

8-29-2019

On the Topic of Spatial Capture-Recapture Modeling

Paul McLaughlin

University of Connecticut - Storrs, paul.mclaughlin@uconn.edu

Follow this and additional works at: <https://opencommons.uconn.edu/dissertations>

Recommended Citation

McLaughlin, Paul, "On the Topic of Spatial Capture-Recapture Modeling" (2019). *Doctoral Dissertations*. 2317.
<https://opencommons.uconn.edu/dissertations/2317>

On the Topic of Spatial Capture-Recapture Modeling

Paul Samuel McLaughlin, Ph.D.
University of Connecticut, 2019

ABSTRACT

Over the past two decades there have been many advancements in modeling capture-recapture (CR) data to account for emerging data collection technology and techniques. Spatial capture-recapture (SCR) models have been introduced to estimate population size and numerous other demographic parameters from spatially explicit CR data. Here we offer a comprehensive review of the development of CR modeling up to and including SCR models.

We then introduce a new SCR model which allows for attractions between individuals via their daily movements. A simulation study is used to demonstrate accounting for these attractions can improve population size estimation. Additionally, we apply our model to an iconic SCR dataset to estimate the population size and attraction parameters of a Bengal tiger (*Panthera tigris tigris*) population.

To conclude we present a reversible-jump Markov chain Monte Carlo (RJMCMC) approach for parameter estimation which has not previously been extended to SCR models. Simulation studies are presented to show the superior computational efficiency

Paul Samuel McLaughlin, Ph.D.
University of Connecticut, 2019

of this proposed approach. We also demonstrate the application of this RJMCMC method to SCR data by estimating the size of an American black bear (*Ursus americanus*) population.

On the Topic of Spatial Capture-Recapture Modeling

Paul Samuel McLaughlin

B.A., Bard College, NY, USA, 2010

M.A., University of Connecticut, CT, USA, 2017

A Dissertation
Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy
at the
University of Connecticut

2019

Copyright by

Paul Samuel McLaughlin

2019

APPROVAL PAGE

Doctor of Philosophy Dissertation

On the Topic of Spatial Capture-Recapture Modeling

Presented by

Paul Samuel McLaughlin, BA, MA

Major Advisor

Dr. Haim Bar

Associate Advisor

Dr. Kun Chen

Associate Advisor

Dr. Ming-Hui Chen

Associate Advisor

Dr. Dipak Dey

University of Connecticut

2019

Dedication

For Sadie.

Acknowledgements

I would like to acknowledge and thank my family, especially my parents and my dog Ruby, for their immeasurable amount of help over the past several years. We did it!

I would also like to thank the University of Connecticut and the statistics department as a whole. I have had wonderful teachers, mentors, peers and students here and it has truly been an honor to be a member of and serve this community.

There have been too many friends to list here who have encouraged me, helped me with statistics and simply kept things fun over the years but to name a few, in alphabetical order, thanks: Brian, Chun, Dooti, Ezra, Greg, Jerry, Qianzhu, Seojin, Shelley, Val, Ved and Yujing.

Thanks Anthony, Megan and Tracy for keeping everything running so smoothly, I could never have done this without your support.

Thank you to my board:

- Dr. Kun Chen: this research would have been impossible without the invaluable experience I gained programming in R while grading for Dr. Chen's course.
- Dr. Dipak Dey: taking Dr. Dey's applied seminar in Bayesian analysis taught me not only the concepts but the intuition behind Bayesian statistics.
- Dr. Ming-Hui Chen: simply put, Dr. Chen has taught me more about what it means to perform statistics than any other person, book or course. I'm extremely

grateful for the experience I gained applying statistics and working with clients under his guidance.

Lastly, I would like to thank and acknowledge the invaluable contribution of my advisor, Dr. Haim Bar, for allowing me to work on this topic and always having faith that we would be successful. It has been an honor and a privilege to learn how to perform research under his guidance and I feel excited and well prepared for my future career almost entirely as a result of his advice.

Contents

Dedication	iv
Acknowledgements	v
1 Review of Capture-Recapture Models for the Estimation of Abundance with an Emphasis on Spatial Capture-Recapture	1
1.1 Introduction	1
1.2 Seminal and Classical Models	3
1.2.1 Two sample experiments	4
1.2.2 Multiple samples with equal catchability	5
1.2.3 Closed population models	7
1.2.4 Open population models	19
1.3 Spatial Capture-Recapture	27
1.3.1 Spatial sampling	29
1.3.2 Single-catch trap experiments	34
1.3.3 Multi-catch trap experiments	37
1.3.4 Proximity detection experiments	40
1.3.5 Area search experiments	43
1.3.6 Individual movement	47

1.3.7	Discussion	49
2	A Spatial Capture-Recapture Model with Dependent Animal Movement	52
2.1	Introduction	52
2.2	A Spatial Capture-Recapture Model with Attractions	55
2.2.1	Model formulation	56
2.2.2	Estimation	57
2.3	Simulation Study	60
2.3.1	CA vs. IB model comparison	61
2.3.2	CA model performance when $\psi_\alpha = 1$	64
2.4	Analysis of Tiger Data	65
2.5	Discussion	70
3	Spatial Capture-Recapture via Reversible Jump Markov Chain Monte Carlo	72
3.1	Introduction	72
3.2	Model	75
3.2.1	Model formulation	76
3.2.2	Estimation	77
3.3	Simulation Study	78
3.3.1	Efficiency of RJMCMC vs. data augmentation	79

3.3.2	RJMCMC vs. data augmentation estimation	85
3.4	Black bear study example	87
3.5	Discussion	89
A	Classical Capture-Recapture Notation	91
A.1	Closed population notation	91
A.2	Open population notation	92
B	Code for Spatial Capture-Recapture Model with Dependent Animal Movement Presented in Chapter 2	94
B.1	σ and λ estimation	94
B.2	σ_π estimation	96
B.3	Unobserved w_i estimation	99
B.4	ψ and ψ_α estimation	101
B.5	Observed s_i estimation	103
B.6	α_{it} estimation	107
	Bibliography	110

List of Tables

1	Model comparison of \hat{N} while varying σ_π	64
2	Parameter estimates from tiger data for CA and IB models	70
3	Model comparison of \hat{N} for 200 simulated datasets	85
4	Parameter estimates from bear data for RJMCMC and DA models	88

List of Figures

1	Graphical representation of estimation model, ellipses represent parameters, rectangles with rounded corners represent functions of parameters	59
2	Side-by-side boxplots of the bias for the estimate of N by σ_π	62
3	Side-by-side boxplots of the MSE for the estimate of N by σ_π	62
4	Side-by-side boxplots of σ estimate bias by σ_π	63
5	Side-by-side boxplots of λ estimate bias by σ_π	63
6	Side-by-side boxplots of N estimate bias by model with $\psi_\alpha = 1$	65
7	Histogram of ψ_α estimate with $\psi_\alpha = 1$	66
8	Histogram of σ_π estimate with $\psi_\alpha = 1$	66
9	Posterior densities of σ from CA and IB models	68
10	Posterior densities of N from CA and IB models	68
11	Posterior density of σ_π from CA model	69
12	Posterior densities of ψ_α from CA model	69
13	Histograms of N 's estimated posterior distribution using data augmentation with increasing M	80
14	Histograms of N 's estimated posterior for two datasets with the same $M(= 110)$	81

15	Histograms of N 's estimated posterior for increasing M and Fisher-Pearson coefficient of skewness	82
16	Side-by-side boxplots of the difference between \hat{N} with $M = 400$ and \hat{N} with increasing values of M starting from $M^{(+skew)}$ for 200 simulated datasets	83
17	Side-by-side boxplots of RJMCMC runtimes vs. DA runtimes using best M (i.e. $M = M^{(+skew)} + 30$) and DA using $M = 400$	84
18	Side-by-side boxplots of RJMCMC vs. DA MSE's for the estimation of N	86
19	Side-by-side boxplots of RJMCMC vs. DA Biases for the estimation of N	86
20	Overlapping histograms of estimated posterior distributions of N from the RJMCMC and DA models	89

Chapter 1

Review of Capture-Recapture

Models for the Estimation of

Abundance with an Emphasis on

Spatial Capture-Recapture

1.1 Introduction

The analysis of capture-recapture (CR) data has had a long history, originating as far back as 1786 with the Lincoln-Petersen estimator, first introduced by Laplace to estimate the size of the population of France (as referenced by Seber [1982]). In recent years there has been a rapid increase in new methods and applications for modeling of CR data, in large part due to technological advances in both computing and data collection (McCrea and Morgan [2014]).

