CAR
42	67	22
	(6,269)	(2,78)
66	44	26
	(4,180)	(3,89)
71	21	17
	(1,82)	(2,56)
85	69	43
	(6,279)	(6,142)
108	41	55
	(4,167)	(9,177)
124	34	32
	(3,135)	(4,105)
130	16	12
	(1,66)	(1,39)
170	9	15
	(0,37)	(1,50)

in unit density across the stream, but we found that the D_{GG} values for the CAR models are no better than those for the spatial independence models. However, this result is not surprising. When detection probability is high as it is here, it is not difficult to get very good estimates of N_i, the unit abundances, and prediction at surveyed units will be little improved by accounting for spatial dependence. Prediction at *unsurveyed* units will improve greatly, however (see Table 2.3), but these units contribute nothing to the calculation of D_{GG}.

As we have noted, with high detection probabilities, estimation of N_i at surveyed sites will be quite precise and, therefore, the greatest factor affecting the estimate of abundance in this example was the prediction at the eight unmonitored sites. It is, therefore, important that we are able to make these predictions as precise as possible. To illustrate how accounting for spatial dependence improves prediction at unmonitored sites, Table 2.3 compares estimates of abundance at these sites computed from versions of Model 3 which assume either spatial independence or use a CAR model for spatial dependence. Although for either model these estimates were not particularly precise, the 95% credible intervals are nevertheless somewhat narrower for the models in which spatial dependence was accounted for. Also, the posterior means themselves are quite different between

the independence and the CAR models, illustrating the influence that neighboring units have on prediction when a spatial model is used.

An important goal of modeling fish populations is creation of maps showing the spatial distribution of fish across the stream or region. Figure 2.1 shows schematic maps of ϕ_i, the spatially dependent errors in the density model, and λ_i, expected unit abundances for the CAR version of Model 3. Colored lines show the density at each unit, with the thickness of the line representing the length of the unit. Lighter colors represent higher fish density. Black bars show uninhabitable areas of stream, mainly runs. The map of the posterior means of the ϕ_i shows that after allowing for cover class effects and unit area, there are clusters of high density around 1250 m, between 1750 and 2000 m, and from around 2250-3000 m along the stream. Precision of the map values is lower at the ends of the stream, as shown in the sd(ϕ_i) map. Translating back to the scale of abundance, there are small patches of high abundance at 1100-1500 m, 1750 m, 2300 m and 2750 m. The map of posterior standard deviations of the λ_i shows brighter colored lines for the eight units at which no data was collected and predicted was required.

Finally, we note that we attempted to fit models which included random effects in the detection component of the model. We found that it not only was very difficult to distinguish the different sources of variation (the MCMC values were highly negatively correlated), but that the models were less stable, converging more slowly and being more sensitive to choices of starting values, prior distributions and algorithms. The D_{GG} statistics were around 0.16 for these models, showing that they also provided a poorer fit to the data. The difficulty in distinguishing between the two sources of variation is perhaps not so surprising when we consider that information about detection comes only from the 65 units with three passes. It appears that with so few repeated passes, we require a larger sample of units to estimate well the detection component of the model. Although we do not recommend models with random effects in the detection probability model, such random variation is certainly present, and in the following section we present a small simulation study in which we attempt to determine how failure to explicitly model this variation affects estimation of other model parameters and total abundance.

2.6 Simulation Study

In the example in the preceeding section we fitted models to repeated count data of fish in stream habitat units. Some additional modeling showed that we can have difficulty distinguishing whether

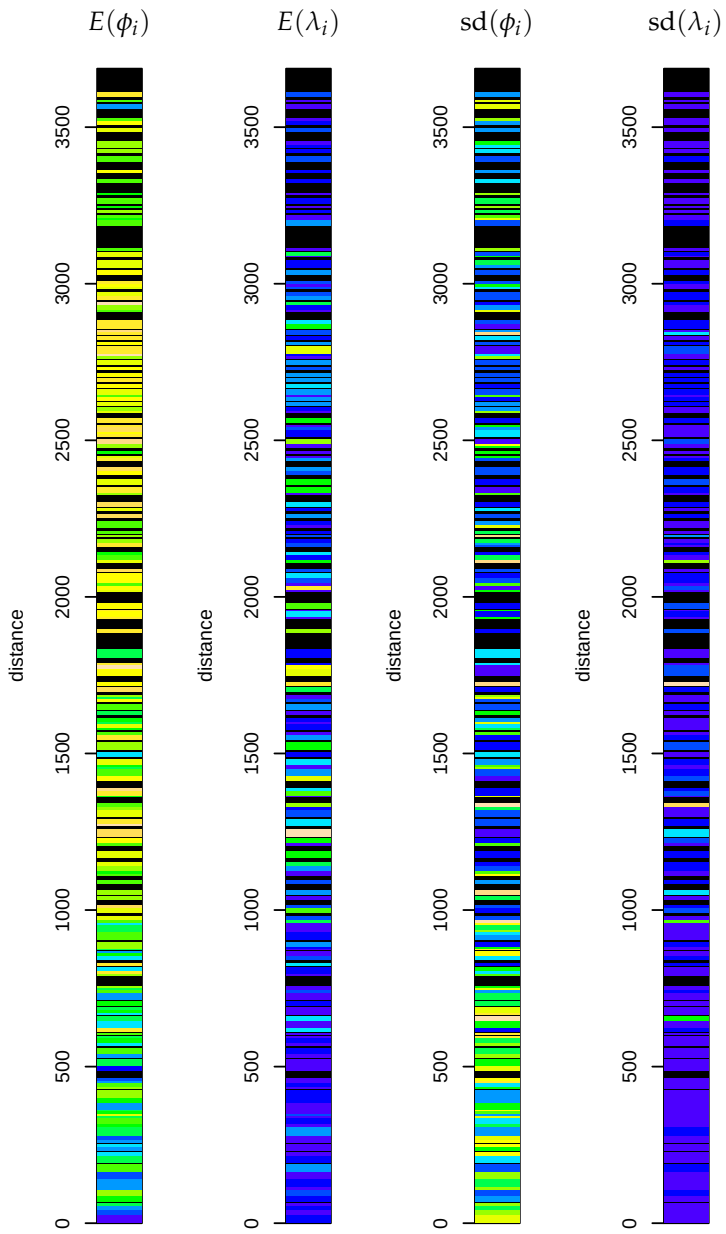

Figure 2.1: Maps of the posterior means of the residuals ϕ_i, expected unit abundances, λ_i, and their standard deviations for McGarvey Creek.

variation in the data is due to variation in abundance or detection probability. We chose to omit random effects in the detection model for our example, but it is unclear what effect this has on estimation of model parameters and total stream abundance. If there is an effect, its magnitude is almost certainly a function of the amount of random variation in detection probability across the stream. Further, we are interested in how estimation of parameters and abundance are in these circumstances affected by study design (for example, number of passes per unit), and detection probabilities. When detection probabilities are high, as was the case for McGarvey Creek, the observed counts are relatively consistent and close to the true abundance. Estimation of abundance under lower detection probabilities is less precise and likely to be more affected by ignoring random effects. In this section we present the results of a simulation study based on the results of the McGarvey modeling which aims to examine these issues. For a selection of study designs, we compare estimation of model parameters, mean detection probability, and abundance for models which assume no random variation in p when it is in fact present.

We generated data under the model without covariates in either the density or detection components, that is, $\mu_i = \mu$ in Equation (2.2) and $\nu_{ij} = \nu$ in Equation (2.3) for $i = 1, \ldots, L$ habitat units. As before, we include a spatially dependent between-unit random effect, ϕ_i, in the density model. For the detection model, we also add a Gaussian random effect, ϵ_i, and this model becomes

$$\text{logit}(p_i) = \nu + \epsilon_i,$$

where $\epsilon_i \sim \text{N}(0, \sigma^2_\epsilon)$. The random effect for density, ϕ_i, followed a proper Gaussian CAR model with weights equal to 1 for adjacent units and 0 otherwise.

The values of most of the parameters in the simulations were chosen to broadly reflect the results of the McGarvey Creek modeling. The density parameter μ was fixed at 1.5, which leads to moderate fish densities similar to McGarvey Creek, while $\sigma^2_\phi = 0.7$, a value based on the results of our additional modeling with detection random effects. The value of ν was set as either 0, or 1, which corresponds to mean detection probabilities of 0.5 and around 0.7. We attempted to perform simulations with even lower values of detection probability, but convergence was prohibitively slow due to the sparse data generated when detection is difficult. The random variation in ϵ_i was given either a 'low' value of $\sigma^2_\epsilon = 0.7$ (based on our additional modeling) or a 'high' value of 1.5, to examine the effect the degree of variation has on estimation.

We looked at survey designs with $L = 200$ contiguous survey units, and therefore similar in

size to the McGarvey Creek survey. For simplicity, units were equally spaced and of equal size. In our simulated designs, every second unit was surveyed with a single pass, and the remainder had either three (as per McGarvey Creek) or five passes, the latter being possibly the largest practical value for most stream fish surveys, while still allowing us to examine the effect of the number of passes on estimation.

For each combination of factors that we consider, we simulated data from $B = 125$ streams – with Bayesian MCMC modeling, any greater number of simulations would be impractical. The true fish abundances we generated were the same for all combinations of parameters we considered, while the detection probabilities only varied with ν and σ_ϵ^2. For each stream, we stored the posterior mean as a point estimate of the parameter along with the posterior standard deviation. The performance of the parameter estimators was measured by calculating estimates of the absolute bias and mean square error (MSE) for each parameter from the posterior means. As each stream had a unique true population size and set of unit detection probabilities, bias and MSE calculations were not possible for abundance and detection probability estimators. Instead, we calculated the mean difference of the posterior mean from the true population size, N, and the mean difference from the true stream mean detection probability, $\bar{p} = \sum_{i=1}^{L} p_i / L$, in order to determine how abundance and detection are affected by different values of the parameters. We also present the average of the posterior standard deviations of N and $\bar{p}$.

Convergence of the MCMC algorithms for each simulation was checked by examining Gelman-Rubin scale-reduction factor (Gelman and Rubin, 1992) for each parameter. For some combinations of parameters, particularly when $\nu = 0$, MCMC algorithms did not converge for certain streams. To minimize the impact convergence failure would have on the results, we excluded the results of 25 streams for which the Gelman-Rubin factor was greater than 1.15 for at least one set of simulations, and so results are based on 100 simulations only. Although these streams are not omitted at random, omitting the same 25 streams for all simulations means comparisons between the simulations should not be significantly affected. Examination of the results from the omitted streams for cases when models did converge show that their omission has only a small effect on the absolute values of the bias and MSE estimates.

The results of the simulations are presented in Tables 2.4 and 2.5. It is clear from Table 2.4 that biases in the estimators of μ and ν are a functions of both detection probability and heterogeneity in detection. There is a negative bias in the estimator of μ, and although this is small when $\sigma_\epsilon^2 = 0.7$, no larger than –0.05, it is substantially worse, up to –0.26, when $\sigma_\epsilon^2 = 1.5$. The bias is around twice

Table 2.4: Estimates of absolute bias and mean square error (MSE) from 100 simulations of data from surveys of $L = 200$ units, $T_i = 3$ or 5 passes per unit for 100 units, and $T_i = 1$ pass for 100 units. In all simulations, $\mu = 1.5$, $\sigma^2_\phi = 0.7$ and $\gamma = 0.85$.

	parameter	$T_i = 3$				$T_i = 5$			
		$\sigma^2_\epsilon = 0.7$		$\sigma^2_\epsilon = 1.5$		$\sigma^2_\epsilon = 0.7$		$\sigma^2_\epsilon = 1.5$	
		bias	MSE	bias	MSE	bias	MSE	bias	MSE
$\nu = 1$	μ	–0.021	0.017	–0.14	0.017	–0.014	0.016	–0.099	0.015
	γ	–0.066	0.008	–0.15	0.015	–0.064	0.008	–0.15	0.014
	σ^2_ϕ	0.13	0.028	0.44	0.042	0.13	0.028	0.42	0.041
	ν	–0.023	0.025	–0.034	0.041	–0.057	0.011	-0.17	0.022
$\nu = 0$	μ	–0.042	0.028	–0.26	0.021	–0.036	0.021	–0.20	0.019
	γ	–0.12	0.013	–0.26	0.022	–0.12	0.012	–0.26	0.020
	σ^2_ϕ	0.31	0.040	0.91	0.074	0.29	0.038	0.88	0.067
	ν	0.040	0.048	0.25	0.050	0.019	0.024	0.11	0.028

Table 2.5: Mean relative deviations from true values and standard deviations of $\hat{N}$ and $\hat{\bar{p}}$ from 100 simulations of data from surveys of $L = 200$ units, $T_i = 3$ or 5 passes per unit for 100 units, and $T_i = 1$ pass for 100 units. The mean of the 100 true N values was 1279, while the mean of the $\bar{p}$ was 0.71 ($\sigma^2_\epsilon = 0.7$) and 0.67 ($\sigma^2_\epsilon = 1.5$) for $\nu = 1$ and 0.50 (both values of σ^2_ϵ) for $\nu = 0$.

	parameter	$T_i = 3$				$T_i = 5$			
		$\sigma^2_\epsilon = 0.7$		$\sigma^2_\epsilon = 1.5$		$\sigma^2_\epsilon = 0.7$		$\sigma^2_\epsilon = 1.5$	
		dev	sd	dev	sd	dev	sd	dev	sd
$\nu = 1$	N	–20	52	–87	55	–13	36	–44	42
	$\bar{p}$	0.012	0.029	0.049	0.031	0.006	0.019	0.022	0.023
$\nu = 0$	N	11	161	–121	117	5	95	–53	86
	$\bar{p}$	0.009	0.054	0.060	0.050	0.004	0.034	0.026	0.034

as great for the lower value of ν. Bias in the ν estimator itself depends more on the true value of ν: when $\nu = 1$, bias is always small and negative and is not greatly affected by changes in σ^2_ϵ, whereas a positive bias is present when $\nu = 0$ that increases dramatically from a low value (< 0.05) to a very high value (up to 0.25) as σ^2_ϵ increases from 0.7 to 1.5. MSEs for these parameters increase as mean detection probability decreases, as we might expect, but are unaffected by σ^2_ϵ. Increasing the number of passes, T_i, from 3 to 5 provides additional information on the detection component of the model and thereby reduces the bias and MSE of ν.

It is less meaningful to consider the bias in σ^2_ϕ when heterogeneity in detection is not modeled. Estimates of σ^2_ϕ are necessarily larger than the true value as this parameter must account for all sources of variation, not just that in density. Thus, our results show that the 'bias' in the estimator increases as σ^2_ϵ increases as we expect. More interesting is the negative bias in the estimator of γ, the spatial dependence parameter. Although this parameter is intrinsically difficult to interpet, the fact that it is consistently underestimated and that the degree of underestimation depends on ν and σ^2_ϵ is of concern as it directly affects the quality of prediction and mapping of fish density. As ν decreases, we have less precise information about the density component of the model, and consequently less information for estimating the strength of spatial correlation. In our model, increasing σ^2_ϵ creates additional noise in the density model, making it more difficult to model spatial dependence. Increasing the number of passes has no effect on the bias and MSE of γ, although increasing L would certainly improve the estimation of γ.

In Table 2.5 we examine how the estimators of abundance and mean detection probability vary across the simulations. When variation in detection across the stream is low ($\sigma^2_\epsilon = 0.7$), the estimates of N and $\bar{p}$ are very close to the true values on average. The average true N for the 100 simulated streams was 1279, and a mean deviation from true N of up to 20 is less than 2%, very small relative to the size of the standard deviations. However, estimates of N shift much further from the truth when σ^2_ϵ increases to 1.5. In this case there is a significant underestimation of abundance and an overestimation of detection probability. This is the most important result of our simulation study: unmodeled heterogeneity in detection probability can lead to underestimation of abundance. We note also, though, that by increasing the number of passes from 3 to 5 for the multiple-pass units, we can mitigate to some extent the magnitude of underestimation. Finally, the results of Table 2.5 show that the standard deviations are larger when ν is lower, and decrease as the number of passes increase, as expected.

2.7 Discussion

Bayesian hierarchical modeling provides a very flexible framework for modeling data from repeated count and removal surveys of multiple habitat units in streams. At the local level, covariates measured on individual units are easily incorporated into models for density and detection, and spatial models can be used to allow for small-scale spatial dependence in the distribution of unit densities. We considered conditional autoregressive models for modeling the spatial structure. Although there are questions concerning the interpretation of the correlation parameter in these models (Banerjee et al., 2004), these models are structurally intuitive and, importantly, lead to fast MCMC algorithms for Bayesian modeling. Unlike the simpler autoregressive model used by Wyatt (2003), it is straightforward, through definition of the weight matrix, to allow the strength of correlation between neighboring units to be dependent on their relative distances, which is appealing when the nearest neighboring habitat units may be separated by stretches of uninhabitable stream. CAR models are not generally used for prediction, but if covariates are available for unsurveyed units, and these units are included when defining the correlation structure via the weight matrix, then we may predict density or abundance in these units. Further, prediction will be more precise than if the spatial dependence had been ignored, and this will in turn lead to greater precision in estimates of overall abundance in the surveyed stream.

While it is straightforward to construct complex hierarchical models for count and removal data, depending on the study design, there may be a practical limitation on model complexity. Surveys of riverine fish abundance typically use relatively few passes per unit, which makes it difficult to model random variation in the detection component of our models. As an alternative, we considered models for which heterogeneity in detection appeared in the model only through covariates. Our simulations show that applying this more restricted model can lead to significant underestimation of abundance when a large degree of heterogeneity remains unaccounted for. This demonstrates the importance of gathering as much data as possible on variables which may affect detection probability and including these variables as covariates in the detection component of the model. By doing this we will reduce the amount of residual variation and mitigate any impact this may have on abundance estimation. While increasing the number of passes for some units can also improve estimation of N, in designing a survey this would likely be at the expense of the number of units to be sampled, L.

Much of the capture-recapture literature deals with how to model heterogeneity in detection

probabilities between individual animals (e.g. Burnham and Overton, 1978; Norris and Pollock, 1996; Shriner, 2001; Dorazio and Royle, 2003). Although such heterogeneity is almost certainly present to some degree, we are unable to model it using data from a repeated count survey as individual animals are not identified. Even if we could attempt such modeling, Link (2003) demonstrates that the results are highly model dependent, even when alternative models fit equally well.

We have presented a framework for the modeling of repeated count data observed at the level of the habitat unit, incorporating covariates measured on individual units and allowing for small-scale spatial dependence. We can easily extend the models for density to two-stage sampling schemes designed for regional monitoring of fish density. In such monitoring, individual streams or segments of streams are first selected, and then habitat units within these streams are surveyed as for the single stream situation described above. Many covariates cannot be measured at the fine scale of an individual unit, yet variables such as landscape, soil and climate characteristics will be available on a coarser scale from existing data or from Geographic Information System databases. Further, we may expect average densities in nearby streams to be similar relative to densities in far apart streams, even allowing for measured covariates. This could be due to unknown or unmeasured covariates or, for migratory fish species, due to the patterns of fish migration across the stream network. We suggest adding an additional error term to the log-linear model to account for this large-scale spatial dependence. Suppose we sample S streams and in stream s we survey L_s habitat units. On unit i of stream s we measure the covariate vector $\mathbf{X}_{si}$, while for stream s itself we have the additional covariate vector $\mathbf{Z}_s$. Finally, denote the random error component associated with stream s by η_s. Thus we extend Equation (2.2) to $\log(d_{si}) = \mu_{si} + \phi_{si} + \eta_s$ for $i = 1, \ldots, L_s$, $s = 1, \ldots, S$. We assume η_s is Gaussian with mean zero, and that it is independent of ϕ_{si}. As before, the mean process μ_{si} will be a function of the covariate vectors and a parameter vector $\boldsymbol{\alpha}$ say. In the case that this function is linear, we can partition the parameters into two vectors, $\boldsymbol{\alpha}_x$ associated with $\mathbf{X}_{si}$ and $\boldsymbol{\alpha}_z$ associated with $\mathbf{Z}_s$ and write $\mu_{si} = \mathbf{X}'_{si}\boldsymbol{\alpha}_x + \mathbf{Z}'_s\boldsymbol{\alpha}_z$. As with the small-scale error term, we must construct a model for the correlation structure of the vector $\boldsymbol{\eta}$. Again, a CAR model is one possibility when the stream system is divided into discreet sampling units.

In some situations we may only have presence-absence data rather than counts, i.e., instead of a sequence of counts of animals seen in each unit, we have a sequence of zeros and ones, with ones denoting that at least one animal was seen. Presence-absence surveys may be prefered because such data may be inexpensive to collect compared to other methods, or where individuals cannot be distinguished reliably for the purpose of counting. MacKenzie et al. (2002) constructed

the likelihood for data from independent units, and focused on estimating the probability a unit is occupied (ψ_i for unit i) and the probability of detecting at least one animal on one survey pass of a unit (ρ_i, say). We note that by using a logit link function as we have done for individual detection probability, models incorporating covariates and, for ψ, spatial dependence, can be fitted using the Bayesian approach we have outlined above. However, because there is less information in presence/absence data, we expect that the number of units sampled and the number of repeated passes made on each unit will have to be large in order to obtain good estimates and distinguish different sources of variation. Royle and Nichols (2003) noted that ρ_i is a function of individual detection probability p_i and unit abundance N_i: $\rho_i = 1 - (1 - p_i)^{N_i}$. By exploiting this relationship, they use maximum likelihood estimation to estimate abundance from presence/absence data collected at independent sites. We could do likewise, once again applying the log-linear model for density and the logit model for detection to allow for covariates and spatial dependence. However, it is clear from the work of Royle and Nichols (2003) that very large sample sizes and many passes of each unit are required for estimation, and even then more complex models, akin to those we have examined, can be very unstable.

Bayesian hierarchical spatial modeling of count and removal surveys of stream dwelling fish can be extremely useful in helping understand the relationship between fish density and environmental variables while allowing for spatial dependence. Using such models can improve our ability to predict fish density at unsurveyed locations, and thereby aid fisheries scientists in understanding the distributional status of fish species by allowing us to produce spatial maps of predicted density across an individual stream or a regional network of streams. For removal data, Wyatt (2003) reviewed similar models for mapping abundance in streams divided into sections of fixed width, using GIS data as covariates. We have shown how modeling can proceed when the survey units are habitat units such as pools and runs by using CAR models for spatial dependence. Such models are flexible enough to account for varying distances between the habitat units, and the larger scale GIS covariates that were the focus of Wyatt's work can be easily incorporated when the streams or sections of streams which contain the habitat units are themselves units of sampling at a regional level.

Acknowledgements

We thank the California Department of Fish and Game, Salmon and Steelhead Trout Restoration Account, for funding this work. We also owe particular thanks to Dr Dave Hankin of Humboldt

State University for his invaluable input and to his field staff for collecting the Coho Salmon survey data used in our analyses.

Bibliography

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, *in* B. N. Petrov and F. Csaki (eds), *Proceedings of the Second International Symposium on Information Theory*, pp. 267–281.

Banerjee, S., Carlin, B. P. and Gelfand, A. E. (2004). *Hierarchical Modeling and Analysis for Spatial Data*, Boca Raton: Chapman and Hall.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems, *Journal of the Royal Statistical Society, Series B* **36**: 192–236.

Brakensiek, K. E. (2002). *Abundance and survival rates of juvenile coho salmon (oncorhynchus kisutch) in Prairie Creek, Redwood National Park*, Master's thesis, Humboldt State University.