While many parameters of interest can be investigated through the use of CR data

(such as survival rates, animal dispersion, ect.) the focus of this current review is on the estimation of abundance or the size of a given population. In particular, this review will put emphases on spatial capture-recapture (SCR). This is done for two main reasons: (1) there has been an explosion of research concerning SCR over the past decade since it's introduction and (2) since SCR seems to be a natural progression of CR analysis given data containing any spatial information.

While CR data can involve virtually any type of subject (i.e. plants, animals, computer bugs, etc.) the general term 'individual' will be used throughout this review. Likewise, while there are numerous methods for sampling individuals (trapping, photographing, hair snaring, etc.) for the most part this review will use the term 'capture', occasionally resorting to 'observed' or 'detected' when the occasion dictates.

The main focus of this paper is on the review and development of likelihoods or hierarchical designs and estimation of their parameters. The paper is structured as follows: section 1.2 will focus on seminal and classical models, outlining the development of traditional CR and major breakthroughs before the induction of SCR; section 1.3 will focus briefly on spatial sampling and then the major areas of analysis and modeling techniques within the field of SCR.

1.2 Seminal and Classical Models

This section focuses on the development of seminal and classical models in the field of CR analysis, while also including some modern extensions and advances in estimations of these models. Early work is briefly touched upon here but it was not until around the mid 1950's that CR analysis started receiving rigorous statistical treatment. Classical CR analysis can be broken up into two main fields; closed population and open population.

Open population modeling began development around the mid 1960's with the publication of the ground breaking work by Cormack, Jolly and Seber. This period resulted in what is now known as the Cormack-Jolly-Seber and the Jolly-Seber models. For a good review of the development and impact of these models see Schofield et al. [2016]. As will be noted in section 1.2.4, since the main focus of open population modeling is survival and birth rates of populations, less emphasis is put on open population compared to closed population modeling here.

Closed population models mainly emerged around a decade after open population models with an emphasis on estimating the size of a given population. The major themes of closed population analysis laid out by Otis et al. [1978] consist of modeling capture probability in terms of heterogeneity, time variation, or capture response of an individual.

A common dilemma in CR analysis is finding a way to handle the construction of explicit likelihoods where vectors of interest are not indexed in terms of the unknown

abundance parameter N , while still somehow estimating N . The most common traditional approach was to either leave N out of the likelihood and estimate it as a derived parameter or to model frequencies of capture histories as apposed to modeling the captures themselves.

1.2.1 Two sample experiments

As referenced in the introduction, the first estimator of population size using CR data was credited to both Lincoln and Petersen for implementing the idea separately in 1896 (Petersen [1896]) and 1930 (Lincoln [1930]) respectively. The Lincoln-Peterson estimator assumes a closed population and equal probability of capture for all individuals over two sampling periods.

Letting n_1 denote the number of individuals observed during the first sampling period, n_2 the number of individuals observed during the second sampling period and m_2 the number of individuals that were recaptured during the second period, the Lincoln-Peterson estimator for the population size N is defined $\hat{N}_{LP} = \frac{n_1 n_2}{m_2}$.

Originally the Lincoln-Petersen estimator was introduced using the naive idea that $\frac{m_2}{n_2} \approx \frac{n_1}{N}$ and both of these ratios are giving estimates for the probability of a capture. Considering n_1 and n_2 as fixed and assuming m_2 follows a Hypergeometric distribution, Chapman [1951] showed the Lincoln-Petersen method actually corresponds to the Maximum Likelihood Estimator (MLE) and is consistent for N . In the same publication

Chapman also introduced the modified estimator $\hat{N}_{chap} = \frac{(n_1+1)(n_2+1)}{m_2+1} - 1$ which he verified as having less bias than \hat{N}_{LP} , being unbiased for the case where $n_1 + n_2 > N$ and having negligible bias as long as $m_2 > 7$.

The obvious drawback to the Lincoln-Petersen and Chapman estimators are their simplifying assumptions of closed populations and equal probability of capture of all individuals. This being said, the advantage to these estimators are clearly their simplicity both in implementation and interpretation of their results. In the case where this analysis is being run by a non-statistician it may often be a fair trade off to choose one of these methods or their extensions depending upon how severely the researcher believes these strict assumptions may be violated. Chao et al. [2008] presented the Lincoln-Petersen estimator from a 2x2 contingency perspective and gave an estimator for the dependence between the two samples (i.e. an assessment of the equal-catchability assumption). Fairly recently Grimm et al. [2014] included the Lincoln-Petersen estimator and a few of its extensions in a comparison of CR estimators when used on actual field data. Grimm et al. [2014] Also gave an extension of the Lincoln-Petersen estimator for use with more than two sampling occasions which pools captures from multiple occasions in hopes of accounting for heterogeneity better.

1.2.2 Multiple samples with equal catchability

Pollock [1991] credits Schnabel and Darroch for notable early works in closed population estimation. Schnabel [1938] denotes N as the total population size, M_i the number of

tagged individuals by the time the i^{th} sample was made, n_i the number of individuals and m_i the number of recaptured individuals in the i^{th} sample and discusses two models, one assuming $m_i \sim \text{Binom}(n_i, \frac{M_i}{N})$ and the other $m_i \sim \text{Poisson}(M_i n_i / N)$. Approximately¹ for the Binomial model and exactly for the Poisson the MLE for N is given by $\frac{\sum_{i=1}^n n_i M_i}{\sum_{i=1}^n m_i}$.

Darroch [1958] introduced an important foundational model for closed population CR data analysis. Following the notation of Otis et al. [1978] here and denoting t as the total number of samples taken, p_i as the probability that any individual is caught during the i^{th} sample, Darroch's approach models (X_ω) the number of individuals with capture history ω with a Multinomial distribution, yielding the likelihood

$$L[N, p_1, \dots, p_t | \mathbf{X}_\omega] = \frac{N!}{(N - M_{t+1})! \prod_\omega X_\omega!} \prod_{i=1}^t p_i^{n_i} (1 - p_i)^{N - n_i} \quad (1.1)$$

where n_i is the number of individuals caught in the i^{th} sample and M_{t+1} is the number of individuals caught in total (see Appendix A.1 for a more detailed description of the X_ω notation).

Darroch suggests a method for calculating the MLE for N under this model, however Otis et al. point out this will not actually correspond to the true MLE and provide their own method for calculating the MLE on pg106 of Otis et al. [1978]. In spite of this error, and although this was a fairly simple early model, Dorrach laid a foundation built upon by many of the more complex models which followed. Additionally, Dorrach was

¹Approximate due to the fact that a finite number of terms are taken of an infinite sum while maximizing the likelihood equation

already considering the rather sophisticated idea of capture probability varying over the samples by choosing p_i instead of opting for the less complex model which would assume a constant capture probability for all samples.

Orians and Leslie [1958] suggested a goodness of fit test for the Darroch model described above, with the null hypothesis being that capture probability varies just among samples and the alternative stipulating the probability will vary among samples and individuals as well. This test has been built upon in several proceeding publications, Chao and Huggins (Amstrup et al. [2010]) note these tests will require a large amount of data however, and cite a test proposed by Stanley and Burnham [1999] as a more efficient option.

1.2.3 Closed population models

Closed population models make the simplifying assumption that the population is closed to deletions (deaths or individuals migrating out of the population) and addition (births or individuals migrating into the population). Although the more basic models discussed in the previous subsections have also made the closure assumption, the models commonly referred to as "closed population models" generally allow for multiple samples and unequal catchability across samples or individual heterogeneity.

For a list of popular and breakthrough estimation techniques for closed population models see Pollock [1991] and Chao [2001]. For this section unless otherwise stated the notation will be consistent with Otis et al. [1978] (see Appendix A.1 for reference).

A set of models which allowed for unequal catchability and laid the groundwork for modern closed population CR modeling were first introduced by Pollock [1974] (an unpublished Ph.D thesis referenced by Otis et al. [1978]). In total, Pollock discussed eight models: a base model (M_0); three models which allowed capture probability to vary with time (M_t), trap response (M_b) and heterogeneity (M_h); and all the combinations of these models (M_{bh} , M_{th} , M_{tb} , M_{bth}).

M_0 and M_t closed population models

The time varying model M_t and base model M_0 correspond to the Darroch model (the likelihood for which is given in equation 1.1) and the simplified version of the Darroch model which assumes equal probability across capture periods respectively (*i.e.* $p_1 = p_2 = \dots = p_s = p \in (0, 1)$). The rest of the models will be discussed further below as their development and estimation became the main focuses of open population CR articles for many years following their introduction.

M_b closed population model

The trap response model M_b assumes an equal probability of capture p for all individuals who have not been captured and c for all those who have been captured. This approach hopes to avoid bias from animals either being trap 'happy' (being more likely to be captured after their first capture) or trap 'shy' (being less likely to be captured after their first capture). Again building upon the base model M_0 , the trap response model's

likelihood is given by

$$L[N, p, c | \mathbf{X}_\omega] = \frac{N!}{(N - M_{t+1})! \prod_\omega X_\omega!} p^{M_{t+1}} (1 - p)^{N - M_{t+1} - M_\cdot} c^{m_\cdot} (1 - c)^{M_\cdot - m_\cdot} \quad (1.2)$$

with $m_\cdot = \sum_{j=1} m_j$ and $M_\cdot = \sum_{j=1} M_j$. An interesting note is the fact that for this model the subsequent recaptures of individuals (m_2, \dots, m_t) contain no information relevant to the estimation of N and p and will only be used to estimate c . The first captures for each individual alone constitute as sufficient information for N and p , thus in terms of estimating N , the M_b model is statistically equivalent to a removal model (where individuals are removed from the population after capture). Because of this, credit for the derivation of the MLE for N is given to Moran [1951] (a method which was refined by Zippin [1956]) who proposed a method for estimating N in a removal model context.

Moran's conceptual discussion of his removal model in is quite interesting, with the general idea being if the total population is small enough relative to the number of individuals captured and removed, then the total population size can be inferred from the rate at which the captures (or in the case of m_b first captures) decrease over time. A major drawback to this approach in terms of the M_b model is that the recaptures are not being used for the estimate of N or p , thus this model is not truly a CR model at all. Seber [1982] also notes that the MLE for N will generally not exist if the number of first captures does not decrease over time, thus this model should clearly not be utilized

for very large populations or in situations where p is assumed to be very small.

M_h closed population model

The heterogeneity model M_h assumes the probability of capture $\{p_i, i = 1, 2, \dots, N\}$ will vary among individuals but will remain constant across sampling occasions. An initial approach to estimate N under these assumptions was developed by Burnham in an unpublished Ph.D thesis and is outlined in Burnham and Overton [1978]. Burnham assumed the capture of each individual on each day $\{X_{ij}, i = 1, \dots, N, j = 1, \dots, t\}$ followed a Bernoulli(p_i) distribution and to reduce the number of unknown parameters assumed p_i followed some parametric distribution F . This leads to the conditional distribution of X_{ij} given p_i being defined as

$$pr[\mathbf{X}|\mathbf{p}] = \prod_{i=1}^N \prod_{j=1}^t p_i^{x_{ij}} (1 - p_i)^{1-x_{ij}} = \prod_{i=1}^N p_i^{y_i} (1 - p_i)^{t-y_i}$$

where y_i denotes the number of times individual i was caught during the experiment.

Integrating out p Burnham obtained

$$pr[\mathbf{X}|\mathbf{F}] = \prod_{i=1}^N \left\{ \int_0^1 p^{y_i} (1 - p)^{t-y_i} dF(p) \right\} \quad (1.3)$$

$$= \left\{ \int_0^1 (1 - p)^t dF(p) \right\}^{N-M_{t+1}} \prod_{j=1}^t \left\{ \int_0^1 p^j (1 - p)^{t-j} dF(p) \right\}^{f_j} \quad (1.4)$$

with f_j being the number of individuals who were caught exactly j times. Notice in Equation 1.4 that the product index has changed to be in terms of t , which is allowed since it is known $y_i \in \{0, 1, 2, \dots, t\}$ as a result of assuming each individual can only be captured once per sampling period. Once a suitable distribution for F is chosen, Equation 1.4 can then be used to derive a MLE for N . As was noted in Burnham and Overton [1978], Burnham implemented this approach assuming a Beta distribution for F but found the results to be unsatisfactory. This method is still included here since the technique of changing the distribution from being in terms of the individual counts to the capture frequencies is relevant to the derivation of similar likelihoods which can arise when modeling CR data. Additionally, Burnham used this setup to argue the capture frequencies f_1, f_2, \dots, f_t are sufficient for estimating N given an appropriate F under the assumption of heterogeneity.

Burnham and Overton [1978] suggested a non-parametric approach for estimating N under the assumptions of the M_h model and presented Jackknife estimators, which aimed to reduce the bias of the naive estimator M_{t+1} and reassuringly depend upon f_1, f_2, \dots, f_t . Burnham and Overton refer to the 'order' of the Jackknife estimator as the number of samples which are dropped when re-sampling and give the first five of these in Table 1 in Burnham and Overton [1978], the first two are given by

$$\hat{N}_{J1} = M_{t+1} + \frac{t-1}{t} f_1$$

$$\hat{N}_{J_2} = M_{t+1} + \frac{2t-3}{t}f_1 + \frac{(t-2)^2}{t(t-1)}f_2$$

the trend of the order of the Jackknife estimator being associated with the order of the capture frequency continues for higher order estimators, which also continue to have reduced bias. Burnham and Overton [1979] suggest a stopping rule for choosing the order of the Jackknife estimator since they note that the standard error will increase with order. One major drawback to the Jackknife method is it provides no estimate or even confirmation for the existence of heterogeneity.

Advances in estimating abundance under closer

Over the nearly 40 years since their publication there has been a great deal of work towards obtaining abundance estimators for the models introduced by Otis et al. [1978]. Aside from a few recommendations of restricting p'_i 's to reduce nuisance parameters, Otis et al. [1978] offered little advice for estimating the model parameters for the combination of the closed population models (namely M_{bh} , M_{th} , M_{tb} and M_{bth}). Chao et. al. gave an extremely useful summary of common estimation methods used for each model in Table 1 in Chao [2001]. Some of the most successful or interesting of these methods are discussed below. Any methods for modeling abundance under more general models (M_{tbh} being more general than M_{bh} , which in turn is more general than M_h for instance) can be used for their less general sub-models, however Huggins and Chao note in Amstrup et al. [2010] that usually the more general models may have smaller bias but larger variance,

whereas a less general model will tend to have smaller variance.

For the M_{tb} it is assumed the probability of capture varies with each trapping period (j) and that each individual (i) will have a behavior response to their first capture, thus it assumes $p_{ij} = p_j$ for first capture and $p_{ij} = c_j$ for all recaptures. Assuming the capture histories (X_ω) follow a Multinomial distribution, Otis et al. [1978] showed that $\{u_1, \dots, u_t, m_2, \dots, m_t\}$, with u_j and m_j denoting the number of unmarked and marked captures in the j^{th} sample respectively, are minimally sufficient for $N, p_1, \dots, p_t, c_2, \dots, c_t$.