Burnham, K. P. and Overton, W. S. (1978). Estimation of the size of a closed population when capture probabilities vary among animals, *Biometrika* **65**: 625–633.

Carroll, R. J. and Lombard, F. (1985). A note on n estimators for the binomial distribution, *Journal of the American Statistical Association* **80**: 423–426.

Casella, G. (1986). Stabilizing binomial n estimators, *Journal of the American Statistical Association* **81**: 172–176.

Cressie, N. (1993). *Statistics for Spatial Data*, 2nd edn, New York: Wiley.

Diggle, P. J., Tawn, J. A. and Moyeed, R. A. (1998). Model-based geostatistics (with discussion), *Applied Statistics* **47**: 299–350.

Dorazio, R. M., Jelks, H. L. and Jordan, F. (2005). Improving removal-based estimates of abundance by sampling a population of spatially distinct subpopulations, *Biometrics* **61**: 1093–1101.

Dorazio, R. M. and Royle, J. A. (2003). Mixture models for estimating the size of a closed population when capture rates vary among individuals, *Biometrics* **59**: 351–364.

Forsyth, D. M., Link, W. A., Webster, R., Nugent, G. and Warburton, B. (2006). Nonlinearity and seasonal bias in an index of brushtail possum abundance, *Journal of Wildlife Management* **(to appear)**.

Gelfand, A. E. and Ghosh, S. (1998). A minimum predictive posterior loss approach, *Biometrika* **85**: 1–11.

Gelfand, A. E. and Vounatsou, P. (2003). Proper multivariate conditional autoregressive models for spatial data analysis, *Biostatistics* **4**: 11–25.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models, *Bayesian Analysis* **(to appear)**.

Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (2004). *Bayesian data analysis*, 2nd edn, Boca Raton: Chapman and Hall.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences, *Statistical science* **7**: 457–472.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian resolution of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**: 721–741.

Ghosh, S. K. and Norris, J. L. (2005). Bayesian capture-recapture analysis and model selection allowing for heterogeneity and behavioral effects, *Journal of Agricultural, Biological and Environmental Statistics* **10**: 35–49.

Gosky, R. M. (2004). *Bayesian Analysis and Matching Errors in Closed Population Capture Recapture Models*, PhD thesis, North Carolina State University.

Hankin, D. G. (2004). Trial implementation of systematic sampling protocols for assessing presence/absence of juvenile coho salmon (oncorhynchus kisutch) in small streams in northern california. Unpublished report, Humboldt State University.

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

Johnson, N. L. and Kotz, S. (1969). *Discrete Distributions*, 1st edn, Boston: Haughton Mifflin Company.

Lichstein, J. W., Simons, T. R., Shriner, S. A. and Franzreb, K. E. (2002). Spatial autocorrelation and autoregressive models in ecology, *Ecological Monographs* **72**: 445–463.

Link, W. A. (2003). Nonidentifiability of population size from capture-recapture data with heterogeneous detection probabilities, *Biometrics* **59**: 1123–1130.

Link, W. A., Cam, E., Nichols, J. D. and Cooch, E. G. (2002). Of BUGS and birds: Markov Chain Monte Carlo for hierarchical Bayesian modeling in wildlife research, *Journal of Wildlife Management* **66**: 277–291.

MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Royle, J. A., and Langtimm, C. A. (2002). Estimating site occupancy when detection probabilities are less than one, *Ecology* **83**: 2248–2255.

Monahan, J. F. (2001). *Numerical Methods of Statistics*, Cambridge: Cambridge University Press.

Neal, R. M. (2003). Slice sampling, *Annals of Statistics* **31**: 705–767.

Norris, J. L. and Pollock, K. H. (1996). Nonparametric MLE under two closed capture-recapture models with heterogeneity, *Biometrics* **52**: 639–649.

Olkin, I., Petkau, A. J. and Zidek, J. V. (1981). A comparison of n estimators for the binomial distribution, *Journal of the American Statistical Association* **76**: 637–642.

Pascutto, C., Wakefield, J. C., Best, N. G., Richardson, S., Bernardinelli, A., Staines, A. and Elliot, P. (2000). Statistical issues in the analysis of disease mapping data, *Statistics in medicine* **19**: 2493–2519.

Plummer, M., Best, N., Cowles, K. and Vines, K. (2004). *The coda package, Version 0.9-1*. http://www-fis.iarc.fr/coda/.

Pollock, K. H. (2002). The use of auxiliary variables in capture-recapture modelling: an overview, *Journal of Applied Statistics* **29**: 85–102.

Pollock, K. H., Nichols, J. D., Brownie, C. and Hines, J. E. (1990). Statistical inference for capture-recapture experiments, *Wildlife Monographs* **107**: 1–97.

Routledge, R. D. (1981). The unreliability of population estimates from repeated incomplete aerial surveys, *Journal of Wildlife Management* **45**: 997–1000.

Royle, J. A. (2004). N-mixture models for estimating population size from spatially replicated counts, *Biometrics* **60**: 108–115.

Royle, J. A., Link, W. A. and Sauer, J. R. (2002). Statistical mapping of count survey data, *in* J. M. Scott, P. J. Heglund, M. L. Morrison, J. B. Haufler, M. G. Raphael, W. A. Wall and F. B. Samson (eds), *Predicting Species Occurences: Issues of Accuracy and Scale*, Washington: Island Press.

Royle, J. A. and Nichols, J. D. (2003). Estimating abundance from repeated presence-absence data or point counts, *Ecology* **84**: 777–790.

Seber, G. A. F. (1982). *The estimation of animal abundance and related parameters*, 2nd edn, New York: Macmillan.

Shriner, S. A. (2001). *Distribution of breeding birds in Great Smoky Mountains National Park*, PhD thesis, North Carolina State University.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002). Bayesian measures of model complexity and fit, *Journal of the Royal Statistical Society, B* **64**: 583–639.

Spiegelhalter, D., Thomas, A., Best, N. and Lunn, D. (2003). *WinBUGS User Manual, Version 1.4*.

Thomas, A., Best, N., Arnold, R. and Spiegelhalter, D. (2002). *GeoBUGS User Manual, Version 1.1 Beta June 2002*.

Thompson, S. K. (2002). *Sampling*, 2nd edn, New York: Wiley.

Wikle, C. (2002). Spatial modelling of count cata: a case study in modelling breeding bird survey data on large spatial domains, *in* A. B. Lawson and G. T. Denison (eds), *Spatial Cluster Modelling*, Boca Raton: Chapman and Hall/CRC.

Wyatt, R. J. (2002). Estimating riverine fish population size from single- and multiple-pass removal sampling using a hierarchical model, *Canadian Journal of Fisheries and Aquatic Science* **59**: 695–706.

Wyatt, R. J. (2003). Mapping the abundance of riverine fish populations: integrating hierarchical bayesian models with a geographic information system (GIS), *Canadian Journal of Fisheries and Aquatic Science* **60**: 997–1006.

Zippen, C. (1956). An evaluation of the removal method of estimating animal populations, *Biometrics* **12**: 163–189.

Chapter 3

Bayesian Spatial Modeling of Data from Bird Surveys

Raymond A. Webster, Kenneth H. Pollock, and Theodore R. Simons

ABSTRACT

When analyzing data from large bird surveys, it has been common in the past to ignore important factors such as variation in bird detection probabilities across space, and spatial dependence in bird density. We present a unified framework for modeling bird survey data collected at spatially replicated survey sites in the form of repeated counts, "removal" counts, or "capture" history counts, that simultaneously models spatial variation in bird density and variation in detection probabilities due to changes in covariates across the landscape. The models have a complex hierarchical structure that makes them suited to Bayesian analysis using Markov chain Monte Carlo (MCMC) algorithms. In order to ensure that these algorithms are computationally efficient, we use a form of conditional autogressive model proposed by Hrafnkelsson and Cressie (2003) for modeling spatial dependence. We apply our models to survey data for three bird species in the Great Smoky Mountains National Park. Our methods lead to maps of predicted relative density which are an improvement over those that would follow from ignoring spatial dependence. Modeling shows that variation in detection probability can also affect inference, particularly when a species is relatively difficult to detect. Our work also highlights the importance of good survey design for bird species modeling. We briefly discuss extensions of the models to include temporal effects or allow for multiple species. We point out that these types of bird survey data, particularly removal and capture-recapture counts (which require individual birds to be identified), are very prone to errors in bird identification. Although we obtain similar results for all three types of survey data, which provides indirect evidence that the effect of such errors may be small, the consequences of identification errors in the data requires further investigation.

3.1 Introduction

The analysis of data from large surveys of bird populations is often based on models which make simplifying assumptions and ignore important aspects of the survey design. The goals of such bird surveys include the estimation of density, mapping the distribution of animals across a region, and modeling the relationships between bird density and habitat variables such as elevation and vegetation. Although bird surveys typically involve making observations on many line or point transects ("survey sites"), data are often aggregated over all sites for analysis (Farnsworth et al., 2002; Alldredge, 2004) or the analysis ignores spatial structure in the data by assuming survey sites are spatially independent (Shriner, 2001). Ignoring spatial dependence will influence estimates of the effects of any covariates on bird density (Cressie, 1993), while accounting for such dependence can lead to improved prediction of density at unsurveyed locations. Further, uncertainty in the detection of animals has either been ignored or dealt with in an ad-hoc manner (Shriner, 2001; Lichstein et al., 2002). Detection probability will vary across the landscape, with the time of the year, according to the local weather conditions, and with individual observers and between different bird species. Not allowing for such variation may lead to over or underestimation of relative bird density at some locations, affecting inference from fitted models, or resulting in misleading comparisons of bird species populations or inaccurate assessment of population changes over time. We present a unified modeling framework for the analysis of repeated count, removal and capture-recapture data from bird surveys carried out at spatially replicated sites that allows for imperfect detection and spatial dependence in the data.

While some surveys involve a simple count of birds of each species along a line transect or at a point transect (Royle et al., 2002; Wikle, 2002), count data alone are not sufficient to distinguish variation in density from variation in detectability. In our work, we consider three other types of survey data gathered by making a sequence of independent observations during multiple survey periods at each site: repeated counts, in which all birds seen or heard are counted in each survey period; "removal" counts, which differ from repeated counts in that individual birds observed in previous periods are not counted in the current period; and "capture-recapture" data, in which observed birds are again identified individually and a record is kept of whether an individual was observed in each survey period or not. There is great potential for errors in such surveys, particularly when individuals are identified by sound, but although we discuss this throughout our work, the models we present assume that observations are made without error. We also assume that the population

is closed to births, deaths, immigration, and emigration during the course of the survey of a single survey site, which are reasonable assumptions as repeated observations are usually made within a very short time period (e.g., 10 minutes for Farnsworth et al., 2002; Alldredge, 2004). Given this, we can make use of binomial models for repeated count and removal data from multiple survey sites, and in a hierarchical framework, combine these with Poisson-lognormal mixture models for relating abundance to covariates. Royle (2004) and Webster et al. (2006) describe this hierarchical framework for repeated count data, while Wyatt (2002), Dorazio et al. (2005), and Forsyth et al. (2006) apply similar models to removal data.

Closed population capture-recapture models have been extensively studied (see, for example, Otis et al., 1978; Pollock et al., 1990), including those which attempt to model heterogeneity in detectability between individual animals (e.g., Chao, 1987; Norris and Pollock, 1996; Coull and Agresti, 1999; Pledger, 2000; Dorazio and Royle, 2003). In the surveys of multiple sites that form the focus of our work, we only model such heterogeneity by allowing for site differences due to covariates or for changes in detectability across the multiple survey periods for each site using a logistic-linear model. This approach is similar to that taken by Huggins (1989) and Alho (1990), who considered logistic models with covariates measured on individual animals, and in Section 3.2 we show that this allows us to fit capture-recapture models within the same framework as models for repeated count and removal data. Modeling heterogeneity between individuals within sites as a random effect is a more difficult problem, and as shown by Link (2003), is highly dependent on the choice of model. We do not consider this class of heterogeneity models in our work.

The complexity of these models makes them difficult to fit using maximum likelihood estimation, but their hierarchical structure means that they are relatively easily implemented using Bayesian modeling with Markov chain Monte Carlo (MCMC) algorithms, for example, by combining Gibbs Sampling (Geman and Geman, 1984) with Metropolis-Hastings steps (Hastings, 1970). Accounting for spatial structure with large data sets requires specifying models that lead to efficient MCMC algorithms. Algorithms based on conditional autoregressive (CAR) models (Besag, 1974) are much faster than those for geostatistical models described by Cressie (1993) and used by Royle et al. (2002) for modeling bird count data. Lichstein et al. (2002) used a CAR model for modeling bird survey data, but their approach required fitting preliminary geostatistical models to estimate the distance at which spatial dependence between two points becomes negligible. In Section 3.3 we propose the use of a variation of standard CAR models for spatial modeling in two-dimensional space (Hrafnkelsson and Cressie, 2003), in which this distance becomes a parameter

of the models.

In Section 3.4 we apply the models to part of a large bird survey data set from the Great Smoky Mountains National Park. The data are in the form of capture-histories at multiple sites, and from this we also extract the appropriate statistics for repeated count and removal analysis of the data. To illustrate the application to the different types of survey data, we fit the hierarchical spatial models to all three versions of the data. A comparison of the three methods is particularly interesting given that it is likely that identification errors exist, but will have a lesser effect when data are aggregated in the form of counts than when we require individuals to be distinguished for analysis using capture-recapture models. These data were analyzed previously by Shriner (2001), Farnsworth et al. (2002), and Alldredge (2004), while Lichstein et al. (2002) looked at a similar data set from the adjacent Pisgah National Forest, North Carolina. Previous analyses have focused on certain aspects of the modeling while neglecting others. Shriner (2001) modeled presence-absence data, but attempted to remove the problem of estimation of detection by assuming that birds within a certain radius of the observer had detection probabilities of 1, with birds detected beyond this radius excluded from the modeling. This author also ignored spatial dependence in her analysis and predicted presence at unobserved sites from covariate models alone. Farnsworth et al. (2002) did consider detection in removal versions of the data, but the data were aggregated over all sites and so the spatial structure was lost. A CAR model was used by Lichstein et al. (2002) to model spatial dependence in count data, but only total site counts of unique individuals were used rather than repeated counts, and as with Shriner (2001), a cut-off radius was used in an attempt to ensure that detection was perfect within this radius. Finally, Alldredge (2004) examined sources heterogeneity in detection probability for the capture-recapture data, but the spatial locations of the observations were ignored. Our modeling framework allows us to simultaneously model all components of the data structure, including covariates, detection probabilities, and spatial dependence.

In Section 3.2 we describe our hierarchical models for bird density and detection probability, while we discuss spatially modeling of bird survey data in Section 3.3. We present detailed analyses of data for selected species from the Great Smoky Mountains bird survey, including maps of the spatial distribution of relative bird density. Finally, the implications of our work and possible avenues of future research are discussed in Section 3.5.

3.2 Population and detection models

Bird survey data are often collected as a sequence of independent repeated counts, "removal" counts, or "capture-recapture" data at multiple survey locations throughout a large area of interest. Our use of quotation marks here is to indicate that the data conforms to this particular type, even though birds are generally neither physically removed nor captured. In each of the three cases, the data at each survey site can be summarized in the form of a vector of animal counts, the form and distribution of which depends on the survey design. Suppose we survey L sites, and make T independent observations at each site. Further, we assume that these T observations are made within a short enough space of time that the population is closed to migration, births, and deaths. Let N_i be the "abundance" of animals at survey site i, and p_{ij} the probability an animal is detected at site i on occasion j. In fact, in most situations N_i is a measure of relative abundance, as we discuss below.

In a repeated count survey, a record is kept of the number of animals observed at a site on each of the T survey periods and these counts are assumed to be made independently at each site. A removal survey differs only in that birds observed in one period are "removed" from the population to be counted on subsequent periods at each site. Such a survey rarely involves actual removal of birds, and instead we assume that it is possible to track individual birds over the T survey periods. Let y_{ij} be the observed bird count site i for period j, so that at each site we observe a data vector $\mathbf{Y}_i = [y_{i1}, \ldots, y_{iT}]'$. For both repeated count and removal data, we can apply a binomial model to the observed counts. For $i = 1, \ldots, L$ and $j = 1, \ldots, T$,

$$y_{ij}|N_i, p_{ij} \sim \text{Bin}(N_i - M_{ij}, p_{ij}), \tag{3.1}$$

where $M_{ij} = \sum_{j<j'} y_{ij'}$ is the number of animals removed prior to occasion j, with $M_{i1} = 0$. For count data, $M_{ij} = 0$ for $j = 1, \ldots, T$ as no animals are removed during the T survey occasions. This binomial model has been used by Royle (2004) for multiple-site count data, and by Wyatt (2002) and Forsyth et al. (2006) for removal data.

In surveys that yield the capture-recapture type of data, a record is kept of whether each observed bird was detected in a given survey period or not. Our data, therefore, consist of values x_{jk}, where $x_{jk} = 1$ if bird k is seen or heard ("captured") in period j, and 0 otherwise. For our models, which allow detection probability to vary between sites and between periods within a site, we can summarize the capture histories of animals as a vector of counts for each site, with each count being the number of animals with a particular unique capture history (Otis et al., 1978).

For example, for $T = 3$ periods, there are 8 possible capture histories, $\omega = 111$, 110, 101, 100, 011, 010, 001 and 000. Let the number of animals with history ω at site i be denoted by $X_{i\omega}$, and let $\mathbf{X}_{i\omega} = [X_{i,111}, X_{i,110}, \ldots, X_{i,000}]$ for the three period example. If $\boldsymbol{\pi}_{i\omega} = [\pi_{i,111}, \pi_{i,110}, \ldots, \pi_{i,000}]$, where

$$\begin{aligned}
\pi_{i,111} &= P(\omega = 111) = p_{i1}p_{i2}p_{i3}, \\
\pi_{i,110} &= P(\omega = 110) = p_{i1}p_{i2}(1 - p_{i3}), \\
&\vdots \\
\pi_{i,000} &= P(\omega = 000) = (1 - p_{i1})(1 - p_{i2})(1 - p_{i3}),
\end{aligned}$$

then $\mathbf{X}_{i\omega}$ has a multinomial distribution given by

$$\mathbf{X}_{i\omega}|N_i, \boldsymbol{p}_{i\omega} = \text{Multinomial}(N_i, \boldsymbol{\pi}_{i\omega}). \tag{3.2}$$

Note that $X_{i,000}$ is unknown since animals with capture history 000 are unobserved, and in constructing a likelihood function, we replace this with $X_{i,000} = N_i - \sum_{\omega \neq 000} X_{i\omega}$.

Although the data models differ for count, removal and capture-recapture data, in all three cases the distribution of the data at a site depends on the same set of parameters, N_i and $p_{i1}, \ldots, p_{iT}$ for $i = 1, \ldots, L$. The next step in our hierarchical modeling is to specify models for these parameters. For N_i we use a Poisson-lognormal mixture to model variation in abundance across the sites as a function of covariates and spatially dependent Gaussian errors (see Diggle et al., 1998; Banerjee et al., 2004) As well as being very flexible, this mixture model allows for over-dispersion in the site abundances, which is generally present in animal data due to the clustering of animal populations. The model is as follows:

$$\begin{aligned}
N_i|\lambda_i &\sim \text{Poisson}(\lambda_i) \\
\log(\lambda_i) &= \mu_i + \phi_i
\end{aligned} \tag{3.3}$$

Here ϕ_i is a zero-mean Gaussian error term through which we model spatial dependence (Section 3.3) and μ_i represents the covariate model, most commonly of linear form, $\mu_i = \mathbf{X}_i'\boldsymbol{\alpha}$, where $\mathbf{X}_i$ here represents the vector of covariates measured on unit i, and $\boldsymbol{\alpha}$ is the corresponding parameter vector. In the case where the animals are observed within a well-defined area around the survey transect,

we may replace λ_i in Equation (3.3) with $A_i\lambda_i$, where A_i is the surveyed area around transect i, and λ_i becomes the density of birds per unit area. In most line and point transect surveys, however, there is no clear boundary beyond which birds are unobservable. The N_i and λ_i in this case correspond to measures of relative abundance and density, and by allowing detection probability to vary between sites, we can still make appropriate inference on the parameters of the covariate model and of variation in density across space, between species, or over time.

In the detection model we use a linear logistic model to allow for variation in detection due to covariates,

$$\text{logit}(p_{ij}) = \nu_{ij},$$

where ν_{ij} is a function of covariates that vary between sites and across the T periods within the sites. For example, if we have values of the covariates at each of the L sites and we fit a linear model, then $\nu_{ij} = \mathbf{Z}_i'\boldsymbol{\alpha}$, where $\mathbf{Z}_i$ is the covariate vector for site i, and $\boldsymbol{\beta}$ is a vector of parameters. In this work we do not consider models which allow for a random effect in the detection model as per Forsyth et al. (2006). Our experience is that it can be difficult to distinguish variation in animal density from that due to detectability, particularly when there are few survey periods (i.e., T is small, common for bird surveys), and attempting to do so does not improve the fit of the models.

Because of their complexity and hierarchical structure, we use Bayesian modeling to make inference on the model parameters. In Bayesian modeling, the goal is to compute the form of the distribution of the parameters given the data and some prior information. The distribution thus determined is known as the posterior distribution and for a parameter vector $\boldsymbol{\theta}$, this is found by solving the equation

$$\wp(\boldsymbol{\theta}|\mathbf{Y}) = \frac{\wp(\mathbf{Y}|\boldsymbol{\theta})\wp(\boldsymbol{\theta})}{\int \wp(\mathbf{Y}|\boldsymbol{\theta})\wp(\boldsymbol{\theta})d\boldsymbol{\theta}},$$

where $\wp(\theta|\mathbf{Y})$ is the joint posterior density, $\wp(\mathbf{Y}|\boldsymbol{\theta})$ is the likelihood of the data and $\wp(\boldsymbol{\theta})$ is a prior density through which prior knowledge about $\boldsymbol{\theta}$ can be incorporated. This requires specification of prior distributions for the parameters, which may be informative, allowing for past knowledge, or non-informative, in which case we wish them to have little effect on the inference. The idea of allowing previous research and expert opinion to inform our understanding of the present data seems very appealing, particularly when the present information is poor as is often the case in wildlife population research. However, deciding how much weight to give to these two sources of information is an inherently subjective process, and clear justification must be given for any informative priors. In cases where there is no useful prior information about a parameter, or if

we wish the focus of our analysis to be on the information contained in the present data alone, we can select non-informative or vague priors so that the form of the posterior is approximately determined by the likelihood alone. This is the approach we take in the present work. We discuss our particular choice of prior distributions in more detail in our example in Section 3.4.

For the models presented in our work, we use MCMC algorithms based on Gibbs sampling (Geman and Geman, 1984) with Metropolis or Metropolis-Hastings steps (Hastings, 1970) to generate samples from the posterior distributions of the parameters. After discarding data generated prior to convergence of the algorithms (the 'burn-in'), the MCMC samples can be used to estimate posterior summary statistics such as means, standard deviations, quantiles or the form of densities themselves (for example, using kernel density estimation). For a discussion of convergence assessment and other diagnostics, we refer the reader to Gelman et al. (2004). As a simple model diagnostic, we perform a cross-validation by re-fitting models with 10% of sites omitted, and creating plots comparing the observed values of a summary statistic at the omitted sites with the corresponding predictions and their 95% prediction intervals. We would expect 95% of the observed values to fall inside the credible intervals, and failure of this would be evidence that the model is inconsistent with the data. As we have multiple observations at a site, we chose the mean count at a site for repeated count data, and the total number of unique observed birds for removal data and capture-recapture data, as the site summary statistics for prediction.