As $\{u_1, \dots, u_t, m_2, \dots, m_t\}$ is of dimension $2t - 1$, while $\{u_1, \dots, u_t, m_2, \dots, m_t\}$ is of dimension $2t$, the parameters are non-identifiable under this model. An obvious solution to combat this issue is to assume some form of a relationship between p_j and c_j to reduce the number of nuisance parameters. One of the more successful examples of this was developed by Chao et al. [2000], where they set $c_j/p_j = \phi$ for $j = 2, \dots, t$, reducing the parameter set to $\{N, \phi, p_1, \dots, p_t\}$, which being of dimension $t + 2$ will allow for identifiability for studies with three or more samples. Under these assumptions, the likelihood can now be written

$$L[N, \phi, \mathbf{p} | \{\mathbf{N}_\omega\}] \propto N(N-1)\dots(N - M_{t+1} + 1)p_1^{u_1}(1 - p_1)^{N-u_1}\phi^m \times \prod_{j=2}^t p_j^{n_j}(1 - p_j)^{N-M_{j+1}}(1 - \phi p_j)^{M_j - m_j} \quad (1.5)$$

Chao et. al. then derived a MLE for N by maximizing Equation 1.5 and a conditional

MLE (CMLE) by noting

$$L[N, \phi, \mathbf{p}] = L_1[N, Q]L_2[\mathbf{p}, \phi]$$

with $Q = \prod_{j=1}^t (1 - p_j)$ denoting the probability of not being observed in the experiment, and then maximizing $L_2[\mathbf{p}, \phi]$ with respect to \mathbf{p} and ϕ to obtain \hat{Q} before maximizing $L_2[N, \hat{Q}]$ with respect to N . Through simulations Chao et. al. concluded the CMLE showed less bias when given a sufficient number of captures but a larger bias than the MLE otherwise.

For M_h it is assumed capture probabilities are heterogeneous, or in other words they vary among individuals. Chao et al. [1992] and Lee and Chao [1994] proposed abundance estimators, known as sample coverage estimators, which involve estimating these heterogeneity effects in terms of p_1, p_2, \dots, p_N coefficient of variation (CV). The CV is a measure of the heterogeneity, with $CV = 0$ corresponding to the case of no heterogeneity (equal catchability) and larger CV corresponding to greater heterogeneity. Sample coverage (C) is defined as the proportion of total individual capture probability made up by the captured individuals, or mathematically

$$C = \frac{\sum_{i=1}^N p_i \cdot \mathbf{I}(i^{th} \text{ individual is captured})}{\sum_{i=1}^N p_i}.$$

Multiple estimators for C are given by Chao et al. [1992] and Lee and Chao [1994],

one of which is given below to illustrate that unlike earlier estimators they are dependent on more than just the first capture frequency. Estimating C as

$$\hat{C} = 1 - \frac{f_1}{\sum_{j=1}^t j f_j}.$$

an estimator for the CV is then given as

$$\hat{\gamma} = \sqrt{\max \left[\frac{tM_{t+1}}{(t-1)\hat{C}(\sum_{j=1}^t j f_j)^2} \sum_{j=1}^t j(j-1)f_j - 1, 0 \right]}.$$

Under M_h , N can then be estimated as

$$\hat{N} = \frac{M_{t+1}}{\hat{C}} + \frac{f_1}{\hat{C}} \hat{\gamma}^2.$$

And so \hat{N} is given by a weighted average of an intuitive bias estimator, namely the total number of individuals observed and a function of an estimator for the heterogeneity (in this case the CV). For M_{th} similar estimators are given by Lee and Chao [1994] with a different estimator for the CV.

Two big advantages to the sample coverage approach are (1) as noted above, and unlike for some other non-parametric approaches the estimator for N relies on all f_1, \dots, f_t and (2) this method gives an explicit estimator for the level of heterogeneity. A downside to this method, at least for those outlined by Chao et al. [1992] and Lee and Chao [1994], is some of the derivations of the estimators may be a bit unrigorous and do not fall back

on more widely accepted methods such as MLE, Jackknife or Bootstrap procedures. Chao and Huggins also note in Amstrup et al. [2010] that these estimators do not seem to perform well without sufficient sample sizes.

Norris and Pollock [1996] presented a non-parametric MLE method for estimating abundance under the assumptions of either M_h or M_{bh} . For the M_h model the authors utilize a multinational distribution for the capture histories (X_ω) and assume the capture probabilities for each individual come from an unknown distribution F , giving the likelihood

$$L[N, F | \mathbf{X}_\omega] = \frac{N!}{(N - M_{t+1})! \prod_\omega X_\omega!} \left[\prod_{i=1}^t (P_{\omega, F})^{X_\omega} \right] (P_{0, F})^{N - M_{k+1}} \quad (1.6)$$

where $P_{\omega, F}$ and $P_{0, F}$ are the probabilities of observing the capture history ω and an individual not being observed under F respectively. Considering the function

$$T_1(N, F) = \binom{N}{M_{k+1}} \left[\prod_{j=1}^N pr(Y_j = y_j) \right], \quad (1.7)$$

where Y_j denotes the number of samples in which individual j was caught and

$$pr(Y = y_j) = \int_0^1 \binom{k}{y_j} p^{y_j} (1 - p)^{k - y_j} dF(p).$$

Norris and Pollock then note that this is proportional to the likelihood given in Equation 1.6 and argue for most datasets Equation 1.7 will be maximized by unique MLE's for

(N, F) . To obtain this maximum the authors suggest testing fixed $N \in \{M_{t+1}, \dots, N^*\}$ where N^* is chosen to be large enough to hopefully not underestimate N .

A nice benefit to Norris and Pollock's method, in addition to not having to specify F , is the fact that this method will also produce \hat{F} which could be useful for many reasons, such as bootstrap estimates relating to properties of N . Pledger [2000] present a more general extension of this approach which can also be used to estimate M_{tth} .

Estimation of parameters under M_{tth} is a very difficult problem. One of the most intuitive approaches is to use a generalized linear models framework to model capture probability with the effects of time, heterogeneity and behavioral response represented by covariates. Huggins [1989] defines γ_{ij} as the probability of capturing individual i in sample j conditionally given that they have been captured at least once during the experiment and the capture histories for all individuals for all samples up to sampling occasion $j - 1$, and models this as

$$\gamma_{ij} = \frac{\exp(\beta_0 + \beta_1 z_i + \beta_2 x_j + \alpha z_{ij})}{1 + \exp(\beta_0 + \beta_1 z_i + \beta_2 x_j + \alpha z_{ij})}$$

where z_i is a vector of individual covariates such as weight or age, x_j is a vector of environmental covariates such as temperature or rainfall and z_{ij} is a vector of individual covariates which may change over time such as capture history. Huggins considers the

likelihood for only the n individuals whom have been observed which is given by

$$L = \prod_{i=1}^n \prod_{j=1}^t \gamma_{ij}^{x_{ij}} (1 - \gamma_{ij})^{1-x_{ij}}$$

with x_{ij} being the indicator of whether individual i was captured during sample j . The key to Huggins approach is to use this likelihood of the the observed individuals only to obtain estimates for the model parameters which at this stage do not include the abundance N . Huggins suggests deriving the MLE's of the parameters from this likelihood and argues these MLE's will still display good properties asymptotically for the true parameters of the full likelihood included unobserved individuals. For small samples Huggins recommends a bootstrap procedure instead. Once the model parameters have been estimated from the partial likelihood given above, Huggins then recommends using a Method of Moments estimate to obtain \hat{N} . The most obvious drawback with this approach again seems to be that it may not perform well with small samples.

Another common approach for reducing bias caused by heterogeneity is to use mixture models. Pledger [2000] introduced a general set of models which divides individuals into G groups separated by their specific capture heterogeneity. Pledger defines π_g as the probability of an individual belonging to group $g = 1, \dots, G$ with $\prod_{i=1}^G \pi_g = 1$. The author also notes that in practice setting $G = 2$ will often be sufficient (this could be viewed as having different capture probabilities for male and female individuals for instance).

For the most general model which lets p_{ij} vary with behavior response, time and heterogeneity, Pledger lets $p_{ij} = \theta_{tbg}$ and models θ_{tbg} as

$$\log\left(\frac{\theta_{tbg}}{1 - \theta_{tbg}}\right) = \mu + \tau_t + \beta_b + \eta_g + (\tau\beta)_{tg} + (\beta\eta)_{bg} + (\beta\eta)_{bg} + (\tau\beta\eta)_{tbg}, \quad (1.8)$$

with μ as a constant and the factors τ_t for time, β for behavior effect, η_g the heterogeneity group and the rest of the terms corresponding to relevant interactions. If there are any relevant covariates that can be included in the model they can simply be added to Equation 1.8. The likelihood for this most general model can then be written

$$L(N, \pi, \theta | \mathbf{X}) \propto \frac{N!}{(N - D)!} \prod_{i=1}^N \sum_{g=1}^G \left[\pi_g \prod_{j=1}^t \theta_{tbg}^{x_{ij}} (1 - \theta_{tbg})^{1-x_{ij}} \right]$$

where x_{ij} denotes the indicator of whether or not individual i was caught in sample j . Pledger suggests maximizing this likelihood using EM algorithm for unknown group membership, and covers using likelihood ratio tests to decide between nested models.

1.2.4 Open population models

As the name would suggest, open population models do not require the closer assumption of Section 1.2.3 and allow for addition (through birth or migration) into the population and deletion (through death or migration) out of the population (although generally it is not assumed distinctions can be made between the two types addition or the two

types of deletions). Traditionally, most open population models assumed homogeneity of capture probabilities (equal catchability) which produced much less sophisticated estimators than those found under closed population models. Additionally, the focus of open population CR model often seems to put more emphasis on estimating survival rates rather than abundance. That being said, open population models are included here for two main reasons: (1) they were critical for the formation of CR analysis in general and (2) depending upon the length of the study, movement of individuals and survival rates of individuals it may not be possible to assume closure of a particular population when attempting to estimate abundance.

For a good overview of the development of open population models see Schofield et al. [2016], for a complete description of the Jolly-Seber likelihood and parameter estimation see Seber [1982], for a brief intuitive description of the Jolly-Seber model see section 8.2.1 of McCrea and Morgan [2014]. The notation used below has kept mostly in line with McCrea and Morgan [2014] and Schwarz and Arnason [1996], for reference see Appendix A.2.

Darroch's open population model

While work on open population modeling was published by C.H.N Jackson and R.A. Fisher as early as the late 1930's (Jackson [1939]) and 1940's (Fisher and Ford [1947]), Pollock and Alpizar-Jara note in Amstrup et al. [2010] that these models contained deterministic elements and should no longer be considered after the introduction of

the models discussed below. One of the earliest influential works in the field of open population modeling, Darroch [1959], was an extension of Darroch's earlier influential closed population model (Darroch [1958]). Darroch [1959] was able to obtain MLE's for one model which allowed for additions to the populations but not deletions and another model which allowed for deletions but not additions. One nice feature of the models presented in Darroch [1958] (which would not remain true for most modern open population models) is that their likelihoods are consistent with the closed population likelihood proposed in Darroch [1958] if it is assumed that the population remains the same size over the whole experiment (i.e. there is no deletions or additions).

Jolly-Seber

The seminal open population CR model was introduced by Jolly [1965] and Seber [1965] and is known as the Jolly-Seber(J-S) model. The J-S model allows for additions due to birth or migration in, deletions due to death or migration out and additionally allows for removals between samples either by design or accidental death of individuals (often referred to as "loss on capture").

As is thoroughly covered in Schofield et al. [2016], while the two models in Jolly [1965] and Seber [1965] were formulated differently (i.e. there are differences in the way random variables are defined and which variables are conditioned on) the resulting final likelihoods are equivalent. The description below follows that given by McCrea and Morgan [2014] which is mainly in line with Seber's formulation in Seber [1982].

The J-S likelihood can be broken up into three main components, lets say $L_{JS} = L_1 L_2 L_3$, with L_1 corresponding to the probability of first capture, L_2 to the probability of loss on capture and L_3 to the probability of recapture. The probability of capturing u_i unmarked (i.e. first capture) individuals during i^{th} sample can be assumed to follow a Binomial(U_i, p_i) where U_i is the number of unmarked individuals in the population and p_i the probability of capture of any unmarked individuals in the i^{th} sample. Thus L_1 can be written as

$$L_1 = \Pr(\text{first capture}) = \prod_{i=1}^t \binom{U_i}{u_i} p_i^{u_i} (1 - p_i)^{U_i - u_i}. \quad (1.9)$$

Note that this expression allows for p_i , the probability of capture to vary over samples.

The second likelihood component which models the probability of loss on capture can be written

$$L_2 = \Pr(\text{loss on capture}) = \prod_{i=1}^t \binom{n_i}{r_i} v_i^{r_i} (1 - v_i)^{n_i - r_i},$$

which again is a product of binomial distributions where n_i denotes the total number of individuals in the i^{th} sample, r_i denotes the number of individuals lost and v_i denotes the probability of loss on capture. Again note this model allows for these loss on capture probabilities to vary over samples.

The final likelihood component which models the probability of recapturing individuals was originally introduced by Cormack [1964] and often leads to the title Cormack-Jolly-Seber(CJS) model. Where as previous work had focused on estimating population size and assumed individuals were sampled from the entire available population, Cormack's model was interested in survival probabilities and as such lead Cormack to condition on released individuals instead (which in the case of the J-S model can be equivalent to conditioning on the captures u_i).

Denoting m_{ij} as the number of individuals caught in the i^{th} sample and recaptured in the j^{th} sample, m_{ij} is assumed to follow a Multinomial distribution. The probability of a particular pair of m_{ij} occurring, (denoted v_{ij}) relies on the probability an individual that was released surviving until the next sampling period (denoted ϕ_i), and the probability an individual who has survived is captured in the i^{th} sample (denoted p_i). For $i < j$, v_{ij} can then be written

$$v_{ij} = \left(\prod_{k=i}^{j-1} \phi_k \prod_{k=i+1}^{j-1} (1 - p_k) \right) p_j.$$

The third component of the J-S model is then the product of Multinomial distributions

$$L_3 \propto \prod_{i=1}^{t-1} \prod_{j=i+1}^t v_{ij}^{m_{ij}} \left(1 - \sum_{j=i+1}^t v_{ij} \right)^{R_i - \sum_{j=i+1}^t m_{ij}},$$

where R_i denotes the number of individuals released after the i^{th} sample (recall this can be less than the number captured n_i due to loss on capture) which are often referred to as "cohorts".

Abundance cannot be estimated directly from the J-S model, rather MLE's are found for \mathbf{p} , ϕ and \mathbf{U} (the sizes of unobserved population at each sample time), the size of the marked population after the i^{th} sample can then be estimated recursively using $M_{i+1} = (M_i + u_i)\hat{\phi}_i$ and the abundance can then be estimated as the sum $\hat{N}_i = \hat{M}_i + \hat{U}_i$. The resulting MLE for $\hat{U}_i = \frac{u_i}{\hat{p}_i}$ gives the abundance estimator for the period after the i^{th} sample as

$$\hat{N}_i = \hat{U}_i + \hat{M}_i = \frac{\hat{M}_i n_i}{n_i - u_i}$$

which corresponds to the Lincoln-Peterson estimator discussed in section 1.2.1.