3.3 Spatial Models

The hierarchical structure of the models described in the previous section makes them particularly suitable for Bayesian modeling using MCMC algorithms. However, for these algorithms to be computationally efficient for large data sets, we must take care in our choice of model for the spatial dependent random effect, ϕ_i. The most intuitive approach would be to use a standard geostatistical model to describe the spatial structure, such as an exponential or the more general Matérn model. Within a Gibbs sampling algorithm, however, updates require the inversion of $L \times L$ matrices, which creates an impractical computational burden when L is large. One common covariance model, the spherical (Cressie, 1993), leads to a covariance matrix with sparse structure, and we considered this for modeling ϕ_i. However, we found that even this model could be prohibitively slow and obtaining convergence of the MCMC algorithms was also difficult.

A more practical alternative is to use conditional autoregressive (CAR) models (Besag, 1974),

which have become popular in recent years because MCMC iterations are very efficient to perform due to the conditional nature of the model specification. In a CAR model, ϕ_i is assumed to be normally distributed with the mean defined to be the weighted average of the elements of the vector of errors with the ith omitted, $\boldsymbol{\phi}_{-i} = [\phi_1, \ldots, \phi_{i-1}, \phi_{i+1}, \ldots, \phi_L]'$. Following Banerjee et al. (2004) we begin by defining a symmetric $L \times L$ matrix of weights, $\mathbf{W}$, with elements $w_{ii'}$. Now define $w_{i.} = \sum_{i'=1}^{L} w_{ii'}$, the row totals of $\mathbf{W}$. A 'proper' Gaussian CAR model then follows by specifying the distribution of ϕ_i conditionally as

$$\phi_i | \boldsymbol{\phi}_{-i}, \gamma, \sigma_\phi^2 \sim \mathrm{N}\left(\gamma \sum_{i' \neq i} \frac{w_{ii'}}{w_{i.}} \phi_{i'}, \frac{\sigma_\phi^2}{w_{i.}} \right). \tag{3.4}$$

Unconditionally, $\wp(\boldsymbol{\phi})$, the joint probability density of the vector $\boldsymbol{\phi} = [\phi_1, \ldots, \phi_L]'$, is given by

$$\wp(\boldsymbol{\phi} | \gamma, \sigma_\phi^2) \propto \exp\left\{ -\frac{1}{2\sigma_\phi^2} \boldsymbol{\phi}' \mathbf{M}^{-1} (\mathbf{I} - \gamma \mathbf{C}) \boldsymbol{\phi} \right\}, \tag{3.5}$$

where $\mathbf{M}$ is a diagonal matrix with ith diagonal element $1/w_{i.}$, and $\mathbf{C} = \mathbf{MW}$, the scaled weight matrix (Banerjee et al., 2004). Equation (3.5) specifies the kernel of a multivariate normal distribution with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}_\phi = \sigma_\phi^2 (\mathbf{I} - \gamma \mathbf{C})^{-1} \mathbf{M}$. Let $\tau_\phi = 1/\sigma_\phi^2$ be the precision parameter of the model. With weights defined above, $\boldsymbol{\Sigma}_\phi^{-1} = \tau_\phi \mathbf{M}^{-1} (\mathbf{I} - \gamma \mathbf{C})$ will be symmetric, and provided that $|\gamma| < 1$, this matrix will also be non-singular and $\boldsymbol{\Sigma}_\phi$ will be a valid covariance matrix. In this case, (3.4) defines a proper prior distribution in Bayesian hierarchical modeling (Banerjee et al., 2004), and hence this model specification is known as a proper Gaussian CAR model.

The parameter γ governs the strength of spatial dependence, with a value of zero leading to independence of the ϕ_i. However, it is difficult to interpret the magnitude of a non-zero γ in any meaningful way as we can for a correlation coefficient (e.g., Banerjee et al., 2004), and in general any degree of spatial dependence will lead to a large value for γ. For this reason, some authors prefer to fix $\gamma = 1$, leading to what is known as an intrinsic CAR model (Banerjee et al., 2004).

As each value of ϕ_i is a function of its neighbors, the CAR model has been most widely used when the survey region is divided into contiguous subregions because in such cases the neigborhood structure is relatively straightforward to define. For example, the simplest case would be to let $w_{ii'} = 1$ if sites i and i' share a common boundary, and zero otherwise. When the data are observed at points in continuous space, more care must be taken in specifying the neighborhood

structure. A convenient approach, and the one we take here, was proposed by Hrafnkelsson and Cressie (2003), who introduced a neighborhood radius (which we call $d_{\max}$) as an additional model parameter: points within $d_{\max}$ of a survey point are considered neighbors of that point and are assigned positive weights, with the magnitude of the weight being a decreasing function of distance from the survey point.

Let $r_{ii'}$ be the Euclidean distance of site i from site i', and let δ be the magnitude of the spatial dependence at $r_{ii'} = d_{\max}$. Hrafnkelsson and Cressie (2003) defined the weights, $w_{ii'}$, assuming a minimum distance between two points of 1 unit. For an arbitrary minimum distance, $r_{\min}$, let

$$\psi = -\frac{\log(\delta)}{\log(d_{\max}/r_{\min})}.$$

Then the $w_{ii'}$ for site i are given by

$w_{ii} = 0, i = 1, \ldots, L$
$w_{ii'} = r_{ii'}^{-\psi}, r_{ii'} \leq d_{\max}$
$w_{ii'} = 0, r_{ii'} > d_{\max}.$

The parameter δ is fixed at a low value so that the drop in the function relating weight to distance between points is not steep. We follow Hrafnkelsson and Cressie (2003) by setting $\delta = 0.05$. Depending on the value of $d_{\max}$, some points may be without neighbors and would need to be excluded from the analysis unless additional unsurveyed points were added to the observed data and prediction made at these points. When fitting a model with covariates, that approach would require covariate measurements at each of the unsurveyed points, which may not always be available. Alternatively, we could set a lower bound on $d_{\max}$ (d_L, say) such that at most only a small number of points have no neighbors within this distance. We do this in our example in Section 3.4.

Hrafnkelsson and Cressie (2003) also demonstrate that if the underlying spatial model is in fact of geostatistical form, such as a Matérn model (Cressie, 1993), then the above CAR specification is likely to be a good approximation. This is an attractive feature if one feels more comfortable with using geostatistical models for data in continuous space. The Matérn model in particular includes a smoothness parameter governing the continuity of spatial process, and the work of Hrafnkelsson and Cressie (2003) implies that using their modified CAR model will also allow for the degree of smoothness of a given underlying spatial process.

3.4 Example: Great Smoky Mountains Bird Survey

An extensive multi-species survey of the bird population of the Great Smoky Mountains National Park was carried out from 1996 to 1999. Over 3000 point survey sites were positioned around low-use hiking trails and a survey of the bird population was conducted at each of these locations. The survey at each point lasted 10 minutes during which time an observer would attempt to record all the birds that were seen or heard. A three-period capture history for each individual bird was created by dividing the observation period into three smaller intervals, the first of 3 minutes, the second of 2 minutes and the third of 5 minutes duration, and a record was kept of whether an individual bird was observed in each of these intervals or not. This survey design also allows us to compute the count statistics required for analysis by repeated count and "removal" survey models.

We restrict ourselves to analyzing data from 1750 survey points monitored in 1997, the first year for which a full survey was conducted (the 1996 survey covered a smaller area of the park than subsequent surveys), and focus on three species as examples: the Black-throated Green Warbler (BT), Ovenbird (OB) and Black and White Warbler (BL). The first two are common and easy to detect - the data we analyzed had 1445 distinct BT and 1615 OB recorded. BL is less abundant and more difficult to detect, with only 469 individuals observed in 1997. Our main interests in analysing this data set include exploring the relationships between density and habitat characteristics and determining if failure to allow for factors affecting detection probability affects model inference, and in developing maps predicting density of a species across the park. Our analyses will also improve upon the methods of that Shriner (2001), Lichstein et al. (2002) and others have applied to these and similar survey data. The most important factors affecting density can be summarized with two variables, elevation and topographic relative moisture index (TRMI). The latter is an index based on forest moisture gradients and has been shown to be related to forest species composition (Pinder et al., 1997), and we use it here instead of more complex categorical variables for forest composition. Using a simple linear model, we model the relative bird density at the ith point transect using

$$\log(\lambda_i) = \alpha_0 + \alpha_1 E_i + \alpha_2 TRMI_i + \phi_i,$$

where E_i and $TRMI_i$ are respectively, the elevation and TRMI values for the ith site. The survey sites are not surrounded by a clearly-defined area beyond which we cannot detect birds. This means that 'density' at a transect, λ_i, should be interpreted as an index of relative bird density as noted in Section 3.2. The term ϕ_i represents the spatially dependent error which we model using

the proper CAR model of Section 3.3.

Given this model, we fit a sequence of three increasingly complex logistic models for detection probability. We begin with a base model with no covariates, then allow for seasonal effects, and finally for local conditions at a point at the time the survey was taken.

Model 1 $\text{logit}(p_{ij}) = \log(\mathcal{T}_j) + \beta_0$

Model 2 $\text{logit}(p_{ij}) = \log(\mathcal{T}_j) + \beta_0 + \beta_1 \mathcal{S}_i$

Model 3 $\text{logit}(p_{ij}) = \log(\mathcal{T}_j) + \beta_0 + \beta_1 \mathcal{S}_i + \beta_2 \mathcal{A}_i + \beta_3 \mathcal{Q}_i + \beta_4 \mathcal{C}_{1i} + \beta_5 \mathcal{C}_{2i}$

The term $\log(\mathcal{T}_j)$ is an offset, where $\mathcal{T}_j = 3$, 2 or 5 minutes for period $j = 1$, 2 or 3, respectively. This allows the probability of detection to vary with the duration of survey period in such a way that the odds ratios are proportional to the ratios of the period lengths, $\mathcal{T}_j$. $\mathcal{S}_i$ takes the value 0 if a point was surveyed prior to June 21, and 1 otherwise, and thus represents a contrast between detection probabilities in spring and summer. We expect this to be important as the activity of birds, and, therefore, their chance of being detected, changes between seasons, and the sites were surveyed at different times from spring to summer. The remaining variables in Model 3 allow for effects of local conditions on detection probability. $\mathcal{A}_i$ is the air temperature at transect i, while $\mathcal{Q}_i$ takes the value 0 for no measured background noise (e.g., due to wind, water) and 1 otherwise. $\mathcal{C}_{1i}$ and $\mathcal{C}_{2i}$ take the values 1 for moderate and high cloud cover respectively, and 0 for low cloud cover. All continuous variables (elevation, TRMI and air temperature) were standardized prior to analysis.

Our MCMC algorithm is detailed in Appendix B. We used non-informative uniform prior distributions for the covariate coefficients in both the density and detection models, and a flat prior for the spatial neighborhood radius $d_{\max}$. This parameter was constrained to be greater than $d_L = 0.5$km, which led us to exclude only 24 survey sites for having no neighbors within distance d_L. We sampled values from the posterior of $\tau_\phi = 1/\sigma_\phi^2$, for which we used a prior density of $\wp(\tau_\phi) \propto \tau_\phi^{-3/2}$, which corresponds to a uniform prior on σ_ϕ as recommended by Gelfand et al. (2006), although with our large sample sizes the exact choice of prior was not important. We generally ran between 10000 and 20000 iterations following a burn-in period for BT and OB models, and at least twice as many when modeling the more sparse BL data.

Figure 3.1 shows the posterior means and 95% credible intervals of the coefficients of elevation and TRMI in the density model (α_1, α_2), and as a relative measure of detectability, the mean detection probability for the first 3-minute survey period, averaged over all L survey sites ($\bar{p}_3$). Full tables of posterior means and credible intervals are presented in Appendix A. Considering first the more abundant species, BT and OB, for both there is a clear negative relationship between density and

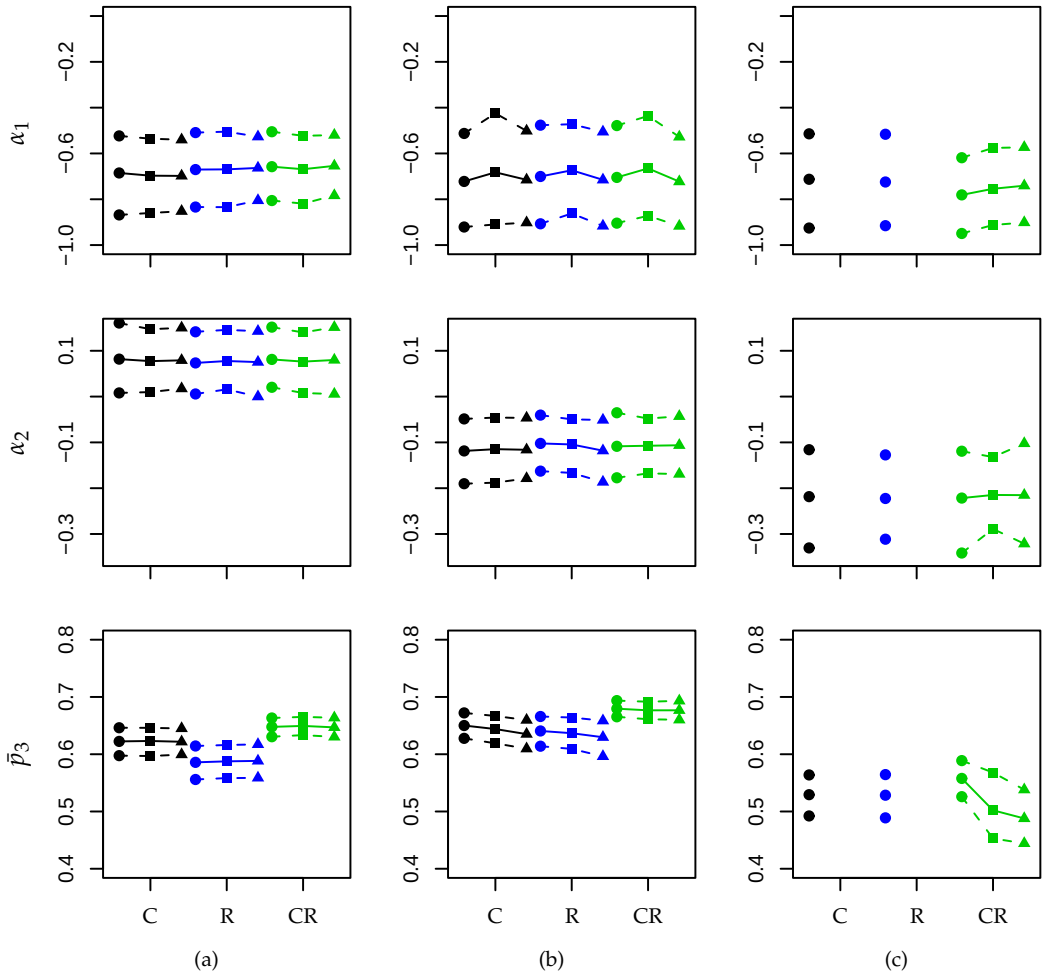

Figure 3.1: Posterior means (solid lines) and 95% credible intervals for selected parameters for (a) Black-throated Green Warbler (b) Ovenbird and (c) Black and White Warbler, for count(C), removal (R) and capture-recapture (CR) data, and for detection Models 1 (●), 2 (■) and 3 (▲).

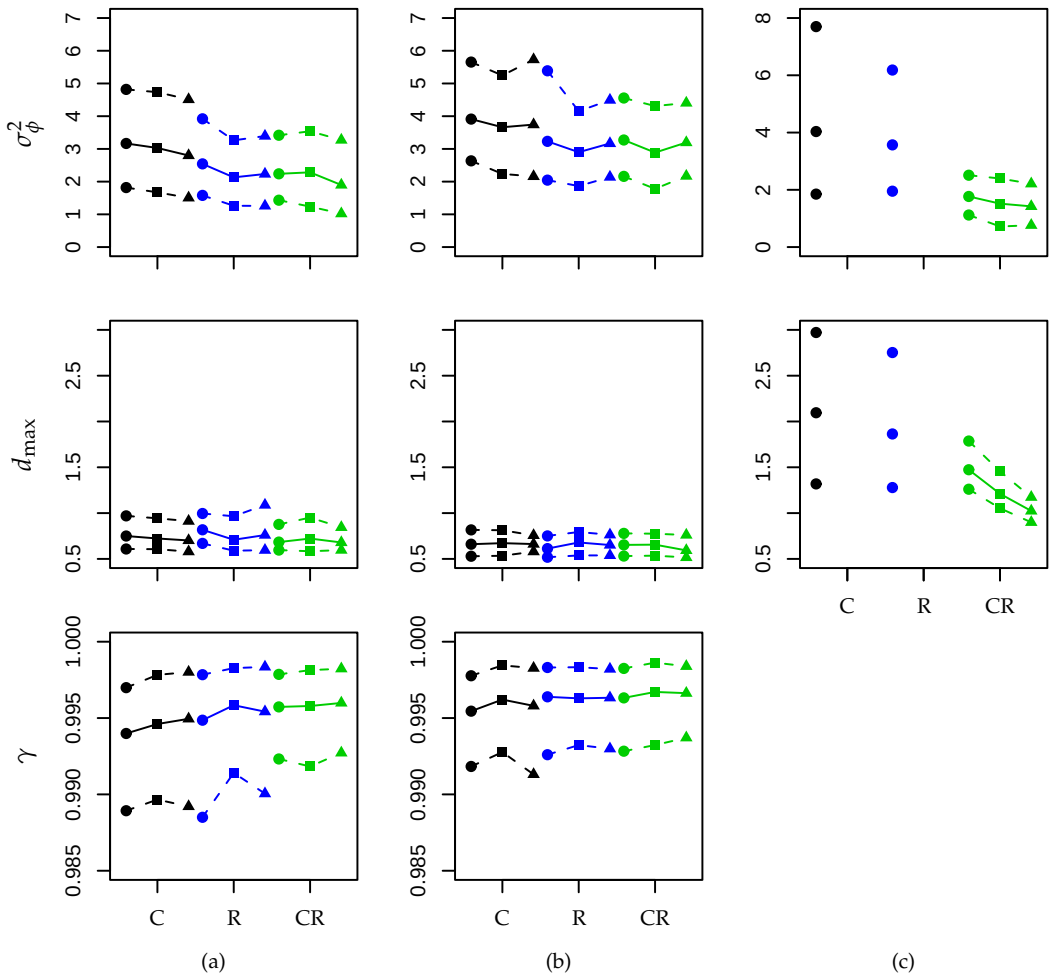

Figure 3.2: Posterior means (solid lines) and 95% credible intervals for selected parameters for (a) Black-throated Green Warbler (b) Ovenbird and (c) Black and White Warbler, for count(C), removal (R) and capture-recapture (CR) data, and for detection Models 1 (●), 2 (■) and 3 (▲).

elevation, and thus these species are more abundant at lower elevations. BT density is positively related to TRMI, while the relationship is negative for OB, although in both cases the magnitude of the coefficients and the wide credible intervals (they almost include 0 for BT) show that TRMI is a less important predictor than elevation. The coefficients of elevation and TRMI are consistent across the three types of data, repeated count, removal, and capture-recapture, with only small differences between the three detection models for each data type. Detection probabilities were higher in summer than in spring for BT (β_1 significantly greater than 0, Appendix A), while the importance of season is less clear for OB. There is evidence, however, from all three versions of the data that local noise at the time of the survey decreases the probability of detection of OB ($\beta_2 < 0$, Appendix A). Accounting for these detection effects does not appear to affect the estimation of the density coefficients for these species.

It is also interesting that, although the values of the density coefficients are similar between models for the three types of data, there are differences in the detection probabilities for repeated count, removal and capture-recapture models for BT and OB. In particular, detection probabilities (as illustrated by $\bar{p}_3$ in Figure 3.1) are higher on average for the capture-recapture model. Given the difficulty in distinguishing individual birds in this survey, we think it is likely that errors in the data would be most serious for capture-recapture data. If this is the case, the higher probabilities imply that observers are mistaking multiple birds for the same bird across the three survey periods, and, therefore, they record too few distinct individuals.

In Figure 3.2, we examine the posterior means and credible intervals of the three parameters of the spatial model for the ϕ_i, σ^2_ϕ, $d_{\max}$ and γ. For BT and OB it is apparent that the conditional variance parameter, σ^2_ϕ, is highest for count data and lowest for capture-recapture data, which we might expect, given that count data contains the least information about detection probabilites and capture-recapture is the most informative in this respect. But we also see that the although the parameter γ moves within a narrow range, the posterior means tend to be higher when those of σ^2_ϕ are lower. The correlation between these parameters became apparent when we examined their MCMC values. Perhaps less clear is that $d_{\max}$, particularly for BT, appears to be higher when σ^2_ϕ is higher, and, therefore, is negatively correlated with γ. It is possible that with these types of data, we should reduce the number of parameters in the spatial model.

In fact, we were forced to simpifly the spatial model when modeling the sparse BL data. Convergence of the BL models was much more difficult. Fewer birds were observed and, from our previous experience with fitting models to similar datasets, convergence of MCMC alogrithms can

be much slower when detection probabilities are relatively low and the number of observation periods at each sampling site is small (recall $T = 3$ here). Given the apparent convergence difficulties, and that even in case of BT and OB the spatial model may be over-parameterized, we chose to simplify the model by fixing $\gamma = 0.995$ in the proper CAR model, similar to the posterior mean values for BT and OB. This is close to an intrinsic CAR model, but setting up an algorithm directly for such a model is more difficult when $d_{\max}$ is a parameter, as its full conditional density would now depend on the improper joint prior density for the residual vector $\boldsymbol{\phi}$ (Appendix B). This simplification improved matters, but even then, for some of the more complex detection models with repeated count and removal data, the algorithms apparently failed to converge. No results are presented in the figures and tables for the cases for which convergence was uncertain.

From Figure 3.1c we can see that again BL density decreases with increasing elevation, and the coefficient of elevation, α_1, has similar posterior means to BT and OB. TRMI seems to be a more important variable for this species, however, with a much larger, negative value for the posterior mean of α_2. As expected, the estimated detection probabilities were lower for BL than for the other bird species. When the MCMC algorithms converged, Figure 3.2c shows that although the parameter σ^2_ϕ has a fairly similar posterior distribution for all three species (although somewhat lower values are seen for BL with capture-recapture data), the values of $d_{\max}$ are much higher. This would imply that even quite distant survey points (up to 2 km away) directly influence density at a given location. However, we have no plausible biological explanation as to why this may be the case.

For the capture-recapture models applied to BL data, the only ones for which the algorithms converged for all three detection models, there is some evidence that allowing for variation in detection probability due to covariates affects the results for the detection component of the model (Figure 3.1c). In particular, the posterior mean and credible intervals for α_1, the elevation coefficient, are slightly higher under detection Models 2 and 3 than with Model 1. The effect is not large but no effect at all was apparent for the easier-to-detect species, BT and OB. A clearer corresponding decrease in average detection probabilities are also apparent when detection probabilities are not constant across the survey sites. There is some evidence of a relationship between detection probability and season (Table 3.3, Appendix B), and accounting for this relationship has some affect on our inference on the detection model.