The J-S model is impressively general, allowing for capture probability, survival rates and birth rates to vary over time. While Nichols notes in Amstrup et al. [2010] that J-S model was practically the only model being used for estimating survival rates for 15-20 years after it's publication, he also notes that it may often be too general, resulting in increased variance of estimates. A good deal of work that came after the J-S model comprised of efforts to impose restrictions on the parameters for more focused application or attempts at including heterogeneity in open population models.

Advances in open population models

As noted by Schwarz and Arnason [1996] a major drawback of the J-S model is the fact that the birth rates² (denoted B_i for the i^{th} sample) is not explicitly involved in

²The term 'birth rate' is used to mean the rate at which individuals enter the population between sampling occasions, whether it be due to actual birth or migration

the likelihood, which can make it difficult to ensure $B_i \geq 0$ and impossible to force B_i to remain constant over time. To combat this issue Schwarz and Arnason [1996] used the idea of a super population, say of size N^* , which represents the total number of individuals that at some point during the experiment were available for capture. They then consider $u_i \sim \text{Multinomial}(N^*; \Psi_1 p_1, \dots, \Psi_t p_t)$, with Ψ_{i+1} (the probability of an individual entering the population, surviving and not being captured before the time the $i + 1$ sample takes place) defined as

$$\Psi_{i+1} = \begin{cases} \beta_0 & \text{for } i = 0, \\ \Psi_i \phi_i (1 - p_j) + \beta_i & \text{for } i \geq 1, \end{cases}$$

where $\beta_i (= \frac{B_i}{N})$ represents the probability of an individual from the super population entering the population between sample i and sample $i + 1$. The first component of the likelihood for the J-S model given in (1.9) can then be replaced by

$$\binom{N^*}{N^* - u} \left(1 - \sum_{i=1}^t \Psi_i p_i \right)^{N^* - u} \prod_{i=1}^t (\Psi_i p_i)^{u_i}$$

with $u. = \sum_{i=1}^t u_i$ being the total number of individuals captured during the experiment. Schwarz and Arnason note that this likelihood will produce the same estimates for ϕ and \mathbf{p} as the J-S likelihood while being numerically easier to optimize. In addition, it is now possible to impose restrictions on the birth rate B_i and survival rate ϕ_i , and the authors show that under the restrictions of birth only ($\phi_1 = \dots = \phi_T = 1$) or death only

($\beta_1 = \dots = \beta_T = 0$) this new model produces the same estimates as birth only and death only models previously proposed by K.P. Burnham³.

One approach that has allowed for including heterogeneity in open models has been to perform primary and secondary sampling. This approach assumes primary samples are spaced far enough apart for the population to be effected by additions or deletions while the secondary samples are close enough to assume the population is closed. For example, imagine sampling the first five days of every month, producing twelve primary samples each with five secondary samples. The main advantage added by the closure assumption for the secondary samples is the ability to use closed population models which can incorporate heterogeneity into the capture process.

Kendall et al. [1995] were the first to give formal likelihoods using this concept. The authors consider the probability of capture within the i^{th} primary sample and the j^{th} secondary sample (p_{ij}) and the probability of being captured at least once within the i^{th} primary sample (p_i). The likelihoods are then made up of two components, the first using the J-S model to estimate survival rates using data involving captures across primary samples and the second modeling the capture process within each secondary sample. Kendall et al. [1995] implemented this approach deriving MLE and conditional MLE estimates for several of the closed population models introduced by Otis et. al. in Otis et al. [1978].

Kendall et. al. note in simulations the reduction in bias introduced by heterogeneity

³In an unpublished manuscript, a matching estimator for the death only model can be found in Pollock et al. [1990]

are especially significant when there are low capture rates. A downside of this approach is the fact that loss on capture cannot be included in these models as the closer assumption is made within the secondary samples.

A major drawback to the J-S model is the fact that it only allows for permanent immigration in or out of the population. In the original models there is no way to distinguish between immigration in and birth or immigration out and death. Kendall et al. [1997] relax this assumption by extending the models which use primary and secondary sampling methods described above.

The authors utilize the idea of a super-population of size N_i^0 which consists of all individuals who have a negligible chance of being in the trapping area during i^{th} primary sampling period. Additionally, N_i denotes the population size of permanent residents (those which are present in the trapping area for all of the secondary samples of the i^{th} primary period) and γ_i denotes the probability a member of the super-population is not in the trapping area during the i^{th} sampling period, so that $E[N_i] = (1 - \gamma_i)N_i^0$.

1.3 Spatial Capture-Recapture

Spatial capture-recapture (SCR) models, also referred to as spatially explicit capture-recapture (SECR) models, fall under the category of capture-recapture models which assume heterogeneity of capture probability among individuals. Any CR dataset which has captures occur at multiple trapping locations and keeps track of these locations

can be considered SCR data. Regardless of whether spatial covariates or individual movement is considered, a core concept shared by essentially all SCR models is the idea that heterogeneity of captures probabilities is caused in part due to an individual's location relative to a given trap (i.e. an individual closer to the trap will in general be easier to capture than an individual who is farther away).

A major innovation utilized in most SCR models is to assume individuals possess latent variables known as activity centers (also referred to as home ranges or home centers) which in general is a unique point associated with the individual's location throughout the experiment. These centers are often assumed to be distributed according to some spatial point process, which allows for a statistical rigorous estimation of individual density throughout a trapping area from CR data.

SCR likelihoods in general consist of two major components, (1) the spatial point process for distribution of activity centers and (2) the probability of capturing an individual. One can either take the approach of estimating the density of activity centers or estimating the number of activity centers (which is equivalent to estimating abundance) with most SCR likelihoods easily converting between these two approaches.

In the following subsections models are considered in terms of four different types of trapping procedures; single-catch, multi-catch, proximity traps and area searchers. This distinction is made as there are unique dependency structure between traps and individuals which arise from these three types of trapping experiments. For instance, for single trap experiments once an individual is captured in a particular trap no other

individual can be caught in that trap nor can that individual be caught in any other trap during that trapping period, while for proximity traps multiple individuals can be caught by a single trap and individuals can be caught by multiple traps in the same trapping period.

While some models will overlap or are capable of handling more than one type of trapping experiment an effort was made to highlight models which have distinct approaches for handling common issues which arise when modeling SCR data. In general, the most difficult of these issues is estimating latent variables and specifying models with parameter vectors indexed by the unknown population size N . While in classical CR this was often achieved by modeling the frequencies of capture histories, SCR modeling often requires more creative solutions as capture probabilities are considered in terms of individual covariates. The four most common approaches for overcoming these challenges highlighted below are that of inverse prediction, conditioning on observed data, augmenting data, and the 'full' likelihood approach.

This section begins with a brief review of spatial sampling which was somewhat of predecessor to SCR.

1.3.1 Spatial sampling

As traditionally spatial sampling models were applied to data which was collected on one occasion (and thus did not contain recaptures) they generally are not consider as SCR models. Spatial sampling is still briefly included here for two reasons: (1) to clarify

the differences between SCR and spatial sampling models and (2) since spatial sampling models were introduced prior to the emergence of SCR models and have influenced their development. In some sense, many SCR models can be thought of as the bridge between spatial sampling and CR models.

Plot sampling

In plot sampling, "plots" of subareas are chosen from the entire sampling area and exhaustively sampled (usually on just one occasion) with the assumption being made that all individuals within these subareas will be observed (i.e. the probability of capture of an individual given the individual is within the plot is 1). As noted by Borchers et al. [2002] there are two perspectives one can approach plot sampling from: (1) a design perspective where the randomness is introduced through randomly selecting plots and (2) from a modeling perspective where randomness is introduced through the modeling of the location of individuals within the sampling region. In general the likelihoods based around the design approach assume the position and number of individuals are fixed and cannot be used directly for inference involving abundance. Because of this, only the modeling approach is covered here, for a brief discussion and example of the design approach see Borchers et al. [2002] pages 56-60.

The simplest approach to plot sample modeling outlined in Borchers et al. [2002] is two assume individuals are uniformly distributed throughout the total sample area (A). Given the total area of the chosen plot(s) is a , then the probability of observing

an individual during the experiment will be given by a/A . The probability of capturing n individuals during the experiment can then be modeled as a *Binomial* $(N, \frac{a}{A})$ random variable, which leads to the MLE estimator of $\hat{N} = nA/a$.

Although this approach may seem overly simple, this type of plot sampling is still widely used especially in botany where multiple sampling periods are unnecessary and the assumption of all individuals within the plot being caught is fairly reasonable. Borchers gives a brief outline of a more general model not assuming certain capture or equal distribution of individuals in Borchers [2012].

Distance sampling

Distance sampling modeling again generally only involves one sampling period and attempts to use the distance of captured individuals from an observer to estimate capture probability. The two major types of distance sampling models are point transect and line transect models. In point transect surveys, an observer will stay at a fixed point and attempt to detect (visually or audibly) individuals in a complete circle around themselves with radius w (which is either the maximum distance the observer can detect individuals or is simply a chosen cutoff point). In line transect surveys, observers move vertically along a specified rectangular plot, of width w , usually noting the perpendicular distance to detected individuals. Neither type of distance sampling generally assumes individuals within specified detection areas are guaranteed to be observed but it is often assumed the probability of detection of individuals at distance zero from the observer is equal

to 1. For a good introduction to distance sampling models see Borchers et al. [2002], a particular model by Borchers et. al. is discussed below as it was one of the early links between distance sampling and CR.

Borchers et al. [1998] developed a model referred to as a mark-recapture distance sampling (MRDS) model. This particular model considers combining CR modeling with line transect distance sampling models (for a MRDS model which utilizes point transect sampling see Kissling and Garton [2006]). Borchers et al. [1998] still assume there will only be one sample period, however there will now be two observers performing independent line transect surveys of the same subarea. By defining the event that an individual is detected by both observers as a recapture, the authors essentially convert the survey into a two sample CR survey. Borchers et. al. define the possible capture histories for individuals as

$$c_i = \begin{cases} 1 & \text{if detected by observer 1,} \\ 2 & \text{if detected by observer 2,} \\ 3 & \text{if detected by both observers.} \end{cases}$$

The probability of an individual having a particular capture history can then be written

$$P_c = \begin{cases} p_1(x_j, \mathbf{z})(1 - p_2(x_j, \mathbf{z})) & \text{if } c=1, \\ (1 - p_1(x_j, \mathbf{z}))p_2(x_j, \mathbf{z}) & \text{if } c=2, \\ p_1(x_j, \mathbf{z})p_2(x_j, \mathbf{z}) & \text{if } c=3. \end{cases}$$

With x_j denoting the perpendicular distance between the observer and j^{th} individual, \mathbf{z}_j denoting covariates associated with the capture of the j^{th} individual and $p_i(x, \mathbf{z})$ denoting the probability of an individual being detected by the i^{th} observer. Finally, the authors consider separating individuals and their captures into k separate bins based on (x_j, \mathbf{z}_j) and define

$$P_{ck} = \int \int_{I_k} P_c(x, \mathbf{z}) \pi(x, \mathbf{z}) dx d\mathbf{z} = Pr(\text{individual in bin } k \text{ has capture history } c),$$

$$P_{..} = \int \left(\int \int_{I_k} \left(\sum_{c=1}^3 P_c(x, \mathbf{z}) \right) \pi(x, \mathbf{z}) dx d\mathbf{z} \right) dk = Pr(\text{individual is observed}),$$

where $\pi(x, \mathbf{z})$ denotes a joint density for x and \mathbf{z} , $\int \int_{I_k} \dots dx d\mathbf{z}$ is integrating over all possible values of x and \mathbf{z} in the k^{th} bin and $\int \dots dk$ is integrating over all possible bins.

The likelihood for N and parameters θ and ϕ which are associated with the chosen

detection function $p_i(x, \mathbf{z})$ and the distribution $\pi(x, \mathbf{z})$ respectively is then given by

$$L(N, \theta, \phi; \mathbf{n}_c) = \frac{N!}{(N - n_{..})! \prod_c \prod_k n_{ck}!} (1 - P_{..})^{N - n_{..}} \prod_c \prod_k (P_{ck})^{n_{ck}},$$

where n_{ck} is the number of individuals caught with capture history c in bin k and $n_{..}$ is the total number of individuals observed. Clearly this likelihood is taking a similar approach to that of classical closed population CR by modeling the number of individuals with capture history c (n_c) with a Multinomial distribution. As such, similar inference techniques as those used for the classical closed population models will apply here as well.

While there is nothing inherently spatial in this model (with only the distance to individuals being utilized as opposed to their location), spatial covariates could in theory be included in \mathbf{z} . More recently Borchers et al. [2015] introduced a unifying model for distance sampling and SCR models. Borchers et al. [2015] model represents the first time a single model could truly take advantage of distance sampling data for the purposes of spatial modeling techniques discussed in the proceeding chapters.

1.3.2 Single-catch trap experiments

Single-catch traps can only catch and hold one individual at a time. In single-catch experiments it is thus usually assumed that each trap can catch at most one individual per trapping period, and individuals can only be caught at most once per trapping

period. Single-catch experiments are generally the most difficult to model and are the least studied of the different trapping experiments discussed here.

One method presented by Balderamaa et al. [2012] attempts to model single-catch data using a combination of the competing risk approach employed by Borchers and Efford [2008] (discussed below in Section 1.3.3) and the augmented data approach proposed by Royle et al. [2007] (discussed below in section 1.3.4). The Balderamaa et. al. model constructs a likelihood for the probability of capture (for a particular individual at a particular trap during a particular sampling occasion) as a series of conditional terms which take into account the order in which individuals were exposed to the trapping array. While the modeling of the capture probability seems inherently flawed (in that the probability of individuals being captured at a particular trap increases as other traps fill up) it is still noted here as the authors have explicitly included the ordering of individuals in a novel way which could be of use in many SCR models.

Inverse prediction approach

Efford [2004] introduced what is referenced by Borchers [2012] as the first SCR modeling technique. Efford's approach does not rely on a particular likelihood but instead utilizes inverse prediction with Monte Carlo simulation and any CR model with closed form estimators. Although Efford [2004] mainly focuses on single-catch experiments it presents a very generalized method that can be extended to any type of CR experiment. As noted by Borchers et al. [2016], inverse prediction has mainly been phased out in

favor of MLE or Bayesian approaches but still remains one of the only methods capable of explicitly handling single-catch data.

Efford's approach considers the density of individual activity centers as a result of a spatial point process with unknown intensity D . He assumes capture probability between a particular individual i and trap j on trapping occasion t follows the half normal model given below

$$p_{ij} = g_0 \exp\left(-\frac{d_{ij}^2}{2\sigma^2}\right). \quad (1.10)$$

Where g_0 denotes a base probability of capture if the individual's activity center is centered at trap j , d_{ij} is the distance between individual i and trap j and σ is a scale factor to adjust the rate at which capture probability will decline with distance.

Efford notes that jointly estimating $\theta = (D, g_0, \sigma)$ is analytically intractable (which to date for single-catch experiments with discrete time point data still seems to be true), instead he extends an inverse prediction method first applied to CR data by Carothers [1979]⁴.