Figures 3.3-3.5 give the results of a cross-validation exercise to assure ourselves that the models are consistent with the observed data. For the Black-throated Green Warbler (BT) data, detection

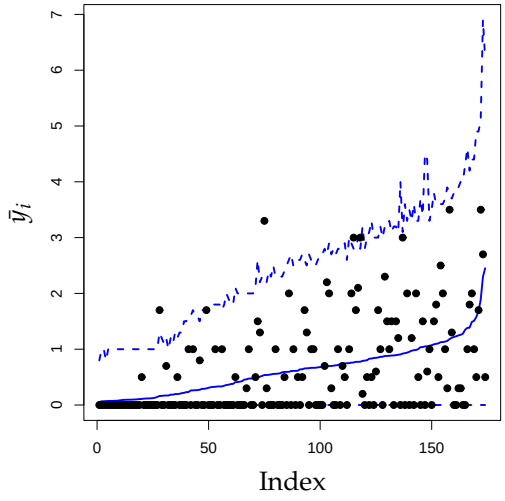

Figure 3.3: Comparison of posterior means (solid line) and 95% prediction intervals (dashed lines) with observed values (circles) of $\bar{y}_i$, the weighted mean count at site i, for 10% of the survey sites when the models are fitted to repeated count data with these sites omitted for the Black-throated Green Warbler (BT).

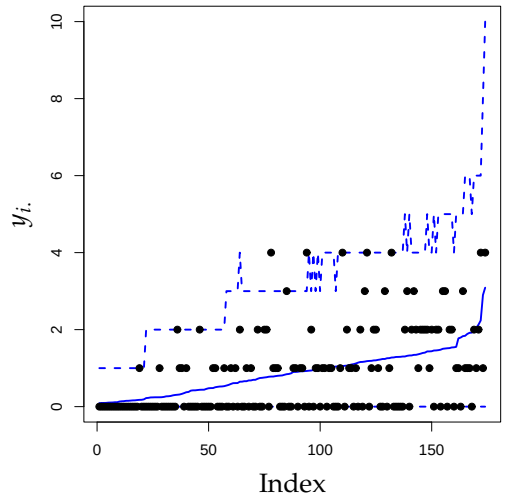

Figure 3.4: Comparison of posterior means (solid line) and 95% prediction intervals (dashed lines) with observed values (circles) of $y_{i.}$, the number of unique birds detected at site i, for 10% of the survey sites when the models are fitted to removal data with these sites omitted for the Black-throated Green Warbler (BT).

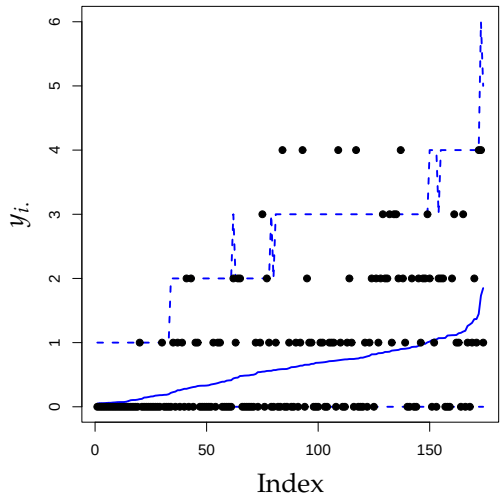

Figure 3.5: Comparison of posterior means (solid line) and 95% prediction intervals (dashed lines) with observed values (circles) of $y_{i.}$, the number of unique birds detected at site i, for 10% of the survey sites when the models are fitted to capture-recapture data with these sites omitted for the Black-throated Green Warbler (BT).

Model 1 was re-fitted with a random sample of 10% of the observations removed, that is, 175 of the original 1750 survey sites were treated as if no count data were available. The 'missing' counts were predicted using the re-fitted model, and for each of these 175 survey sites we computed posterior means and 95% prediction intervals of the weighted average of the three counts for repeated count data, and the total of the three counts for removal data and capture-recapture data. That is, for $i = 1, \ldots, 175$, for the repeated count model we computed

$$\bar{y}_i = \frac{\sum_{j=1}^{3} \mathcal{T}_j y_{ij}}{\sum_{j=1}^{3} \mathcal{T}_j},$$

while for removal and capture-recapture models we calculated

$$y_{i.} = \sum_{j=1}^{3} y_{ij}.$$

The figures show comparisons of these predicted statistics with the corresponding observed values for the 175 prediction sites. For clarity of presentation, we plot the values sorted by posterior mean, and the x-axis just represents the site index (1 to 175) of the sorted means. In each figure,

the solid blue line represents the posterior mean of the predicted values, and the dashed lines the corresponding lower and upper 95% prediction intervals, although the former was 0 for all models and all prediction locations. The solid circles are the observed values, and our interest lies in whether these fall within the credible intervals. In all three cases, few prediction intervals fail to include the corresponding observation, and thus there is no evidence that our models are inconsistent with the observed data. This was also true for the other species and detection models that we examined.

One of the most important goals of fitting spatial models to animal survey data is to produce maps of the population distribution in some region of interest. In the case of the bird survey, we are interested in mapping bird density across the Great Smoky Mountains National Park. Although our models do not allow us to directly estimate density, the Poisson intensity λ from Equation (3.3) provides an index of relative density, and maps of predicted values of λ will show how animal density varies across the region. Also of interest are maps of the spatially dependent error process, ϕ, which show patterns of spatial variation once we have allowed for covariates in the density model. This can give us insight into other factors that may affect animal distribution, such as covariates that have not already been included in the model.

As examples of the type of maps that can be produced by fitting the models we have presented in our work, we generated prediction maps of bird distribution by fitting detection Model 1 to the repeated count data for each of the three species. The original 1997 survey locations were augmented with a regular grid of 7683 sites spaced 500 m apart within the boundaries of the Great Smoky Mountains National Park. Values of elevation and TRMI at these prediction locations were extracted from a Geographic Information System (GIS) database, and are shown in Figure 3.6. Although it was technically straightforward to modify our MCMC algorithm (Appendix B) to include prediction on the grid, our algorithm runs too slowly for this to be practical. Instead, we used the WinBUGS package (Spiegelhalter et al., 2003) to fit a simplified version of our Bayesian model. The neighborhood radius parameter, $d_{\max}$, was fixed to be approximately the value of the posterior mean from our original modeling, i.e., 750 m for BT and OB, and 2000 m for BL, while we used an intrinsic CAR prior (i.e., $\gamma = 1$) since the posterior means of γ were always close to 1 for BT and OB. (We are concerned that the intrinsic CAR prior leads to improper posterior distributions (see Sun et al., 1999), and it is our intention to verify the results of the mapping by re-fitting the models with a proper CAR prior in future work.)

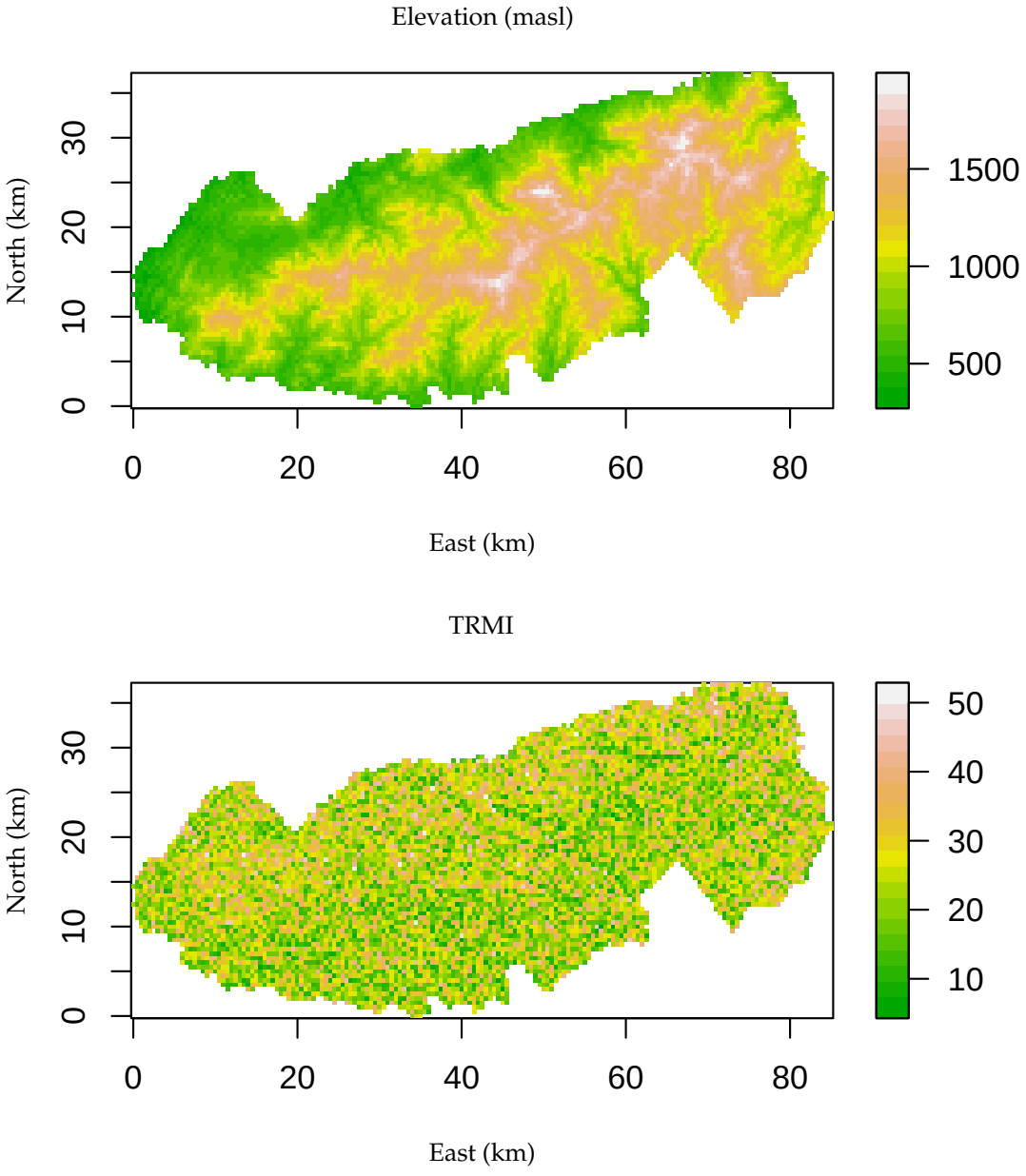

Figure 3.6: Elevation (meters above sea level) and TRMI at the 7683 grid locations used for prediction maps of bird distribution.

Figure 3.7 shows prediction maps of the distribution of BT. From the results in Table 3.1, Appendix B, elevation is the more important of the two covariates (recall that the covariates were standardized prior to analysis) and without allowing for spatial dependence, we would expect a map of predicted intensity, λ, to largely mimic the map of elevation in Figure 3.6. Although this species is not predicted to be abundant at higher altitudes, at lower altitudes we predict only three patches of high density (pink), in the west, south-east, near the eastern part of the northern boundary. There are also some areas of moderate density (yellow), notably in the central north and just east of the south-eastern high density patch. The map of predicted ϕ shows that, allowing for elevation and TRMI, we predict a patch of much higher than average density in the south-eastern section of the park, along with a few other high density regions, including a large area near the park's center. Identification of these areas of relatively high density may help the biologist to better understand the factors affecting the distribution of this species. It is important when attempting to interpret the prediction maps to consider the adjacent maps of standard deviation (sd) of prediction. The map of $sd(\widehat{\lambda})$ shows that the regions of highest intensity are also those of greatest uncertainty in prediction, something we would expect from the lognormal model for intensity, λ. Of even more interest in this case is the map of standard deviation of ϕ. The areas where prediction is precise (green) show up as a network of lines, which in fact correspond to the trails upon which the survey points were located. That is, we only have precise prediction in the neighborhood of the original survey sites. This has very important implications for survey design. If the goal is to create precise maps of a large region such as the Smoky Mountains National Park, then we should strive to locate the survey points throughout the region, rather than clustering them within narrow corridors such as trails, as was the case for the current survey.

The predicted distribution of OB is shown in Figure 3.8. This species appears to be concentrated in the south-western region of the park, according to the prediction map of λ, but this region contains few survey points, and again with the exception of areas close to survey trails, prediction is very imprecise (maps of $sd(\widehat{\lambda})$ and $sd(\widehat{\phi})$).

Finally, Figure 3.9 gives maps of the predicted distribution of the less abundant BL species. From the map of $\widehat{\lambda}$, this species appears to be found mainly in the park's southern areas of low elevation. The exception to this is the south-east section of the park, in which we predict BL to be largely absent, in contrast to the high density predicted for BL in the same area. Again, the lack of precision in prediction away from the survey sites implies that we must interpret the prediction maps with great caution.

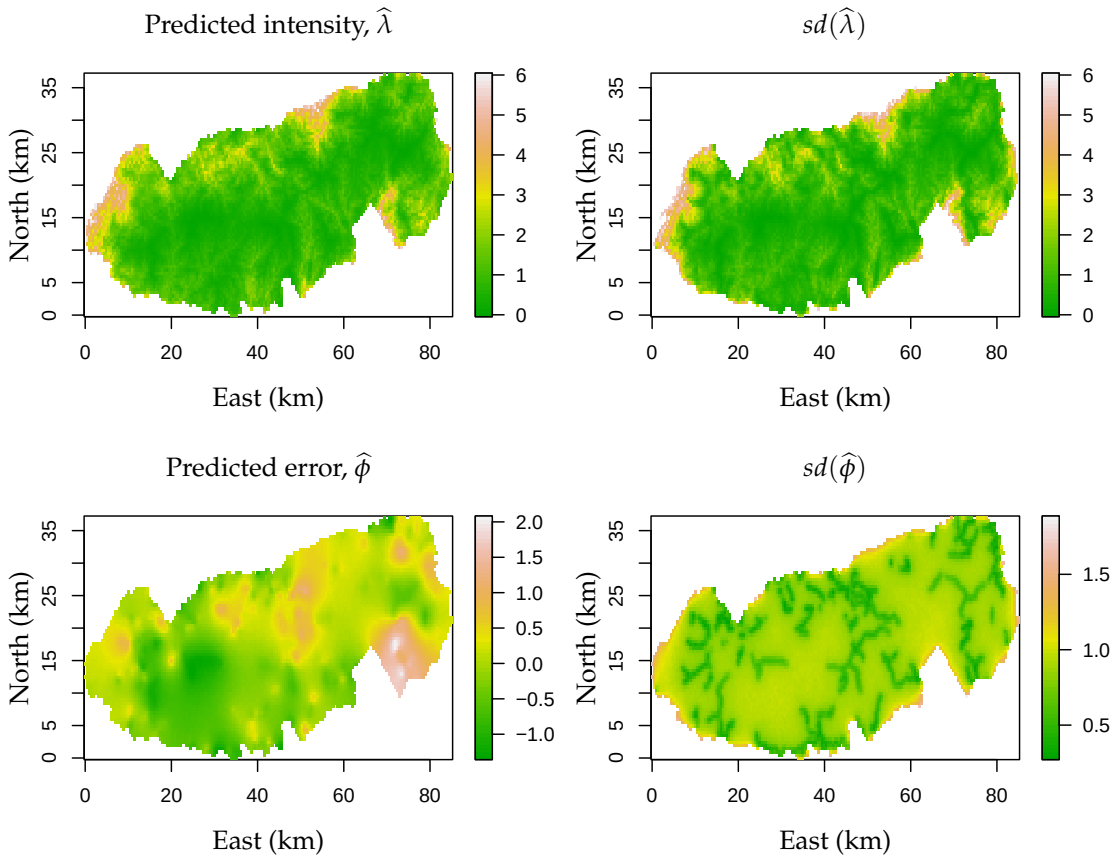

Figure 3.7: Maps of predicted Poisson intensity, $\widehat{\lambda}$ and its standard deviation (top), and the predicted spatial error process, $\widehat{\phi}$ and its standard deviation (bottom) derived from fitting detection Model 1 to repeated count data for the Black-throated Green Warbler (BT).

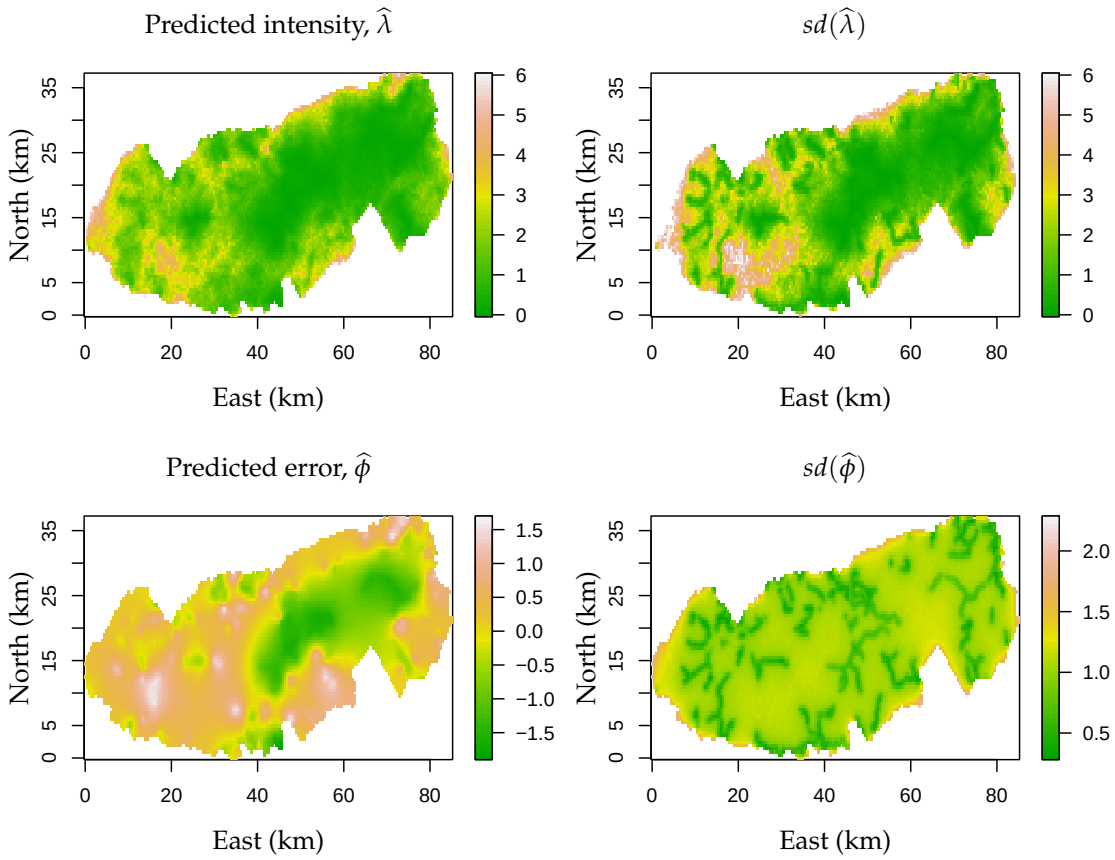

Figure 3.8: Maps of predicted Poisson intensity, $\widehat{\lambda}$ and its standard deviation (top), and the predicted spatial error process, $\widehat{\phi}$ and its standard deviation (bottom) derived from fitting detection Model 1 to repeated count data for the Ovenbird (OB).

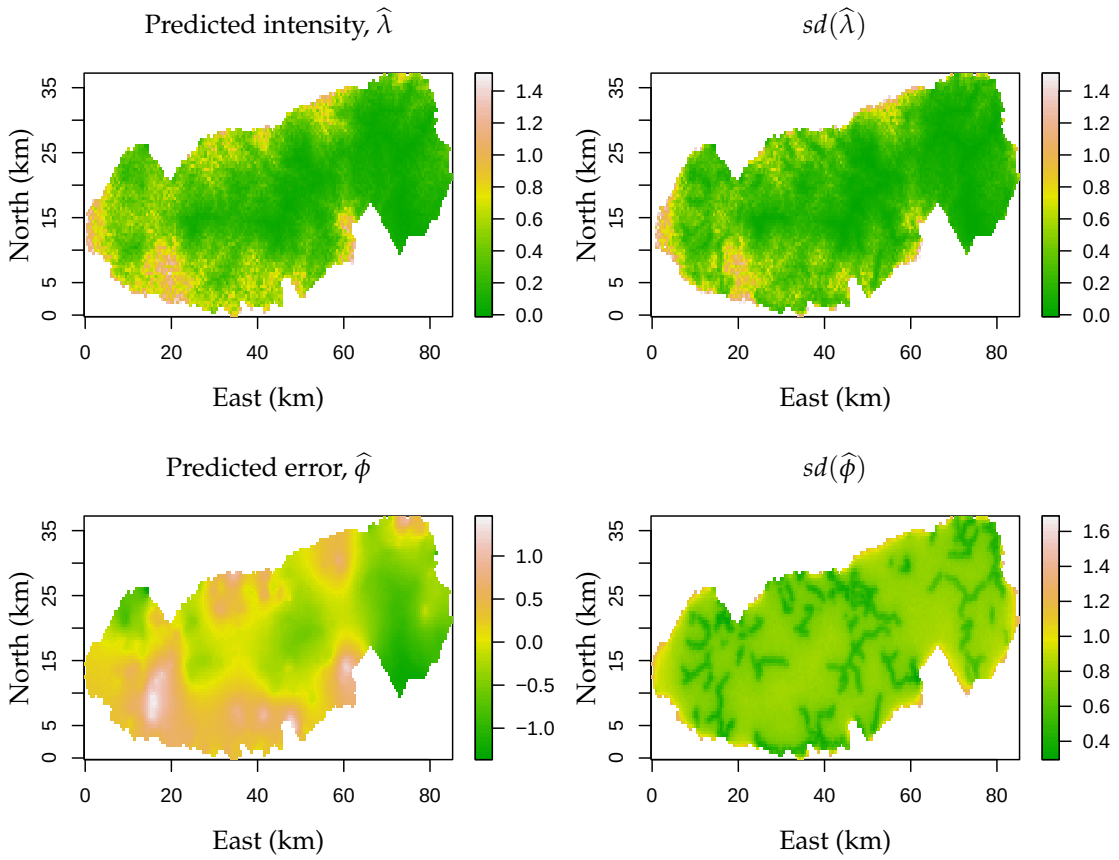

Figure 3.9: Maps of predicted Poisson intensity, $\widehat{\lambda}$ and its standard deviation (top), and the predicted spatial error process, $\widehat{\phi}$ and its standard deviation (bottom) derived from fitting detection Model 1 to repeated count data for the Black and White Warbler (BL).

3.5 Discussion

In this work, we have described a framework for spatial modeling of closed-population bird survey data, where the observations come from repeated counts, removal counts or capture-recapture data recorded at multiple sites within the survey region. The models simultaneously allow for variation in detection probability and for spatial dependence in distribution of birds, and thereby lead to more reliable inference on the causes of variation in bird density across the landscape. When covariates are available at unsurveyed locations, these models will lead to improved mapping of birds species distribution. We have found some evidence that for the Black and White Warbler, the hardest to detect species of those we examined, allowing for heterogeneity in detection probabilities by incorporating covariates in the detection model affects not only the mean detection probability across the survey region, but also affects the estimates of coefficients in the density component of the model. Although the effect was not strong in this case, we believe that in general it is very important to account for variation in detection probability across a region when attempting to model the relationship between bird density and covariates, or when producing maps of relative density from spatial models.

We have also shown how allowing for spatial dependence will improve the mapping of bird distribution. If we had simply fitted a covariate model for density and assumed independence of observations as did Shriner (2001), we would have found that our maps largely reflected the negative relationship with elevation for the three species we considered. Accounting for further variables would have refined the maps somewhat, but even without this, modeling spatial dependence allowed us to identify apparent clusters of birds, along with regions of low density, that would not have been predicted by our covariate models alone.

Our mapping exercise also highlighted the importance of good survey design. The Smoky Mountains survey was conducted mainly along high-use trails, and although many sites were surveyed each year, they tended to be relatively close together. Not only does such site location raise concerns of bias, but it also meant that large areas of the park went unsurveyed. The precision of our prediction maps was very much dependent on the design - precision was only good if we were predicting values close to a survey site. While it may be impractical in difficult terrain to lay a regular grid of survey points across a region of interest, ensuring more even coverage is clearly important for good spatial prediction of bird density.

In many instances, the bird population is surveyed over a number of years, and we may therefore wish to add a temporal component to the models presented in this work. Banerjee et al. (2004) review extentions of CAR models to allow for temporal effects and for spatio-temporal interactions. They present examples of both aligned data, when the same survey sites are monitored over time, and misalignment across years, when the sites change over time. The Smoky Mountain survey data, for example, come from some sites that were monitored every year and others that were not.

We also restricted ourselves in this work to single-species modeling. Extending the models to allow for multiple species is also possible, either through adding species into the covariate model, or through a multivariate conditional autoregressive (MCAR) specification (Gelfand and Vounatsou, 2003; Banerjee et al., 2004). Modeling the diversity of species can also be fitted into our framework. Instead of counts of the number of individual animals, our data would be based on counts of the number of unique species observed at each survey site, and modeling could otherwise proceed as before. It may even be possible to combine diversity and abundance modeling in a single model, although the complexity of such a model and the quantity of data that would need to be analyzed may mean such an exercise is difficult in practice.

There are some important assumptions implicit in our models, some of which are quite questionable given the manner in which bird survey data are gathered. We have already discussed the assumption of population closure, and we believe that this is reasonable provided that the survey is designed appropriately, with repeated survey periods being made in quick succession. Of greater concern is the high chance of errors being made in detecting and identifying individual birds. In capture-recapture bird surveys of the type we have discussed, birds are generally not marked, and the accuracy of the capture histories depends entirely on an observer's ability to correctly identify and distinguish individual birds over the $\mathcal{T}$ survey periods. This is an extremely difficult task, particularly given that birds are often observed in dense forest and that many species are identified by sound alone. One might suppose, however, that a repeated count survey would be less prone to error, as an observer needs only to track the individuals within each period in order to produce a count, which should be easier than identifying individuals across all $\mathcal{T}$ periods. Therefore, although repeated count data may lead to lesser precision in parameter estimates than capture-recapture, the lesser bias induced by identification errors may mean that this method is preferable. We plan to investigate the effects of data errors on bird survey models in future work. We note, however, that the results of our analyses were very similar for all three types of data, implying that if errors exist (and we are certain they do), their effect is much the same regardless of the form of

the data being analyzed.

We have not considered presence-absence modeling in this work in order to focus on comparing methods that model factors affecting bird density. We note, however, that hierarchical presence-absence models can be constructed in a similar way to those we discussed in Section 3.2 by modeling the probability that at least one bird is detected rather than some measure of the number of detectable birds, N. But since N (and Poisson intensity, λ) are not direct measures of absolute abundance or density within some well-defined area around a survey site, meaningful interpretation of "presence" or "absence" is difficult in this context. Combining distance and detection data as per Borchers et al. (1998) and Webster and Pollock (2006) allows us to estimate absolute density, and provided we can sensibly define areas over which we wish to conclude a species is present or not, these methods may allow us to consider presence-absence modeling of bird survey data of the kind we have examined here.

Acknowledgements

We thank Dr Sujit Ghosh of North Carolina State University (NCSU) for his invaluable advice on Bayesian modeling and suggestions for our MCMC algorithm, and Ed Laurent, also of NCSU, for providing us with the covariate values from GIS databases for our mapping exercise.

Appendix A

Table 3.1: Posterior means and 95% credible intervals of parameters for modeling of BT data.

	Count			Removal			Capture-Recapture		
parameter	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
α_0	–0.50 (–0.70,–0.33)	–0.51 (–0.70, –0.33)	–0.50 (–0.69,–0.33)	–0.48 (–0.68,–0.29)	–0.47 (–0.67, –0.30)	–0.47 (–0.65,–0.31)	–0.50 (–0.67,–0.33)	–0.51 (–0.69, –0.34)	–0.47 (–0.64,–0.31)
α_1 (Elev)	–0.69 (–0.87,–0.52)	–0.70 (–0.86, –0.54)	–0.70 (–0.85,–0.54)	–0.67 (–0.83,–0.51)	–0.67 (–0.83, –0.50)	–0.66 (–0.81,–0.53)	–0.66 (–0.81,–0.51)	–0.67 (–0.82,–0.52)	–0.65 (–0.78,–0.52)
α_2 (TRMI)	0.082 (0.008, 0.161)	0.077 (0.010, 0.147)	0.079 (0.017,0.150)	0.073 (0.006, 0.141)	0.078 (0.016, 0.146)	0.075 (0.000,0.143)	0.081 (0.020,0.152)	0.076 (0.008,0.140)	0.080 (0.005,0.151)
γ	0.994 (0.989,0.997)	0.995 (0.990, 0.998)	0.995 (0.989,0.998)	0.995 (0.989,0.998)	0.996 (0.991,0.998)	0.995 (0.990,0.998)	0.996 (0.992,0.998)	0.996 (0.991,0.998)	0.996 (0.993,0.998)
$d_{\max}$	0.75 (0.61,0.97)	0.72 (0.61, 0.94)	0.70 (0.58,0.91)	0.82 (0.67,1.00)	0.71 (0.59, 0.97)	0.76 (0.60,1.09)	0.68 (0.60,0.88)	0.72 (0.58,0.95)	0.68 (0.60,0.84)
σ^2_ϕ	3.2 (1.8,4.8)	3.0 (1.7, 4.7)	2.8 (1.5,4.5)	2.5 (1.6,3.9)	2.1 (1.3, 3.3)	2.2 (1.3,3.4)	2.2 (1.4,3.4)	2.3 (1.2,3.5)	1.9 (1.0,3.3)
β_0	–0.60 (–0.70,–0.50)	–0.66 (–0.79, –0.55)	–0.75 (–0.94,–0.52)	-0.75 (–0.87,–0.63)	–0.82 (–0.96, –0.68)	–0.84 (–1.20,–0.55)	–0.49 (–0.56,–0.42)	–0.54 (–0.62,–0.46)	–0.66 (–0.87,–0.44)
β_1 (Season)	–	0.27 (0.06,0.52)	0.23 (0.01,0.46)	–	0.31 (0.03,0.59)	0.32 (0.03,0.65)	–	0.26 (0.08,0.43)	0.28 (0.06,0.51)
β_2 (Temp)	–	–	0.06 (–0.04,0.17)	–	–	0.04 (–0.12,0.17)	–	–	0.04 (–0.05,0.13)
β_3 (Noise)	–	–	0.10 (–0.10,0.29)	–	–	–0.03 (–0.25,0.21)	–	–	0.16 (0.00,0.34)
β_4 (Cloud1)	–	–	–0.06 (–0.28,0.16)	–	–	–0.23 (–0.58,0.16)	–	–	–0.14 (–0.36,0.08)
β_5 (Cloud2)	–	–	0.10 (–0.11,0.31)	–	–	0.25 (–0.10,0.65)	–	–	0.09 (–0.14,0.33)
$\bar{p}$ (3 min)	0.62 (0.60,0.65)	0.62 (0.60,0.65)	0.62 (0.60,0.65)	0.59 (0.56,0.61)	0.59 (0.56,0.62)	0.59 (0.56,0.62)	0.65 (0.63,0.66)	0.65 (0.63,0.67)	0.65 (0.63,0.66)
$\bar{p}$ (2 min)	0.52 (0.50,0.55)	0.52 (0.50,0.55)	0.52 (0.50,0.55)	0.49 (0.45,0.52)	0.49 (0.46,0.52)	0.49 (0.46,0.52)	0.55 (0.53,0.57)	0.55 (0.54,0.57)	0.55 (0.53,0.57)
$\bar{p}$ (5 min)	0.73 (0.71,0.75)	0.73 (0.71,0.75)	0.73 (0.71,0.75)	0.70 (0.68,0.73)	0.70 (0.68,0.73)	0.70 (0.68,0.73)	0.75 (0.74,0.77)	0.76 (0.74,0.77)	0.75 (0.74,0.77)

Table 3.2: Posterior means and 95% credible intervals of parameters for modeling of OB data.

	Count			Removal			Capture-Recapture		
parameter	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
α_0	–0.61 (–0.89,–0.40)	–0.62 (–0.86, –0.40)	–0.61 (–0.83,–0.41)	–0.61 (–0.81,–0.42)	–0.59 (–0.80,–0.38)	–0.61 (–0.82,–0.41)	–0.64 (–0.87, –0.44)	–0.61 (–0.83, –0.40)	–0.63 (–0.84,–0.43)
α_1 (Elev)	–0.72 (–0.92,–0.51)	–0.68 (–0.91, –0.42)	–0.72 (–0.90,–0.50)	–0.70 (–0.91,–0.48)	–0.67 (–0.86,–0.47)	–0.72 (–0.92,–0.51)	–0.70 (–0.90,–0.48)	–0.67 (–0.87,–0.44)	–0.72 (–0.92,–0.53)
α_2 (TRMI)	–0.12 (–0.19, –0.05)	–0.11 (–0.19,–0.05)	–0.12 (–0.18,–0.05)	–0.10 (–0.16,–0.04)	–0.10 (–0.17,–0.05)	–0.12 (–0.19,–0.05)	–0.11 (–0.18,–0.04)	–0.11 (–0.17,–0.05)	–0.11 (–0.17,–0.04)
γ	0.995 (0.992,0.998)	0.996 (0.993, 0.998)	0.996 (0.991,0.998)	0.996 (0.993,0.998)	0.996 (0.993,0.998)	0.996 (0.993,0.998)	0.996 (0.993,0.998)	0.997 (0.993,0.999)	0.997 (0.994,0.998)
$d_{\max}$	0.66 (0.53,0.82)	0.67 (0.53, 0.81)	0.66 (0.58,0.76)	0.60 (0.52,0.75)	0.68 (0.54,0.79)	0.66 (0.54,0.76)	0.65 (0.53,0.78)	0.66 (0.54,0.78)	0.59 (0.52,0.76)
σ^2_ϕ	3.9 (2.6,5.7)	3.7 (2.2,5.2)	3.7 (2.2,5.7)	3.2 (2.0,5.4)	2.9 (1.9,4.2)	3.2 (2.1,4.5)	3.3 (2.2,4.6)	2.9 (1.8,4.3)	3.2 (2.2,4.4)
β_0	–0.48 (–0.58,–0.38)	–0.44 (–0.54, –0.33)	–0.37 (–0.58,–0.18)	–0.52 (–0.63,–0.41)	–0.52 (–0.63,–0.40)	–0.46 (–0.77,–0.18)	–0.35 (–0.41,–0.28)	–0.33 (–0.41,–0.26)	–0.36 (–0.53,–0.18)
β_1 (Season)	–	–0.26 (–0.56, 0.01)	–0.42 (–0.69,–0.18)	–	–0.08 (–0.41,0.25)	–0.24 (–0.69,0.16)	–	–0.10 (–0.29,0.10)	–0.15 (–0.38,0.08)
β_2 (Temp)	–	–	0.07 (–0.03,0.17)	–	–	0.03 (–0.11,0.16)	–	–	–0.01 (–0.08,0.08)
β_3 (Noise)	–	–	–0.21 (–0.37,–0.03)	–	–	–0.26 (–0.52,–0.04)	–	–	–0.16 (–0.31,–0.02)
β_4 (Cloud1)	–	–	0.16 (–0.10,0.39)	–	–	0.39 (0.10,0.71)	–	–	0.21 (0.00,0.38)
β_5 (Cloud2)	–	–	0.06 (–0.17,0.31)	–	–	0.03 (–0.28,0.32)	–	–	0.18 (–0.01,0.36)
$\bar{p}$ (3 min)	0.65 (0.63,0.67)	0.64 (0.62,0.67)	0.64 (0.61,0.66)	0.64 (0.61,0.67)	0.64 (0.61,0.66)	0.63 (0.60,0.66)	0.68 (0.67,0.69)	0.68 (0.66,0.69)	0.68 (0.66,0.69)
$\bar{p}$ (2 min)	0.55 (0.53,0.58)	0.55 (0.52,0.57)	0.54 (0.51,0.56)	0.54 (0.51,0.57)	0.54 (0.51,0.57)	0.53 (0.50,0.56)	0.59 (0.57,0.60)	0.58 (0.57,0.60)	0.58 (0.56,0.60)
$\bar{p}$ (5 min)	0.76 (0.74,0.77)	0.75 (0.73,0.77)	0.74 (0.72,0.76)	0.75 (0.73,0.77)	0.74 (0.72,0.77)	0.74 (0.71,0.76)	0.78 (0.77,0.79)	0.78 (0.76,0.79)	0.78 (0.76,0.79)

Table 3.3: Posterior means and 95% credible intervals of parameters for modeling of BL data.

	Count			Removal			Capture-Recapture		
parameter	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
α_0	–1.74 (–2.03,–1.48)			–1.75 (–2.03,–1.49)			–1.73 (–1.93,–1.54)	–1.65 (–1.83,–1.48)	–1.62 (–1.81,–1.44)
α_1 (Elev)	–0.71 (–0.93,–0.51)			–0.72 (–0.92,–0.52)			–0.78 (–0.95,–0.62)	–0.75 (–0.91,–0.58)	–0.74 (–0.90,–0.57)
α_2 (TRMI)	–0.22 (–0.33,–0.12)			–0.22 (–0.31,–0.13)			–0.22 (–0.34,–0.12)	–0.21 (–0.29,–0.13)	–0.22 (–0.32,–0.10)
$\gamma = 0.995$									
$d_{\max}$	2.10 (1.32,2.97)			1.78 (1.28,2.75)			1.47 (1.26,1.79)	1.21 (1.05,1.45)	1.02 (0.90,1.17)
σ^2_ϕ	4.04 (1.85,7.70)			3.34 (1.95,6.18)			1.77 (1.12,2.51)	1.52 (0.72,2.40)	1.42 (0.76,2.21)
β_0	–0.98 (–1.13,–0.84)			–0.98 (–1.14,–0.84)			–0.87 (–0.99,–0.74)	–0.86 (–0.99,–0.73)	–0.66 (–0.93,–0.36)
β_1 (Season)	–			–			–	–0.99 (–1.86,0.06)	–1.21 (–2.05,–0.51)
β_2 (Temp)	–			–			–	–	0.04 (–0.10,0.18)
β_3 (Noise)	–			–			–	–	–0.20 (–0.48,0.07)
β_4 (Cloud1)	–			–			–	–	0.05 (–0.29,0.39)
β_5 (Cloud2)	–			–			–	–	–0.18 (–0.54,0.15)
$\bar{p}$ (3 min)	0.53 (0.49,0.56)			0.53 (0.49,0.56)			0.56 (0.53,0.59)	0.50 (0.45,0.57)	0.49 (0.44,0.54)
$\bar{p}$ (2 min)	0.43 (0.39,0.46)			0.43 (0.39,0.46)			0.46 (0.43,0.49)	0.41 (0.36,0.47)	0.39 (0.36,0.44)
$\bar{p}$ (5 min)	0.65 (0.62,0.68)			0.65 (0.61,0.68)			0.68 (0.65,0.70)	0.62 (0.57,0.69)	0.61 (0.56,0.66)

Appendix B

Here we present our MCMC algorithms for the analyses of Section 3.4 for count and removal data. The algorithm for capture-recapture data differs only in the likelihood component, with the multinomial model of Equation (3.2) replacing the binomial model of Equation (3.1). The algorithms are based on Gibbs sampling, in which at each iteration we sample from the full conditional distribution of a parameter conditional on the most recent values of all other parameters. For our models the full conditional distributions for most parameters do not have a familiar form, and we instead employ Metropolis or Metropolis-Hastings steps to update these parameters. For our algorithm, for $i = 1, \ldots, L$, we write the full model for count and removal data as:

$$\begin{aligned} y_{ij}|N_i, \boldsymbol{\beta}, \mathbf{Z}_i &\sim \text{Bin}(N_i - M_{ij}, p_{ij}) \\ N_i|u_i &\sim \text{Poisson}(e^{u_i}) \end{aligned}$$

where

$$\text{logit}(p_{ij}) = \log(\mathcal{T}_j) + \mathbf{Z}_i\boldsymbol{\beta},$$

and u_i is modeled using a proper Gaussian CAR prior,

$$u_i|\mathbf{X}_i, \boldsymbol{\alpha}, \tau_\phi, \gamma, d_{\max} \sim \text{N}\left(\mathbf{X}_i\boldsymbol{\alpha} + \gamma \sum_{i' \neq i} \frac{w_{ii'}}{w_{i.}}(u_{i'} - \mathbf{X}_{i'}\boldsymbol{\alpha}), \frac{1}{w_{i.}\tau_\phi}\right).$$

The CAR weights $w_{ii'}$ are functions of $d_{\max}$ as defined in the text and $M_{ij} = \sum_{j<j'} y_{ij'}$ is the number of animals removed prior to occasion j, with $M_{i1} = 0$. For count data, $M_{ij} = 0$ for $j = 1, \ldots, T$ as no animals are removed during the T survey occasions.

The joint posterior distribution of the parameters is given by

$$\wp(\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma, d_{\max}, \tau_\phi|\mathbf{Y}, \mathbf{X}, \mathbf{Z}) \propto \wp(\mathbf{Y}|\mathbf{N}, \boldsymbol{\beta}, \mathbf{Z})\wp(\mathbf{N}|\boldsymbol{u})\wp(\boldsymbol{u}|\boldsymbol{\alpha}, \mathbf{X}, \gamma, d_{\max}, \tau_\phi)\wp(\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma, d_{\max}, \tau_\phi).$$

We choose independent priors for $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, γ ,$d_{\max}$ and τ_ϕ, and so

$$\wp(\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma, d_{\max}, \tau_\phi) = \wp(\boldsymbol{\alpha})\wp(\boldsymbol{\beta})\wp(\gamma)\wp(d_{\max})\wp(\tau_\phi).$$

For the tth update, we use the following steps.

Step 1

For $i = 1, \ldots, L$, sample $N_i^{(t)}$ from $\wp(N_i|\mathbf{Y}_i, u_i^{(t-1)}, \boldsymbol{p}_i^{(t-1)})$ where $\boldsymbol{p}_i = [p_{i1}, p_{i2}, p_{i3}]$ with

$$\text{logit}(p_{ij}) = \log(\mathcal{T}_j) + \mathbf{Z}_i\boldsymbol{\beta},$$

using a Metropolis-Hastings step with a negative binomial candidate and tuning parameter r_n:

- Generate N_i^* from $q(x|N_i^{(t-1)})$, the NegBin $\left(r_n, r_n/(r_n + N_i^{(t-1)})\right)$ density.
- Calculate
$$r = \frac{\wp(\mathbf{Y}_i|N_i^*, \boldsymbol{\beta}^{(t-1)}, \mathbf{Z}_i)\wp(N_i^*|u_i^{(t-1)})q(N_i^{(t-1)}|N_i^*)}{\wp(\mathbf{Y}_i|N_i^{(t-1)}, \boldsymbol{\beta}^{(t-1)}, \mathbf{Z}_i)\wp(N_i^{(t-1)}|u_i^{(t-1)})q(N_i^*|N_i^{(t-1)})}$$
- Let $N_i^{(t)} = N_i^*$ with probability $\min(r, 1)$, otherwise $N_i^{(t)} = N_i^{(t-1)}$.

Step 2

For $i = 1, \ldots, L$, sample $u_i^{(t)}$ from $\wp(u_i|\mathbf{N}_i^{(t)}, u_1^{(t)}, \ldots, u_{i-1}^{(t)}, u_{i+1}^{(t-1)}, \ldots, u_L^{(t-1)}, \gamma^{(t-1)}, d_{\max}^{(t-1)}, \tau_\phi^{(t-1)})$ using a Metropolis step with a Gaussian candidate with tuning parameter a_u:

- Generate u_i^* from $q(x|u_i^{(t-1)})$, the N $\left(u_i^{(t-1)}, a_u\right)$ density.
- Calculate
$$r = \frac{\wp(N_i^{(t)}|u_i^*)\wp(u_i^*|u_1^{(t)}, \ldots, u_{i-1}^{(t)}, u_{i+1}^{(t-1)}, \ldots, u_L^{(t-1)}, \gamma^{(t-1)}, d_{\max}^{(t-1)}, \tau_\phi^{(t-1)})}{\wp(N_i^{(t)}|u_i^{(t-1)})\wp(u_i^{(t-1)}|u_1^{(t)}, \ldots, u_{i-1}^{(t)}, u_{i+1}^{(t-1)}, \ldots, u_L^{(t-1)}, \gamma^{(t-1)}, d_{\max}^{(t-1)}, \tau_\phi^{(t-1)})}$$
- Let $u_i^{(t)} = u_i^*$ with probability $\min(r, 1)$, otherwise $u_i^{(t)} = u_i^{(t-1)}$.

Step 3

Sample $\boldsymbol{\alpha}^{(t)}$ from the full conditional

$$\wp(\boldsymbol{\alpha}^{(t)}|\boldsymbol{u}^{(t)}, \mathbf{X}, \gamma^{(t-1)}, d_{\max}^{(t-1)}, \tau_\phi^{(t-1)}) \propto \wp(\boldsymbol{u}^{(t)}|\boldsymbol{\alpha}^{(t-1)}, \mathbf{X}, \gamma^{(t-1)}, d_{\max}^{(t-1)}, \tau_\phi^{(t-1)})\wp(\boldsymbol{\alpha}^{(t-1)}).$$

Here we choose a flat prior, so $\wp(\boldsymbol{\alpha}^{(t-1)}) \propto 1$, leading to a Gaussian full conditional,

$$\wp(\boldsymbol{\alpha}^{(t)}|\boldsymbol{u}^{(t)}, \mathbf{X}, \gamma^{(t-1)}, d_{\max}^{(t-1)}, \tau_\phi^{(t-1)}) = \text{MVN}\left((\mathbf{X}'\boldsymbol{\Sigma}_\phi^{(t-1)}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}_\phi^{(t-1)}\boldsymbol{u}^{(t)}, (\mathbf{X}'\boldsymbol{\Sigma}_\phi^{(t-1)}\mathbf{X})^{-1}\right),$$

where $\boldsymbol{\Sigma}_\phi$ is a function of γ and (through $\mathbf{C}$, the standardized weight matrix) $d_{\max}$.

Step 4

Again choosing a uniform prior, $\wp(\boldsymbol{\alpha}^{(t-1)}) \propto 1$, sample $\boldsymbol{\beta}^{(t)}$ from its full conditional, $\wp(\mathbf{Y}|\mathbf{N}, \boldsymbol{\beta}, \mathbf{Z})$, using a Metropolis step with a Gaussian candidate with tuning parameter a_β:

- Generate $\boldsymbol{\beta}^*$ from $q(x|\boldsymbol{\beta}^{(t-1)})$, the MVN $\left(\boldsymbol{\beta}^{(t-1)}, a_\beta \mathbf{I}\right)$ density.
- Calculate
$$r = \frac{\wp(\mathbf{Y}|\mathbf{N}^{(t)}, \boldsymbol{\beta}^*, \mathbf{Z})}{\wp(\mathbf{Y}|\mathbf{N}^{(t)}, \boldsymbol{\beta}^{(t-1)}, \mathbf{Z})}$$
- Let $\boldsymbol{\beta}^{(t)} = \boldsymbol{\beta}^*$ with probability $\min(r, 1)$, otherwise $\boldsymbol{\beta}^{(t)} = \boldsymbol{\beta}^{(t-1)}$.

Step 5

Choosing a prior of $\wp(\tau_\phi) \propto \tau_\phi^{-3/2}$ (equivalent to a uniform prior for σ_ϕ), sample $\tau_\phi^{(t)}$ from the full conditional,

$$\wp(\tau_\phi|\mathbf{X}, \boldsymbol{u}^{(t)}, \boldsymbol{\alpha}^{(t)}, \gamma^{(t-1)}, d_{\max}^{(t-1)}) = \text{Gamma}\left(\frac{L-1}{2}, \frac{1}{0.5(\boldsymbol{u} - \mathbf{X}\boldsymbol{\alpha})' \boldsymbol{\Sigma}_\phi^{(t-1)} (\boldsymbol{u} - \mathbf{X}\boldsymbol{\alpha})}\right).$$

Step 6

With a uniform prior, we sample values from the full conditional $\wp(\gamma|\boldsymbol{u}, \boldsymbol{\alpha}, \tau_\phi) \propto \wp(\boldsymbol{u}|\mathbf{X}, \boldsymbol{\alpha}, \gamma, d_{\max}, \tau_\phi)$ using a Metropolis step. As γ must lie within $(-1, 1)$, we simulated values of $\theta = \text{logit}((\gamma + 1)/2)$ using a Gaussian candidate, with tuning parameter a_γ.

- Generate θ^* from $q(x|\theta^{(t-1)})$, the $\text{N}(\theta^{(t-1)}, a_\gamma)$ density, and compute $\gamma^* = \text{logit}^{-1}(\theta^*)$
- Calculate
$$r = \frac{\wp(\boldsymbol{u}^{(t)}|\mathbf{X}, \boldsymbol{\alpha}^{(t)}, \gamma^*, d_{\max}^{(t-1)}, \tau_\phi^{(t)})}{\wp(\boldsymbol{u}^{(t)}|\mathbf{X}, \boldsymbol{\alpha}^{(t)}, \gamma^{(t-1)}, d_{\max}^{(t-1)}, \tau_\phi^{(t)})}$$
- Let $\gamma^{(t)} = \gamma^*$ with probability $\min(r, 1)$, otherwise $\gamma^{(t)} = \gamma^{(t-1)}$.

Step 7

With a uniform prior, we use a Metropolis-Hastings step to sample values from the full conditional $p(d_{\max}|\boldsymbol{u}, \boldsymbol{\alpha}, \tau_\phi) \propto \wp(\boldsymbol{u}|\mathbf{X}, \boldsymbol{\alpha}, \gamma, d_{\max}, \tau_\phi)$. Because we add the constraint $d_{\max} > d_l$ in order to ensure the number of isolated points is small, we simulate values of $D = d_{\max} - d_l$ using a gamma candidate with mean D and variance, a_d, the tuning parameter.

- Generate D^* from $q(x|D^{(t-1)})$, the Gamma $\left(\frac{D^{(t-1)}}{a_d}, \frac{a_d}{D^{(t-1)}}\right)$ density, and compute $d_{\max}^* = D^* + d_l$
- Calculate
$$r = \frac{\wp(\boldsymbol{u}^{(t)}|\mathbf{X}, \boldsymbol{\alpha}^{(t)}, \gamma^{(t)}, d_{\max}^*, \tau_\phi^{(t)}) q(D^{(t-1)}|D^*)}{\wp(\boldsymbol{u}^{(t)}|\mathbf{X}, \boldsymbol{\alpha}^{(t)}, \gamma^{(t)}, d_{\max}^{(t-1)}, \tau_\phi^{(t)}) q(D^*|D^{(t-1)})}$$
- Let $d_{\max}^{(t)} = d_{\max}^*$ with probability $\min(r, 1)$, otherwise $d_{\max}^{(t)} = d_{\max}^{(t-1)}$.

Bibliography

Alho, J. M. (1990). Logistic regression in capture-recapture models, *Biometrics* **46**: 639–649.

Alldredge, M. W. (2004). *Avian Point Count Survyes: Estimating Components of the Detection Process*, PhD thesis, North Carolina State University.

Banerjee, S., Carlin, B. P. and Gelfand, A. E. (2004). *Hierarchical Modeling and Analysis for Spatial Data*, Boca Raton: Chapman and Hall.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems, *Journal of the Royal Statistical Society, Series B* **36**: 192–236.

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

Chao, A. (1987). Estimating the population size for capture-recapture data with unequal catchability, *Biometrics* **43**: 783–791.

Coull, B. A. and Agresti, A. (1999). The use of mixed-logit models to reflect heterogeneity in capture-recapture studies, *Biometrics* **55**: 294–301.

Cressie, N. (1993). *Statistics for Spatial Data*, 2nd edn, New York: Wiley.

Diggle, P. J., Tawn, J. A. and Moyeed, R. A. (1998). Model-based geostatistics (with discussion), *Applied Statistics* **47**: 299–350.

Dorazio, R. M., Jelks, H. L. and Jordan, F. (2005). Improving removal-based estimates of abundance by sampling a population of spatially distinct subpopulations, *Biometrics* **61**: 1093–1101.

Dorazio, R. M. and Royle, J. A. (2003). Mixture models for estimating the size of a closed population when capture rates vary among individuals, *Biometrics* **59**: 351–364.

Farnsworth, G. L., Pollock, K. H., Nichols, J. D., Simons, T. R., Hines, J. E. and Sauer, J. R. (2002). A removal model for estimating detection probabilities from point-count surveys, *The Auk* **119**: 414–425.

Forsyth, D. M., Link, W. A., Webster, R., Nugent, G. and Warburton, B. (2006). Nonlinearity and seasonal bias in an index of brushtail possum abundance, *Journal of Wildlife Management* **(to appear)**.

Gelfand, A. E., Valenti, J., Wu., S. and Sirmans, C. F. (2006). Building statistical models to analyze species distributions, *Ecological Applications* **(to appear)**.

Gelfand, A. E. and Vounatsou, P. (2003). Proper multivariate conditional autoregressive models for spatial data analysis, *Biostatistics* **4**: 11–25.

Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (2004). *Bayesian data analysis*, 2nd edn, Boca Raton: Chapman and Hall.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian resolution of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**: 721–741.

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

Hrafnkelsson, B. and Cressie, N. (2003). Hierarchical modeling of count data with application to nuclear fall-out, *Environmental and Ecological Statistics* **10**: 179–200.

Huggins, R. M. (1989). On the statistical analysis of capture experiments, *Biometrika* **76**: 133–140.

Lichstein, J. W., Simons, T. R., Shriner, S. A. and Franzreb, K. E. (2002). Spatial autocorrelation and autoregressive models in ecology, *Ecological Monographs* **72**: 445–463.

Link, W. A. (2003). Nonidentifiability of population size from capture-recapture data with heterogeneous detection probabilities, *Biometrics* **59**: 1123–1130.

Norris, J. L. and Pollock, K. H. (1996). Nonparametric MLE under two closed capture-recapture models with heterogeneity, *Biometrics* **52**: 639–649.

Otis, D. L., Burnham, K. P., White, G. C. and Anderson, D. R. (1978). Statistical inference from capture data on closed animal experiments, *Wildlife Monographs* **62**: 1–135.

Pinder, J. E., Kroh, G. C., White, J. D. and Basham May, A. M. (1997). The relationships between vegetation type and topography in lassen volcanic national park, *Plant Ecology* **131**: 17–29.

Pledger, S. (2000). Unified maximum likelihood estimates for closed capture-recapture models using mixtures, *Biometrics* **56**: 434–442.

Pollock, K. H., Nichols, J. D., Brownie, C. and Hines, J. E. (1990). Statistical inference for capture-recapture experiments, *Wildlife Monographs* **107**: 1–97.

Royle, J. A. (2004). N-mixture models for estimating population size from spatially replicated counts, *Biometrics* **60**: 108–115.

Royle, J. A., Link, W. A. and Sauer, J. R. (2002). Statistical mapping of count survey data, *in* J. M. Scott, P. J. Heglund, M. L. Morrison, J. B. Haufler, M. G. Raphael, W. A. Wall and F. B. Samson (eds), *Predicting Species Occurences: Issues of Accuracy and Scale*, Washington: Island Press.

Shriner, S. A. (2001). *Distribution of breeding birds in Great Smoky Mountains National Park*, PhD thesis, North Carolina State University.

Spiegelhalter, D., Thomas, A., Best, N. and Lunn, D. (2003). *WinBUGS User Manual, Version 1.4*.

Sun, D., Tsutakaw, R. K. and Speckman, P. L. (1999). Posterior distribution of hierarchical models using car(1) distributions, *Biometrika* **86**: 341–350.

Webster, R. A. and Pollock, K. H. (2006). Modeling observers effects on animal density and detection for combined distance and capture-recapture data, *(in preparation)* .

Webster, R. A., Pollock, K. H. and Ghosh, S. K. (2006). Bayesian spatial modeling of data from unit-count surveys of fish in streams, *(in preparation)* .

Wikle, C. (2002). Spatial modelling of count cata: a case study in modelling breeding bird survey data on large spatial domains, *in* A. B. Lawson and G. T. Denison (eds), *Spatial Cluster Modelling*, Boca Raton: Chapman and Hall/CRC.

Wyatt, R. J. (2002). Estimating riverine fish population size from single- and multiple-pass removal sampling using a hierarchical model, *Canadian Journal of Fisheries and Aquatic Science* **59**: 695–706.

Chapter 4

Modeling Observer Effects on Animal Density and Detection for Combined Distance and Capture-Recapture Data

Raymond A. Webster and Kenneth H. Pollock

ABSTRACT

The estimation of animal density using the methods of distance sampling generally ignores the effects of survey transects or observers on local animal density or detection probability. Typically, animal density is assumed to be uniform around the transect, which fails to allow for attraction or avoidance of the transect by some animal species, while detection probability must be perfect on the transect, something that is unlikely to be true for many species. We propose parametric models for application to combined distance and capture-recapture survey data from both line and point transect surveys that allow for two types of movement: permanent avoidance or attraction to a transect, or temporary displacement of animals in the vicinity of a transect. The models have a simple form, with parameters that quantify the impact transects and observers have on local density. We combine these density models with logistic-linear models for detection probability using the likelihood framework of Borchers et al. (1998) for combined distance and capture-recapture data. This allows us to separately estimate the parameters of both the density and detection components of the model, which is not possible using distance sampling models alone. Through a simulation study, we show that provided sufficient animals are detected, the model parameters have little bias, and lead to improved estimates of density over the simple uniform model, particularly for line transect surveys. Model selection by

AIC generally chooses the correct density model. We apply our models to a point transect survey of birds in the Great Smoky Mountains National Park, and find some evidence of observer effects on local bird density.

4.1 Introduction

Distance sampling (Buckland et al., 2001) has been widely used to estimate animal density from line and point transects, but standard distance models require some restrictive and often unrealistic assumptions. Two important assumptions are that that the distribution of animals is uniform about the transect, and the probability of detecting an animal is 1 on the transect. Regarding the first assumption, movement away from (avoidance) or towards (attraction) the observer or transect prior to detection will affect the local density of animals during the survey, leading to departures from a uniform density near the transect. When the methods of distance sampling are used, this results in negative biases in estimators of animal density when there is avoidance of the observer, while attraction will cause positive bias in density estimation (Buckland et al., 2001).

While the assumption of perfect detection probability on the transect may be reasonable for many large, easily detected animal species, it is unlikely to hold for small mammals, insects and birds. Some animals may become more or less active in the vicinity of the observer, positively or negatively affecting their detection probabilities. For example, some bird species tend to fall silent when an observer is near, while others may increase their singing. As discussed in Buckland et al. (2001), if detection is imperfect on the transect, then distance sampling models will produce density estimators with negative bias.

It is not possible from the distance data alone to distinguish observer effects on detection from those on density because both lead to fewer detection distances close to the transect than expected under the usual assumptions of distance sampling. In order to allow for these effects when estimating animal density, we require additional information. Past approaches have used auxiliary survey data to correct for the effect of movement and imperfect detection at zero distance within the general distance sampling framework (Turnock and Quinn, 1991; Buckland and Turnock, 1992). For some animal species, however, it is possible to record multiple detection data, either in form of double counts or capture-recapture data, along with distances from the observer. With capture-recapture methods, animals are observed over multiple independent survey periods, and a record is kept of whether each observed animal was detected in a given survey period or not (Pollock et al., 1990). In double counting, two observers independently record whether they detect an individual

animal or not, and therefore such a survey yields data of the same type as a two-period capture-recapture survey. It is this type of data that is the focus of our work. Such data can be analyzed using the likelihood approach of Borchers et al. (1998), which does not require perfect detection on the transect and is flexible enough to accommodate models more general than the uniform density function for the density of detectable animals.

We propose two alternatives to the uniform density for the distribution of animals. The first, based on the Gaussian density, assumes that animals avoid the transect even in the absence of an observer, and so there is permanently lower or higher density around the transect. This is likely when a transect is a trail, road, or ridge or some other feature with physical characteristics which affect animal density. The second case is when the observer causes non-random movement of the animals prior to detection, the situation that has been more frequently examined in the literature (see Buckland et al., 2001; Turnock and Quinn, 1991). For this case, we propose a probability density function for animal distances based on the "Mexican hat" function, which has a mode due to movement towards or away from the observer prior to detection. Our models are appealing because the parameters have direct interpretations, quantifying the effects that transects and observers have on local density in terms of their magnitude and range.

For modeling the relationship between detection probability and distance, we apply logistic-linear models, which have been proposed in the capture-recapture literature for modeling heterogeneity in capture probability due to covariates (e.g., Huggins, 1989; Alho, 1990). We consider either a simple linear relationship between the logit of detection probability and distance, or a more flexible quadratic model. The quadratic model may account for transect effects on detectability as it allows detection probability to initially increase with increasing distance from the transect before finally decreasing to zero at larger distances. Such models would violate the assumptions of distance sampling, which require detection functions to be monotonic, decreasing, with perfect detectability on the transect (Buckland et al., 2001). The logistic framework also allows covariates other than distance to be easily incorporated into the models of detection probability. The linear and quadratic are certainly not the only plausible models, and different types of models may better capture the effects of transects on detection probability. However, we restrict ourselves to these simple models in order to focus attention on our models for local density.

We begin by presenting our new parametric models for accounting for observer and transect effects on animal density for line and point transects in Section 4.2. This is followed by a brief review of logistic models for detection probability (Section 4.3) and a discussion of how the density

and detection models can be incorporated into a single likelihood for parameter estimation from capture-recapture data as per Borchers et al. (1998) (Section 4.4). In Section 4.5, we present the results of a simulation study to investigate the properties of parameter estimators, particularly estimators of abundance and density. We also consider how well Akaike's Information Criterion (AIC) (Akaike, 1973) can select the correct model of the three density models, the uniform, Gaussian, and Mexican hat. An application of the models to point transect data from a large bird population survey of the Great Smoky Mountains National Park is presented in Section 4.6, which is followed by a discussion of the implications of our work in Section 4.7.

4.2 Modeling Observer Effects on Animal Density

4.2.1 Animal Distribution Models for Line Transects

In general, line transect sampling (Buckland et al., 2001) assumes that animals are uniformly distributed in two-dimensional space around the line transect. As such surveys require perpendicular distances from the transect to be recorded, it follows that the distribution of distances is also uniform. If x denotes the distance of a randomly selected animal from the population within distance W of the line transect, the probability density function (pdf) of x is

$$\wp_u(x) = \frac{1}{W}, 0 \leq x \leq W \tag{4.1}$$

However, for many animal species, transects are often located for convenience along trails, roads or other features which permit easier access for observers. Although we do not advocate non-random placement of transects if it can be avoided, such surveys are relatively common, and unless the entire population of interest is located in the vicinity of potential transects, such non-random placement will lead to biased estimates of animal density. Animal density is likely to be higher or lower around trails or roads than in more typical habitat. For example, for birds, some raptors may be attracted to roads because of the presence of animals killed by vehicles, while other species will prefer less modified habitat than that found along roadsides.

We propose a modified version of (4.1) based on adding a constant times the Gaussian pdf to the numerator of Equation (4.1) which allows for density to be higher or lower in the vicinity of the

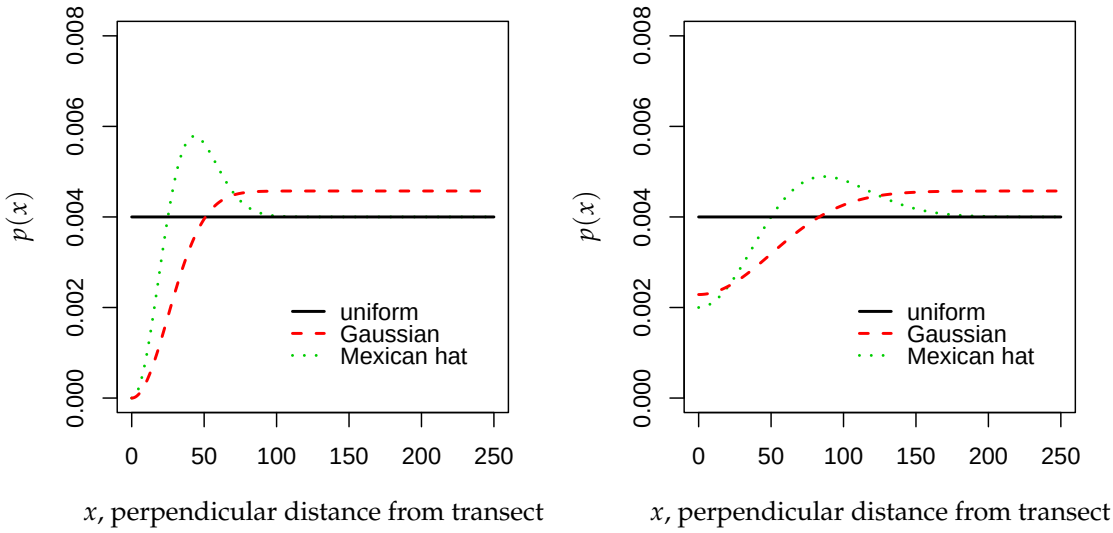

Figure 4.1: Comparison of the uniform, Gaussian and Mexican hat animal density functions within distance $W = 250$ m of a line transect for strong avoidance over a short distance ($\alpha = -1/\phi(0)$, $\sigma = 25$) and moderate avoidance over a long distance from the transect ($\alpha = -1/2\phi(0)$, $\sigma = 50$).

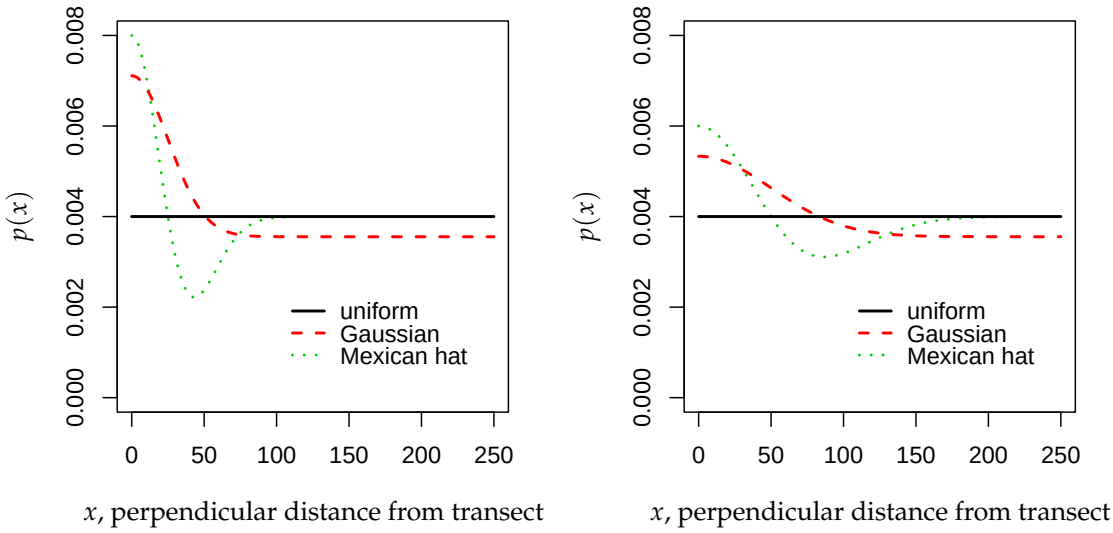

Figure 4.2: Comparison of the uniform, Gaussian and Mexican hat animal density functions within distance $W = 250$ m of a line transect for strong attraction over a short distance ($\alpha = 1/\phi(0)$, $\sigma = 25$) and moderate attraction over a long distance from the transect ($\alpha = 1/2\phi(0)$, $\sigma = 50$).

transect but uniform at greater distances:

$$\wp_g(x) = \frac{1 + \alpha\phi\left(\frac{x}{\sigma}\right)}{W + \alpha\sigma\left[\Phi\left(\frac{W}{\sigma}\right) - \frac{1}{2}\right]}, 0 \leq x \leq W \tag{4.2}$$

where $\phi(x)$ and $\Phi(x)$ are the standard normal pdf and distribution function respectively. The parameter α controls the magnitude of the effect of the trail or road on density on the transect, while σ is a scale parameter that determines how far into the habitat there is an effect on animal density. For greater density at and around the transect, $\alpha > 0$, while $\alpha < 0$ means there are fewer animals near the transect than we would expect under a uniform distribution. When $\alpha = 0$, Equation (4.2) reduces to the uniform density, Equation (4.1). Note that $\sigma > 0$ and for $\wp_g(x) \geq 0$ we require $\alpha \geq -1/\phi(0)$. The denominator of (4.2) is a constant, the form of which ensures that $\int_0^W \wp_g(x)dx = 1$.

Although the Gaussian model allows for the long term effects transects on density, an estimator of density based on this model will still be a biased due to density within distance W of the transects being unrepresentative of the entire survey region. We consider this problem further in the Discussion, Section 4.7.

The other important case is when the observer rather than the transect affects local density, with movement of animals towards or away from the observer prior to detection. This time we modify the uniform density model by adding a component based on the Mexican hat function, which has a mode to the right of $x = 0$. Like the Gaussian pdf, the 1-dimensional Mexican hat function has two parameters and has the following form:

$$m(x) = \left(1 - \frac{x^2}{\sigma^2}\right) e^{-x^2/2\sigma^2}. \tag{4.3}$$

As with the Gaussian model, the Mexican hat pdf of distances is formed by adding a constant, α, times Equation (4.3) to the uniform pdf, and dividing by a factor that ensures integration to 1:

$$\wp_m(x) = \frac{1 + \alpha\left(1 - \frac{x^2}{\sigma^2}\right)\phi\left(\frac{x}{\sigma}\right)}{W\left[1 + \alpha\phi\left(\frac{W}{\sigma}\right)\right]}, 0 \leq x \leq W. \tag{4.4}$$

The parameter α determines the strength of the movement of animals towards or away from the observer, while the scale parameter σ has a similar interpretation as in the Gaussian model, measuring

the range of influence of the observer. In the Mexican hat model, for $\wp_m(x) \geq 0$ we require

$$-\frac{1}{\phi(0)} \leq \alpha \leq \frac{1}{2\phi(\sqrt{3})},$$

and as before, the denominator is a constant ensuring that $\int_0^W \wp_m(x)dx = 1$. Equation (4.4) uses a version of the Mexican hat function which integrates to 0 over the range $(0, \infty)$. This is a very attractive feature when modeling the change in density due to movement, as it means the increase in animal density in one direction (towards or away from the observer) is exactly matched by a decrease in the other direction.

Figures 4.1 and 4.2 compare the three density models for positive and negative values of α respectively and for two values of σ. Strong local avoidance over a relatively short distance, and weaker avoidance over a long distance are shown in Figure 4.1, while Figure 4.2 gives similar graphs for attraction.

4.2.2 Animal Distribution Models for Point Transects

Point transects are particularly common in bird population surveys, and again the methods of distance sampling typically assume a uniform distribution of animals in two-dimensional space. As given in Buckland et al. (2001), this leads to a form for the pdf of radial distance, r, of animals within W of the observer of

$$\wp_u(r) = \frac{2r}{W^2}, 0 \leq r \leq W. \tag{4.5}$$

Here $r^2 = x^2 + y^2$ where x and y are the distances of an animal from the observer along the north-south and east-west axes respectively. Equation (4.5) means that the probability of an animal occurring increases linearly with distance from the point, due to the greater area that surrounds the point within larger distances.

The Gaussian model is less likely to be useful for point transects, as it attempts to model the effect of a permanent modification to the habitat surrounding the transect, something which is unlikely to occur in a circular pattern about a point. Nevertheless, we present a Gaussian model for point transects that, as with Equation (4.2), is derived by modifying the uniform model, and is characterized by the pdf

$$\wp_g(r) = \frac{2r\left[1 + \alpha\phi\left(\frac{r}{\sigma}\right)\right]}{W^2 + 2\alpha\sigma^2\left[\phi(0) - \phi\left(\frac{W}{\sigma}\right)\right]}, 0 \leq r \leq W. \tag{4.6}$$

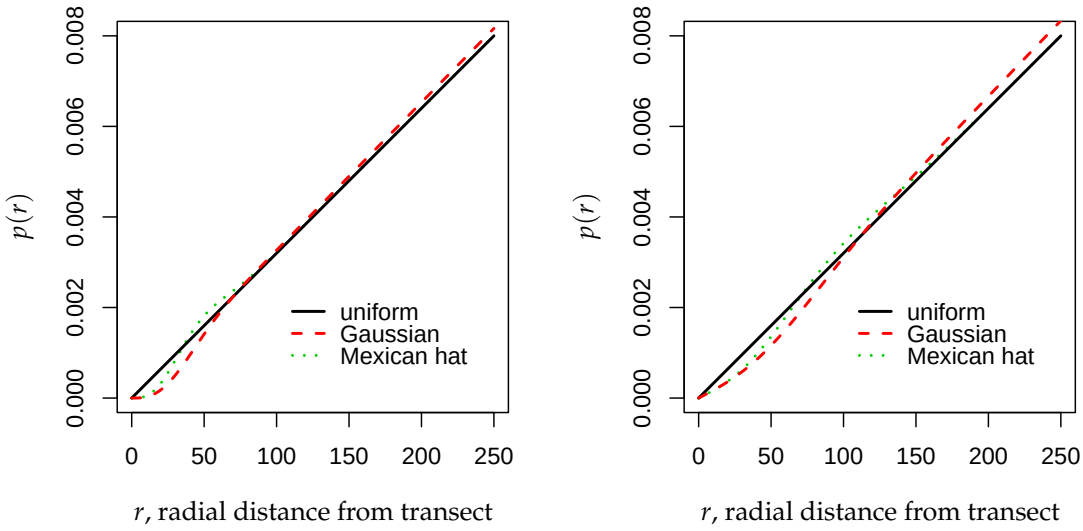

Figure 4.3: Comparison of the uniform, Gaussian and Mexican hat animal density functions within distance $W = 250$ m of a point transect for strong avoidance over a short distance ($\alpha = -1/\phi(0)$, $\sigma = 25$) and moderate avoidance over a long distance from the transect ($\alpha = -1/2\phi(0)$, $\sigma = 50$).

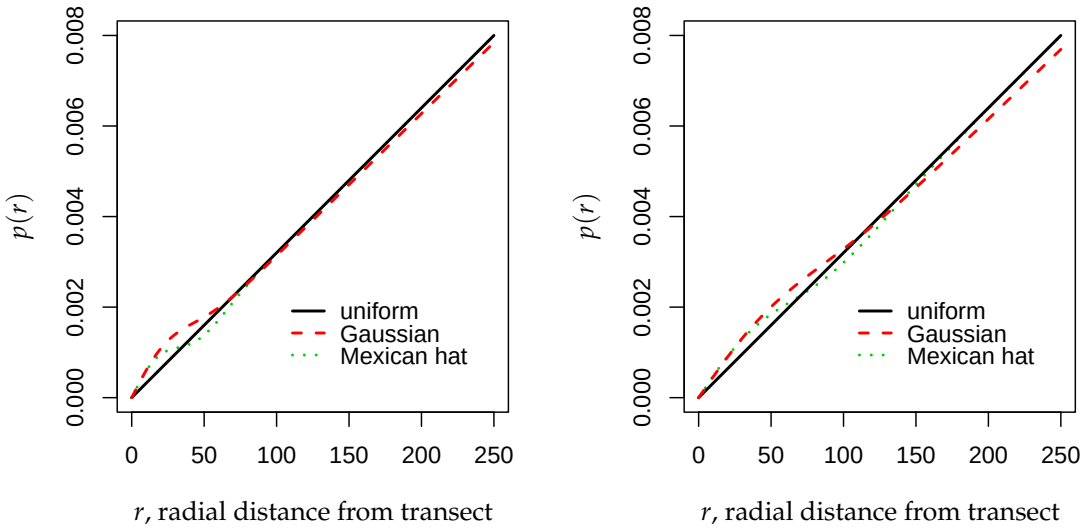

Figure 4.4: Comparison of the uniform, Gaussian and Mexican hat animal density functions within distance $W = 250$ m of a point transect for strong attraction over a short distance ($\alpha = 1/\phi(0)$, $\sigma = 25$) and moderate attraction over a long distance from the transect ($\alpha = 1/2\phi(0)$, $\sigma = 50$).

As for the case of line transects, $\alpha > 0$ leads to a peak in density and $\alpha < 0$ to a trough in density around the observer, while $\alpha = 0$ gives $\wp_g(r) = \wp_u(r)$. For $\wp_g(r) > 0$ we again require $\alpha > -1/\phi(0)$.

More realistic for point transects, where movement towards or away from the observer might be expected prior to detection, is a pdf for density based on the 2-dimensional form of the Mexican hat function. This is usually written as

$$m(x,y) = \left(2 - \frac{x^2+y^2}{\sigma^2}\right) e^{-(x^2+y^2)/2\sigma^2}.$$

Using $r^2 = x^2 + y^2$ to write this in terms of r, the function becomes

$$\begin{aligned} m(r) &= \left(2 - \frac{r^2}{\sigma^2}\right) e^{-r^2/2\sigma^2} \\ &\propto \left(1 - \frac{r^2}{2\sigma^2}\right) \phi\left(\frac{r}{\sigma}\right). \end{aligned} \tag{4.7}$$

We propose using (4.7) to model the deviation from the uniform density due to observer effects for point transects. The Mexican hat model for animal density about a point transect is therefore

$$\wp_m(r) = \frac{2r\left[1 + \alpha\left(1 - \frac{r^2}{2\sigma^2}\right)\phi\left(\frac{r}{\sigma}\right)\right]}{W^2\left[1 + \alpha\phi\left(\frac{W}{\sigma}\right)\right]}, 0 \le r \le W. \tag{4.8}$$

The bounds for α that must be satisfied in order that $\wp_m(r) \ge 0$ over the range $[0, W]$ now become

$$-\frac{1}{\phi(0)} \le \alpha \le \frac{1}{\phi(2)}.$$

We can see from Figures 4.3 and 4.4 that unless σ is large relative to W, the deviations from the uniform model will not be great due to the small proportion of the population that is close to a point transect. One consequence of this is that precise estimation of the density parameters may be more difficult. However, this may not matter so much because when estimating population size or density, the effect of allowing for transect or observer effects on local density will be less marked for point transects than line transects. Looking at it another way, ignoring observer effects by assuming uniformity may be less important when estimating abundance from point transect data than data from line transects.

4.3 Modeling Detection Probability

The key to being able to separate the detection and density components of the model is the additional information we have on detection through the capture-recapture data. Unlike standard distance sampling models, we are not restricted to assuming perfect detection on the transect, nor that detection probability decreases monotonically with distance. Instead, we are able to entertain a much wider class of models for modeling the relationship between detectability and distance. In our work we consider simple logistic-linear models relating an animal's detection probability to its distance from the observer. We assume that the distance of each animal is recorded once, and that the local population is closed to immigration and emigration, births and deaths during the observation period.

Let p_{jk} denote the detection probability of the kth observed animal on capture occasion j, $j = 1, \ldots, T$. The simplest model we consider assumes p_{jk} is a linear function of distance alone, i.e.,

$$\text{logit}(p_{jk}) = a_0 + a_1 r_k$$

where here we use r_k to denote either the perpendicular distance of animal k from the observer for line transects, or the radial distance for point transects. For this model we require $a_1 < 0$ so that detection probability decreases linearly with increasing distance on the logit scale.

When the observer has an effect on detection, either making animals easier or more difficult to detect by his or her presence, the linear model just described may be too simplistic. Instead, a quadratic model can incorporate features of the relationship we would expect when the observer affects detectability. When animals are more difficult to detect near the observer, we may expect the detection probability to initially increase with increasing distance before decreasing as before. When the observer causes nearby animals to be more easily detected, there may be a sharper peak at $r = 0$ in the relationship between detection probability and distance than we might expect from a linear model. This is illustrated with some examples in Figure 4.5. The model is given by

$$\text{logit}(p_{jk}) = a_0 + a_1 r_k + a_2 r_k^2$$

where this time we require $a_2 < 0$.

Another attractive feature of our models is that they can easily incorporate other covariates recorded on the individual animals (e.g., sex, age) or characteristics of the transect location when

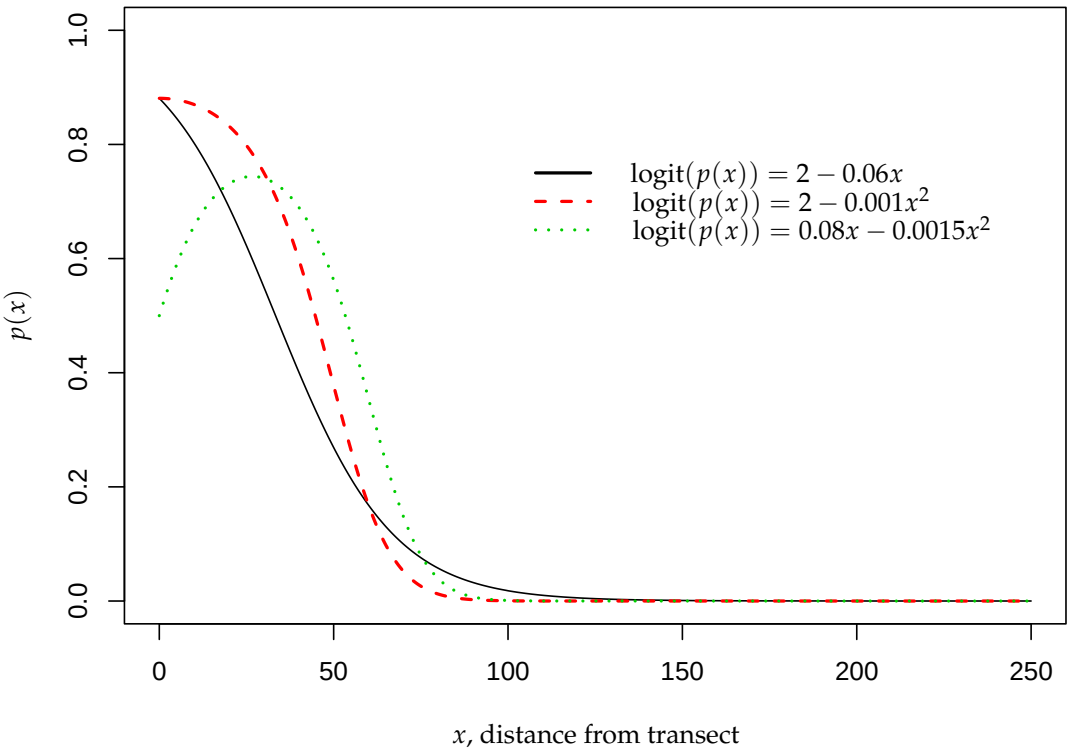

Figure 4.5: Examples of linear-logistic models where detection probability is a function of distance from the transect.

data come from multiple transects (e.g., vegetation, altitude, noise). We can also allow p_{kj} to vary across the T capture occasions, which we do in the example in Sections 4.5 and 4.6 when the $T = 3$ survey occasions are of different duration. However, such a limited class of detection functions may not be appropriate in all situations. Further, unlike our density models, the effect of transects or observers on detection probability is not directly quantified by the model parameters, and interpretation is more difficult. We discuss this further in Section 4.7.

4.4 Maximum Likelihood Estimation

For each distinct animal observed during the study, our data consist of a "capture" history for that animal (actually a detection history, as animals are not physically captured in a survey), and a measure of the distance of each animal from the transect. We assume that observations are recorded

on T sampling occasions, and that the population is closed over the sampling period. Let the true population size within distance W of the transects be denoted by N and let p_{jk} be the detection probability of the kth individual observed on occasion j, $j = 1, \ldots, T$ and $k = 1, \ldots, M$, where M is the number of unique animals detected during the survey. The detection data are written as z_{jk}, where $z_{jk} = 1$ if animal k is detected on occasion j, and 0 otherwise. Most generally we can think of these as coming from a Bernoulli distribution,

$$z_{jk}|p_{jk} \sim \text{Bernoulli}(p_{jk}). \tag{4.9}$$

Together with the assumption of independence of detections, we can use Equation (4.9) to construct the likelihood for our models. Note that animals for which $z_{jk} = 0$ for all j are not observed during the course of the survey.

We make use of the full multinomial likelihood constructed by Borchers et al. (1998) for capture-recapture data when detection probability is a function of covariates. Consider the general case of a single covariate, x (for example, distance), which (following Borchers et al., 1998) we assume has a distribution with density $\wp(x|\boldsymbol{\theta}_x)$ dependent on parameters $\boldsymbol{\theta}_x$. For animal k, x has measurement x_k, and we write the capture probability as a parametric function of x, $p_{jk} = p_j(x_k|\boldsymbol{\theta}_p)$ (in our case, this has logistic-linear form as in Section 4.3). Now denote the probability that an animal is seen at least once by $g(x|\boldsymbol{\theta}_p)$. It follows from the original Bernoulli model, (4.9), that

$$g(x|\boldsymbol{\theta}_p) = 1 - \prod_{j=1}^{T} \left\{1 - p_j(x|\boldsymbol{\theta}_p)\right\}.$$

For our models, this corresponds to the detection function in the distance sampling literature, except this function no longer requires the restriction that $g(0) = 1$. From Borchers et al. (1998), the full likelihood can now be written as

$$\begin{aligned} \mathcal{L}(\mathbf{Z}, \boldsymbol{x}|\boldsymbol{\theta}_x, \boldsymbol{\theta}_p) &= \frac{N!}{M!(N-M)!}(1-p_\cdot)^{N-M} \\ &\quad \times \prod_{k=1}^{M} \left\{ \prod_{j=1}^{T} p_{jk}^{z_{jk}} (1-p_{jk})^{1-z_{jk}} \right\} \wp(x_k), \end{aligned}$$

where for clarity we have suppressed the conditioning on the parameters on the right hand side. Here $p_\cdot$ is the expected probability that an animal is observed at least once during the survey, and

is given by

$$\begin{aligned} p_{\cdot} &= \int P(\text{animal detected}|x)\wp(x)dx \\ &= \int g(x)\wp(x)dx. \end{aligned}$$

This integral must be evaluated numerically, which has the potential to cause computational difficulties in maximizing the likelihood. However, for our examples and simulations, we have found few problems in evaluating the integral provided that the number of observed animals is sufficiently large and that good starting values for the numerical optimization algorithm are chosen when maximizing the likelihood. Pollock (2002) points out that there may also be problems with identifiability and estimation. Whether this is the case or not is of particular interest to us. The information on animal density is relatively poor, as it essentially comes from the histogram of observed distances (see Figure 4.6 below for examples), which reflects the combined density and detection processes. Therefore, in order to estimate the parameters of $\wp(x|\boldsymbol{\theta}_x)$ and distinguish effects of density from those of detection, we require good information from the detection component of the model from the capture history data. We examine this in the simulation study that follows.

4.5 Simulation Study

For both line and point transects we generated a population of N distances from either a Gaussian or Mexican hat density model with $W = 250$ m (the choice of W is arbitrary, but this matches the value used in Section 4.6). In the simulation study, animals were detected over $T = 3$ survey period according to a linear function shown in Figure 4.5, with $a_0 = 2$ and $a_1 = -0.06$. Reflecting the design of the particular bird surveys used in the following section, we also allowed the duration of each survey period to vary, with 3 minutes for the first, 2 for the second and 5 for the final period by including the logarithm of the duration as an offset term. The full logistic model was

$$\text{logit}(p_{jk}) = \log(t_j) + a_0 + a_1 r_k,$$

where $t_j = 3$, 2, or 5 minutes. We also did some simulations based on a quadratic model, but the results were quite similar in terms of bias and precision, and we only briefly discuss these below. The density models had parameter values of $\alpha = -1/\phi(0)$, $-1/2\phi(0)$ and $1/2\phi(0)$ and $\sigma = 25$ or

50, giving a range of observer effects on local density. For each parameter, we computed estimates of the bias and standard deviation of the estimators. We also computed Akaike's Information Criterion (AIC, Akaike, 1973) for each model, and from these we calculated the proportion of times the correct density model was chosen as a measure of our ability to distinguish the uniform, Gaussian and Mexican hat models in different circumstances. For each combination of underlying models and parameter values, we generated 200 simulated data sets.

We found it necessary to impose some restrictions on our simulations. From preliminary simulations, we have found that the population N within distance W of the line or point transect must be large enough to yield in the order of 1000 detected individuals for consistent convergence of maximization algorithms to occur when maximizing the likelihood. Particularly for point transects, the likelihood surface is quite flat and sometimes algorithms will apparently converge to a local maximum or not converge at all, depending on the choice of starting values. In fitting models to real data, we can try different starting values and examine in detail the characteristics of the likelihood surface in order to ensure that it is indeed maximized. In a simulation study it is impractical to examine each maximization in such detail, and we therefore choose a large N for our simulations, 7500 for line transects and 15000 for point transects. This does not necessarily imply our methods are limited to such large local populations, but rather that when animal densities are relatively low, we must take more care in fitting our models. We also restricted the range of α, the attraction/avoidance parameter to be negative when the true value was negative, and positive when the true value was also positive. Our preliminary simulations showed some tendency for apparent convergence to a local maximum with $\widehat{\alpha}$ with the incorrect sign. This is not a particularly limiting restriction as in many instances it is known in advance whether a species avoids or is attracted to observers or transects. Again, even if there is uncertainly in practice, we are able to take more time in examining the fitted models when applied to real rather than simulated data.

4.5.1 Line Transects

Results of the line transect simulations are given in Tables 4.1, 4.2 and 4.3. Table 4.1 gives estimates of absolute bias and standard deviations for the estimators of the density and detection parameters under both Gaussian and Mexican hat density models. For most combinations of α and σ, bias in all estimators is very low and the standard deviation is low relative to the true value of the parameter. Performance of the Gaussian estimation is somewhat poorer than that of the Mexican hat. Bias and standard deviation are consistently higher, particularly when $\alpha = 1/2\phi(0)$ for both $\sigma = 25$ and 50

and also $\alpha = -1/2\phi(0)$ for $\sigma = 25$. The reason for the poorer performance of the Gaussian model in general is that when this model holds for given parameter values, a smaller proportion of the population is closer to the observer than is the case with the Mexican hat, as animals are displaced some distance under the latter model rather than removed entirely. Therefore, fewer birds end up being detected under a Gaussian model, and there are less data available for estimation. We shall see that this effect is even more striking for point transect data.

As the goal of a bird survey is estimation of abundance or animal density, the effect of ignoring observer effects and assuming uniformity of density is of greatest interest. Table 4.2 shows the relative bias in the estimator of N for our models when the underlying true model is either Gaussian or Mexican hat. If the Gaussian model holds true, then for our simulations, relative biases as large as –0.64 occur when we mistakenly assume a uniform model. N is underestimated when α is negative (avoidance) and overestimated when α is positive (attraction). Biases of up to magnitude 0.36 are estimated when the uniform model is fitted to distance data which have a Mexican hat distribution, smaller than for the Gaussian due to many displaced animals still being within detectable distances under this model. Also, the standard deviation of the estimators under the uniform model is consistently much lower than that for the true model: not only can misspecification of the density model lead to large biases in estimation of abundance and density, but very misleading estimates of the precision of parameter estimates will also be obtained.

Table 4.3 shows the proportion of times each of the three alternative density models were selected as "best" by having the lowest AIC value. The AIC easily distinguishes the uniform as a poorer fit for these simulations, correctly failing to select this model almost all of the time. It is more difficult to distinguish the fits of the Gaussian and Mexican hat models. When $\sigma = 25$, the Gaussian model is correctly chosen a majority of the time (from 0.62 to 0.82) but when $\sigma = 50$, the AIC selects the Gaussian and Mexican hat models almost equally as often. When the Mexican hat is the true density, it is selected 0.88-0.98 of the time for $\sigma = 25$ and 0.7-0.8 of the time when $\sigma = 50$. The changing AIC performance with σ and, to a lesser degree, α, is likely to be as much a function of sample size as the values of these parameters. As σ increases, a smaller proportion of the population has high detection probabilities from being close to the transect, particularly for the Gaussian model. We verified this with some supplementary simulations with different values of N (not presented here).

Table 4.1: Estimates of absolute bias and standard deviation (sd) for the detection and density parameters from 200 simulations of data from either Gaussian ($\wp_g$) or Mexican hat ($\wp_m$) model of animal density in the vicinity of line transects. The detection model was linear ($a_0 = 2.0$, $a_1 = -0.06$), $W = 250$ m and true $N = 7500$.

true	$\sigma =$	25						50					
density	$\alpha =$	$-1/\phi(0)$		$-1/2\phi(0)$		$1/2\phi(0)$		$-1/\phi(0)$		$-1/2\phi(0)$		$1/2\phi(0)$	
		bias	sd	bias	sd	bias	sd	bias	sd	bias	sd	bias	sd
	α	0.003	0.007	–0.014	0.11	0.12	0.33	0.002	0.004	–0.10	0.31	0.71	2.19
$\wp_g$	σ	0.2	1.4	1.4	5.6	0.4	6.1	3.2	9.5	9.6	26.3	–0.1	20.7
	a_0	0.00	0.09	0.00	0.07	0.01	0.06	0.03	0.16	0.01	0.10	0.00	0.07
	a_1	0.000	0.001	0.000	0.001	0.000	0.001	–0.001	0.003	0.000	0.002	0.000	0.002
	α	0.005	0.004	0.004	0.092	0.024	0.13	0.004	0.008	–0.018	0.11	0.035	0.22
$\wp_m$	σ	0.1	0.8	0.0	2.4	0.2	1.5	0.5	3.1	1.6	7.1	–0.5	5.4
	a_0	–0.01	0.07	–0.00	0.07	0.00	0.06	0.00	0.11	0.01	0.08	0.00	0.07
	a_1	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.002	0.000	0.002	0.000	0.020

Table 4.2: Comparisons of estimates of relative bias and standard deviation (sd) of $\widehat{N}$ when fitting the uniform, Gaussian and Mexican hat densities to 200 simulated data sets from either the Gaussian ($\wp_g$) or Mexican hat ($\wp_m$) model of animal density in the vicinity of line transects. The detection model was linear ($a_0 = 2.0$, $a_1 = -0.06$), $W = 250$ m and true $N = 7500$.

true	$\sigma =$	25						50					
density	$\alpha =$	$-1/\phi(0)$		$-1/2\phi(0)$		$1/2\phi(0)$		$-1/\phi(0)$		$-1/2\phi(0)$		$1/2\phi(0)$	
		bias	sd	bias	sd	bias	sd	bias	sd	bias	sd	bias	sd
	$\widehat{\wp}_u$	–0.37	110	–0.18	130	0.17	170	–0.64	80	–0.29	130	0.23	170
$\wp_g$	$\widehat{\wp}_g$	0.01	320	0.02	490	–0.01	390	0.08	1750	0.08	1730	–0.01	870
	$\widehat{\wp}_m$	–0.23	230	–0.08	510	0.07	410	–0.08	2110	0.04	1940	0.03	940
	$\widehat{\wp}_u$	–0.05	130	–0.03	140	0.03	160	–0.36	110	–0.19	130	0.21	180
$\wp_m$	$\widehat{\wp}_g$	0.25	250	0.12	210	–0.12	190	0.34	760	0.17	690	–0.13	450
	$\widehat{\wp}_m$	0.00	160	0.00	160	0.00	170	0.01	610	0.02	630	0.00	440

Table 4.3: Proportions of times the uniform, Gaussian, and Mexican hat density models were selected using AIC for 200 simulations of data from either the Gaussian ($\wp_g$) or Mexican hat ($\wp_m$) model of animal density in the vicinity of line transects. The detection model was linear ($a_0 = 2.0$, $a_1 = -0.06$), $W = 250$ m and true $N = 7500$.

true	$\sigma =$	25			50		
density	$\alpha =$	$-1/\phi(0)$	$-1/2\phi(0)$	$1/2\phi(0)$	$-1/\phi(0)$	$-1/2\phi(0)$	$1/2\phi(0)$
$\wp_g$	$\widehat{\wp}_u$	0.00	0.00	0.00	0.00	0.00	0.03
	$\widehat{\wp}_g$	0.82	0.615	0.62	0.565	0.425	0.46
	$\widehat{\wp}_m$	0.18	0.385	0.38	0.435	0.575	0.51
$\wp_m$	$\widehat{\wp}_u$	0.00	0.00	0.00	0.00	0.00	0.00
	$\widehat{\wp}_g$	0.02	0.12	0.085	0.20	0.29	0.305
	$\widehat{\wp}_m$	0.98	0.88	0.915	0.80	0.71	0.695

4.5.2 Point Transects

With point transects, a smaller proportion of animals are within the region of high detection probability surrounding the observer. Because the 'interesting' part of the Gaussian and Mexican hat densities are close to the observer, we can expect estimation and model selection to be poorer than for line transects. This is also why we needed to choose a larger N for the point transect simulations – a smaller choice of N leads to many convergence problems in estimation. Even with this larger population size within W of the observer, biases and standard deviations for the density parameters, α and σ (Table 4.4) are generally greater than for line transects. This is not true for the detection parameters α_0 and α_1 but these are morey a function of the number of detected animals, and thus the number of distance observations, and we have chosen N to yield similar sample sizes for both line and point transects. As with line transects, biases and standard deviations are greatest when $\sigma = 50$ and $\alpha > -1/\phi(0)$ and $\alpha = 1/2\phi(0)$ with $\sigma = 25$. When $\alpha > 0$ there is a much larger range of possible estimates, so we may at least expect more variation in this case.

When $\sigma = 25$ and, therefore, the effect of the transect or observer extends a relatively moderate distance into the habitat, the bias in the estimator of N for the uniform model (Table 4.5) is much smaller than for the corresponding line transect simulations. We anticipated this due to the fact that a smaller proportion of the population is close to a point transect than a line transect. The relative bias estimates in the uniform for $\sigma = 25$ range in absolute value from 0.02 to 0.23, compared to 0.03 to 0.37 for line transects (Table 4.2). When σ increases to 50, large biases appear in the uniform estimators of abundance (0.19-0.56), and the values are now similar to the value for line transects.

Table 4.4: Estimates of absolute bias and standard deviation (sd) for the detection and density parameters from 200 simulations of data from either Gaussian ($\wp_g$) or Mexican hat ($\wp_m$) model of animal density in the vicinity of point transects. The detection model was linear ($a_0 = 2.0$, $a_1 = -0.06$), $W = 250$ m and true $N = 15000$.

	$\sigma =$	25						50					
	$\alpha =$	$-1/\phi(0)$		$-1/2\phi(0)$		$1/2\phi(0)$		$-1/\phi(0)$		$-1/2\phi(0)$		$1/2\phi(0)$	
		bias	sd	bias	sd	bias	sd	bias	sd	bias	sd	bias	sd
$\wp_g$	α	0.032	0.056	–0.09	0.27	0.22	0.61	0.010	0.022	–0.11	0.29	0.64	1.31
	σ	0.7	2.3	0.9	9.1	1.3	9.3	4.2	10.6	5.6	26.2	1.5	19.9
	a_0	0.01	0.10	0.01	0.10	–0.01	0.09	0.02	0.18	0.02	0.15	–0.02	0.10
	a_1	0.000	0.001	0.000	0.002	0.000	0.002	–0.001	0.003	0.000	0.003	0.000	0.002
$\wp_m$	α	0.042	0.081	–0.08	0.32	0.11	0.38	0.015	0.030	–0.09	0.22	0.22	0.52
	σ	0.3	1.5	–0.1	5.5	0.8	6.8	1.3	5.3	4.7	20.9	1.1	11.0
	a_0	0.00	0.08	0.00	0.08	0.00	0.08	0.01	0.15	0.02	0.12	–0.01	0.10
	a_1	0.000	0.001	0.000	0.001	0.00	0.001	0.000	0.002	0.000	0.002	0.000	0.002

Table 4.5: Comparisons of estimates of relative bias and standard deviation (sd) of $\widehat{N}$ when fitting the uniform, Gaussian and Mexican hat densities to 200 simulated data sets from either the Gaussian ($\wp_g$) or Mexican hat ($\wp_m$) model of animal density in the vicinity of point transects. The detection model was linear ($a_0 = 2.0$, $a_1 = -0.06$), $W = 250$ m and true $N = 15000$.

true	$\sigma =$	25						50					
density	$\alpha =$	$-1/\phi(0)$		$-1/2\phi(0)$		$1/2\phi(0)$		$-1/\phi(0)$		$-1/2\phi(0)$		$1/2\phi(0)$	
		bias	sd	bias	sd	bias	sd	bias	sd	bias	sd	bias	sd
$\wp_g$	$\widehat{\wp}_u$	–0.23	360	–0.12	130	0.12	470	–0.56	270	–0.28	340	0.27	510
	$\widehat{\wp}_g$	0.01	1020	0.02	1510	–0.02	1270	0.12	4200	0.09	4370	–0.04	2250
	$\widehat{\wp}_m$	–0.11	850	–0.04	1820	0.04	1290	–0.01	4620	0.04	4480	–0.01	2320
$\wp_m$	$\widehat{\wp}_u$	–0.03	390	0.02	410	0.02	440	–0.34	310	–0.18	380	0.19	490
	$\widehat{\wp}_g$	0.11	560	0.05	640	–0.05	650	0.22	2270	0.16	3890	–0.09	1630
	$\widehat{\wp}_m$	0.01	450	0.01	520	0.00	590	0.04	1940	0.08	3850	–0.03	1640

Table 4.6: Proportions of times the uniform, Gaussian, and Mexican hat density models were selected using AIC for 200 simulations of data from either the Gaussian ($\wp_g$) or Mexican hat ($\wp_m$) model of animal density in the vicinity of point transects. The detection model was linear ($a_0 = 2.0$, $a_1 = -0.06$), $W = 250$ m and true $N = 15000$.

true	$\sigma =$	25			50		
density	$\alpha =$	$-1/\phi(0)$	$-1/2\phi(0)$	$1/2\phi(0)$	$-1/\phi(0)$	$-1/2\phi(0)$	$1/2\phi(0)$
$\wp_g$	$\widehat{\wp}_u$	0.00	0.005	0.015	0.00	0.00	0.07
	$\widehat{\wp}_g$	0.60	0.54	0.480	0.48	0.315	0.37
	$\widehat{\wp}_m$	0.40	0.45	0.505	0.52	0.685	0.56
$\wp_m$	$\widehat{\wp}_u$	0.00	0.005	0.005	0.00	0.00	0.00
	$\widehat{\wp}_g$	0.21	0.335	0.30	0.28	0.435	0.385
	$\widehat{\wp}_m$	0.79	0.66	0.695	0.72	0.565	0.615

Biases in N for the correct model (Gaussian or Mexican hat) are again low when $\sigma = 25$, but are somewhat larger for point transects when $\sigma = 50$, up to 0.09 for the Gaussian and 0.08 for the Mexican hat model.

As expected, selection of the correct model using AIC is more difficult for point transect data (Table 4.6) than line transect data (Table 4.3). When $\sigma = 25$, the Mexican hat model is selected almost as often as the Gaussian when the latter model is in fact correct; when $\sigma = 50$ the Mexican hat is even more likely to be selected than the correct Gaussian model. This may not be of concern, as our simulations show that bias in N is actually lower for the incorrect Mexican hat model in this case (Table 4.5). The Mexican hat model fares better, being selected a majority of the time for all parameter values. Perhaps most importantly, the uniform model is again almost never selected for point transect data.

We also undertook some simulations using a quadratic model to examine if decreased detection near the observer could be distinguished from observer effects on local density. With detected animals of around 1000 or more, we found that the parameters of the detection model were again estimated precisely and with relatively little bias when the correct model was being fitted. This led to estimates of the density parameters that are similar in terms of bias and precision to those for the linear case discussed above.

Table 4.7: Parameter estimates and parametric bootstrap standard errors from fitting models to the black throated green warbler data, assuming avoidance of observers. $W = 250$m and $\widehat{d}$ is the estimated bird density in individuals/km^2.

	Linear detection						Quadratic detection					
	Uniform		Gaussian		Mexican hat		Uniform		Gaussian		Mexican hat	
	value	sd	value	sd	value	sd	value	sd	value	sd	value	sd
$\widehat{d}$	40.7	0.7	41.7	1.12	40.8	0.7	38.3	0.6	39.0	0.7	38.3	0.7
$\widehat{\alpha}$	–		–2.49	0.25	–2.37	0.24	–		–2.49	0.17	–2.39	0.19
$\widehat{\sigma}$	–		7.9	1.8	17.1	1.6	–		7.5	1.5	16.8	1.6
$\widehat{a}_0$	2.21	0.08	2.22	0.10	2.21	0.08	0.25	0.16	0.28	0.15	0.26	0.19
$\widehat{a}_1 \times 10^2$	–5.77	0.10	–5.81	0.13	–5.77	0.10	0.40	0.43	0.31	0.45	0.36	0.56
$\widehat{a}_2 \times 10^4$	–		–		–		–3.95	0.29	–3.92	0.32	–3.93	0.37
AIC	1076.8		1061.0		1024.9		898.7		894.4		855.8	

4.6 Application to Bird Survey Data

We applied our models to capture-recapture survey data from a large survey of bird populations in the Great Smoky Mountains National Park from 1996 to 1999. Over 3000 point transects were surveyed in the study, although we restrict ourselves to data from the first year in which a full survey was undertaken, 1997. At each point the survey lasted 10 minutes during which time an observer would attempt to record all the birds that were seen or heard. The observation period was divided into three smaller intervals, the first of 3 minutes, the second of 2 minutes, and the third of 5 minutes duration, and the observer attempted to keep track of all birds throughout the three intervals. For example, if a bird was seen or heard in the first interval, the observer would then try to detect the same bird in the succeeding two intervals. The intervals in which the bird was observed were recorded, leading to a three-period capture history for each bird, e.g., 101 if the bird was detected in the first and third intervals, but not the second. Along with the capture histories, the distance of each bird from the observer was estimated at first detection. The data we analyze here have been aggregated over all point transects, although we discuss more general models which allow for between transect spatial variation in Section 4.7.

Table 4.7 gives the parameter estimates for the Black-Throated Green Warbler, one of the more abundant species detected in the survey. It is uncertain whether this species avoids or is attracted to observers, and we therefore fitted two sets of models, first ensuring that $\alpha < 0$ when maximizing

the likelihood, and again with $\alpha > 0$. The parameter estimates for the density model in the latter case were quite implausible, implying that birds were attracted to the observer in large numbers from over 100 m away. Table 4.7, therefore, gives the more realistic results we obtained from an avoidance model. Standard deviations were estimated using a parametric bootstrap: 50 data sets were generated under each of the fitted models and the standard deviations of the resulting parameter estimates were computed. Although the results show that the Mexican hat model with a quadratic detection function was the best fitting model according to its AIC, the difference between the 6 models in terms of the estimate of density is very small. Based on the results of our simulations, the effect of incorrectly assuming a uniform density on estimation of animal density should not be great with $\widehat{\sigma} \approx 17$ for the Mexican hat model. The results here support this, with density estimates for the uniform being virtually identical to the Mexican hat for both linear and quadratic models. Nevertheless, the values of the density model parameters are themselves interesting. Taking the Mexican hat model density and quadratic detection model, we can infer that observers have a strong effect on local density ($\widehat{\alpha} = -2.39$, sd=0.19), and with $\widehat{\sigma} = 16.8$ (sd=1.6) birds within approximately 25 metres are displaced to further distances from the observer (based on Figure 4.3 with $\sigma = 25$).

Although the Mexican hat/quadratic model appears to be the best of those we considered, in fact none of the models fit the observed distance data very well. Figure 4.6 compares the histogram of observed distance with the estimated probability density function of observed distances for both linear and quadratic detection models. We can see why the Mexican hat and Gaussian models provide a better fit than the uniform model, as fewer birds were observed at very short distances than expected under the uniform model. From 25 to 75 m, however, many more birds were observed than predicted under all three density models for both linear and quadratic detection, while fewer than expected birds were detected at greater distances. Either a quadratic detection model is too inflexible to model the relationship between detection probability and distance, or for these data, the density models do not have the correct shape. In fact, in terms of the AIC, the Gaussian and Mexican hat attraction models for density provided better fits than the avoidance model, but as noted above, the parameters estimates were not sensible. We analyzed data from a number of other bird species from the same point transect survey and the results, including examination of model fit, were very similar. Part of the problem here is almost certainly due to the small amount of information on local density contained in the distance data. Essentially, that information comes

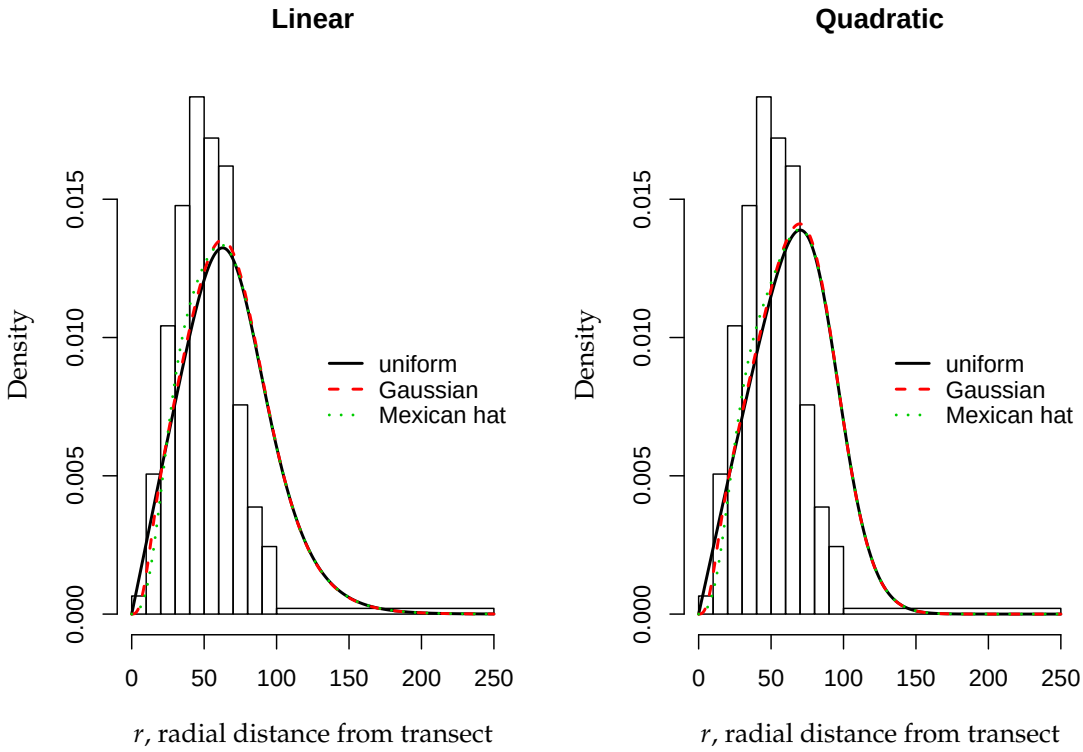

Figure 4.6: Comparison of histogram of observed distances with estimated density under linear and quadratic detection models fitted to Black Throated Green Warbler survey data.

from the left hand side of the histograms in Figure 4.6, where few animals are detected. Complicating matters in this case is that there are quite possibly serious errors in both the recording of detections of individual birds, and in the estimation of distances. Birds are generally recorded by sound alone, making it difficult to distinguish individuals and to record accurate measurements of distances from the observer. The effect of measurement errors is unclear and requires further study.

4.7 Discussion

Distance sampling surveys typically assume that the density of animals about a line or a point transect is uniform. We have presented models which can be used to allow for types of non-uniformity due to attraction towards or avoidance of a transect or observer for both line and point transects when we have both distances and capture histories for observed animals. For line transects, our

simulation study shows that assuming uniformity when there is avoidance or attraction can lead to large biases in estimates of abundance or density, and an overly optimistic assessment of the precision of such estimates. The consequences of incorrectly assuming uniformity are less serious for point transect estimation unless transect and observer effects are very large, due to fewer animals being seen at the short distances over which transects may affect bird behavior.

In our simulation study and our application, we examined estimation of abundance or mean density within a distance W of the transect. When using the Gaussian model, however, we must be more careful when attempting to extrapolate a density estimate to the entire region of interest. Under the uniform and Mexican hat models, birds are not generally displaced beyond W of the transect, and so estimates of mean density can be reasonably applied to the entire survey region, provided of course that there is no other source of bias due to the location of the transects. But in the Gaussian case, when birds are permanently displaced around the transect, the fact that most of the region within which a survey is undertaken will not be within W of a transect will mean that some adjustment must be made. One straightforward approach is to examine the estimated Gaussian density function and estimate the proportion of the area within W for which this function approximates a uniform density (i.e., is horizontal in the case of line transects). From this and our original estimate of abundance within distance W of the transects, we can estimate the mean density for the region for which animal density is unaffected by the transects.

Our density models are appealing in their simplicity and in the ease with which the parameter values can be interpreted. They do require a fairly large amount of data for estimation, but given this, the estimators should be relatively unbiased and the effects of transects on density should be distinguishable from changes in detection probability with distance. In numerically maximizing the likelihood, care must be taken to ensure that the algorithms have correctly converged to the maximum, and we recommend fitting models with various starting values to check this. Our data analysis in Section 4.6 also highlights some of the estimation problems that can arise when information on density near the observer is poor, as is the case with point transects. Errors in both animal detections and in distance measurement will also complicate the fitting and interpretation of our models.

A drawback of our density models is that they have very particular shapes which may not be suitable for all species of animals. Future work should consider alternative models, but preferably ones which do not unduly sacrifice interpretability for complexity. We have focused less attention on our choice of detection model, but we note that it should be possible to develop models which

allow for observer effects more explicitly than polynomial models are able to do. For example, we could consider a parametric model for detection with parameters that measure the strength and range of observer effects in a similar manner to our density models. An interesting observation from our simulation studies, particularly for quadratic models, is that when the fitted model is correct, both density and detection parameters are generally estimated with little bias. The implication of this is that with sufficient data, it is possible to distinguish transect effects on density from those on detection probability. It would be useful to confirm this by examining a wider range of detection models.

Our analyses have been based on aggregating data from multiple line or point transects. Conceptually, it is straightforward to extend the models to allow for spatial variation in animal density. This approach would follow the same route as that taken by various authors in modeling spatially replicated repeated count or removal data (e.g., Royle, 2004; Webster, Pollock and Ghosh, 2006; Webster, Pollock and Simons, 2006), for example, with a Poisson-lognormal mixture model for the values of N at each transect, incorporating both covariate and spatial effects. Such a complex model would inevitably require Bayesian hierarchical modeling using Markov chain Monte Carlo algorithms. This approach is worthy of investigation, as it would lead directly to maps of animal density. However, with a complex form for the likelihood, and numerical integrations at each step of the algorithm, fitting such models in practice may be difficult.

Acknowledgements

We thank Dr Ted Simons of North Carolina State University for supplying the bird survey data.

Bibliography

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, *in* B. N. Petrov and F. Csaki (eds), *Proceedings of the Second International Symposium on Information Theory*, pp. 267–281.

Alho, J. M. (1990). Logistic regression in capture-recapture models, *Biometrics* **46**: 639–649.

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

Buckland, S. T., Anderson, D. R., Burham, K. P., Laake, J. L., Borchers, D. L. and Thomas, L. (2001). *Introduction to Distance Sampling – Estimating Abundance of Biological Populations*, Oxford: Oxford University Press.

Buckland, S. T. and Turnock, B. J. (1992). A robust line transect method, *Biometrics* **48**: 901–909.

Huggins, R. M. (1989). On the statistical analysis of capture experiments, *Biometrika* **76**: 133–140.

Pollock, K. H. (2002). The use of auxiliary variables in capture-recapture modelling: an overview, *Journal of Applied Statistics* **29**: 85–102.

Pollock, K. H., Nichols, J. D., Brownie, C. and Hines, J. E. (1990). Statistical inference for capture-recapture experiments, *Wildlife Monographs* **107**: 1–97.

Royle, J. A. (2004). N-mixture models for estimating population size from spatially replicated counts, *Biometrics* **60**: 108–115.

Turnock, B. J. and Quinn, T. J. (1991). The effect of responsive movement on abundance estimation using line transect sampling, *Biometrics* **47**: 701–715.

Webster, R. A., Pollock, K. H. and Ghosh, S. K. (2006). Bayesian spatial modeling of data from unit-count surveys of fish in streams, *(in preparation)* .

Webster, R. A., Pollock, K. H. and Simons, T. R. (2006). Bayesian spatial modeling of point transect data from bird surveys, *(in preparation)* .