Robustness of Borchers and Efford's multi-catch model

Another approach for dealing with single-catch data is to simply apply a multi-catch model. For high density of individuals it would be expected that a multi-catch model would underestimate density if applied to single-catch data but there is evidence to

⁴mainly noted here as Carothers description of the general approach to the inverse prediction method is slightly easier to follow than that given by Efford

suggest multi-catch models may be surprisingly robust in this circumstance. Efford et al. [2009b] showed through a simulation study that the multi-catch model by Borchers and Efford [2008] (discussed at length in Section 1.3.3) will remain roughly unbiased as long as trap saturation (the number of occupied traps) remains below 86%. Efford et al. even concluded for higher levels of trap saturation that the parameter σ from Equation 1.10 will remain roughly unbiased with the only major loss in precision being a underestimation of g_0 .

1.3.3 Multi-catch trap experiments

Multi-catch traps allow multiple individuals to be caught in one trap. In multi-catch experiments it is usually assumed an individual can be caught at most once per trapping period. Under these assumptions there is no competition between individuals to see which individual is caught first in a particular trap as in single-catch traps, however there is still competition among traps to catch each individual as there cannot be recaptures in the same trapping period.

Conditioning on observed data approach

Borchers and Efford [2008] introduced one of the first SCR models to combine an explicit likelihood with the idea that the individual activity centers are a realization of a spatial point process. Borchers and Efford's approach uses the novel idea of handling the competition among traps by considering capture as a death and modeling captures

with a competing risks survival model. The authors note that this type of model would generally be used for survival analysis when there are K possible causes of death while only one cause will eventually kill the individual (or in this case only one trap can possibly capture an individual). After each trapping period in line with this analogy each captured individual will be brought back to life, or in reality released from a trap.

Borchers and Efford denote the total number of captured individuals as n and assume each of these individuals have a home range center coordinate (\mathbf{s}), with these coordinates being a realization of a inhomogeneous Poisson process with rate parameter $D(\mathbf{s}; \phi)$. Here ϕ is a parameter vector associated with the rate parameter of the Poisson process with authors hoping to model individual density within the trapping area with their likelihood as apposed to abundance.

Capture probabilities rely on a hazard function, with the probability of an individual located at \mathbf{s}_i being captured in any of the $k = 1, \dots, K$ traps on occasion $t = 1, \dots, T$ given by

$$p_{.t}(\mathbf{s}; \theta) = 1 - \exp\left(-\sum_{k=1}^K h(d_k(\mathbf{s}); \theta)\right),$$

in this case considering capture at a single trap with $\theta = (g_0, \sigma)$ and choosing the hazard function

$$h(d_k(\mathbf{s}); \theta) = -\ln\left(1 - g_0 \exp\left(-\frac{d_{ik}^2}{2\sigma^2}\right)\right),$$

will lead to the same half normal capture function considered by Efford [2004] given in Equation 1.10.

Denoting $\omega_{it} = k$ as the event individual i is captured at trap k during the t^{th} sampling occasion, the complete capture history for individual i is then written as the vector $\omega_i = (\omega_{i1}, \dots, \omega_{iT})$. Borchers and Efford then write the likelihood for the capture parameter θ and the density parameter ϕ as the joint distribution for n and $\omega_1, \dots, \omega_n$

$$L(\phi, \theta | n, \omega_1, \dots, \omega_n) = \Pr(n | \phi, \theta) \Pr(\omega_1, \dots, \omega_n | n, \phi, \theta). \quad (1.11)$$

The inhomogeneous Poisson process which was imposed on the activity centers leads to n following a inhomogeneous Poisson process with rate parameter

$$\lambda(\phi, \theta) = \int_{\mathcal{R}^2} D(\mathbf{s}; \phi) p.(\mathbf{s}; \theta) d\mathbf{s},$$

where $p.(\mathbf{s}; \theta)$ denotes the probability of an individual being captured at least once during the experiment. MLE estimates of ϕ and θ can then be obtained by maximizing the likelihood given in Equation 1.11, which leads to the abundance estimator $\hat{N} = \int D(\mathbf{s}; \hat{\phi}) d\mathbf{s}$. Thus this approach avoids some of the intractability which would have been introduced by including N directly in the likelihood by instead modeling density parameters in terms of observed individuals only.

1.3.4 Proximity detection experiments

Proximity detection traps are the least restrictive of the different trapping methods, allowing for multiple individuals to be caught in the same trap and individuals to be caught in multiple traps during a single sampling occasion (or sometimes even the same trap multiple times). As a result of this less restrictive setup it is generally assumed all trapping events are completely independent across individuals and traps.

Augmented data approach

Royle et al. [2009a] Royle et. al. introduced a hierarchical Bayesian approach for modeling data from camera trapping experiments. Their approach utilized data augmentation which was first introduced by Royle et. al. in Royle et al. [2007] to handle the challenge of performing inference for Multinomial models with unknown index.

In Royle et al. [2009a] it is assumed individual activity centers (\mathbf{s}_i) are a realization of a Binomial spatial point process which leads to them being distributed uniformly over the sampling area of interest. For the case where it is assumed individuals can only be captured once per trapping period per trap y_{ikt} denotes the indicator variable of individual i being caught at trap k during the t^{th} trapping period and is modeled

$$y_{ikt} \sim \text{Bernoulli}(\pi_{ik}),$$

where $\pi_{ik} = 1 - \exp(-p_{ik})$ with p_{ik} being the same half normal equation given in Equation

1.10. Under this setup the following linear equation can be achieved by applying a cloglog link

$$\text{cloglog}(w_i \pi_{ik}) = \log(g_0) - \frac{d_{ik}^2}{\sigma}.$$

Royle et. al. then augment the data of observed individuals with rows of zeroes corresponding to $M - n$ unobserved individuals where M is chosen large enough to ensure N will not be underestimated and introduce M latent indicator variables, $\{w_i; i = 1, \dots, M\}$, defined as

$$w_i = \begin{cases} 1 & \text{if individual } i \text{ belongs to the true population of size } N, \\ 0 & \text{o.w.} \end{cases}$$

To estimate the model parameters when N is unknown the authors implemented the model in WinBUGS with $\pi_{ik} = w_i(1 - \exp(-p_{ik}))$ and a hierarchical setup up of a Bernoulli(ψ) prior on w_i , a Uniform(0, 1) prior on ψ .

In practice if the model converges properly for observed individuals it will result in $w_i = 1 \{i = 1, \dots, n\}$ and for augmented individuals $w_i \approx \frac{N-n}{M} \{i = n + 1, \dots, M\}$. An estimate for N can then be given by $\hat{N} = \sum_{i=1}^M \hat{w}_i$.

The augmented data approach is a very versatile approach which ingeniously skirts the issue of dealing with inference with parameter vectors of unknown length by working with the fixed value M . Royle et. al. state that M can be chosen by trial and error, choosing too small of an M will truncate the posterior for N while choosing too large

of an M should hopefully only increase convergence time. Royle et al. [2013a] note a more formal analysis for choosing an optimal M is needed. A possible drawback of this method noted by Schofield and Barker [2014] is the fact that it is difficult to impose priors on N as it is a derived quantity not explicitly contained within the likelihood.

Proximity detection with Poisson point process

The only major limitation for data augmentation which Dorazio [2013] cautions against comes from the fact that the approach was developed assuming a Binomial spatial point process. As a result of this, for populations with activity centers distributed according to a spatial Poisson point process in small regions or with low density where Poisson limiting theory would not apply, data augmentation models may exhibit higher bias.

To combat this issue Dorazio constructed a marginal likelihood for the observed data under the Poisson point process assumption for proximity detection experiments in Dorazio [2013]. This model uses a very similar approach to that outlined in section 1.3.5 for area search data and also can be used to derive MLE or Bayesian estimates for N .

Efford et al. [2009b] extended Borchers and Efford's multi-catch model from Borchers and Efford [2008] (discussed in section 1.3.3). This is essentially an identical model to the multi-catch one with a few small changes to the capture probability function to allow for captures of individuals in multiple traps during the same sample period. Additionally, as with the multi-catch models, this extension can work with a Poisson point process.

1.3.5 Area search experiments

Area search experiments generally involve one or more observers searching a specified area and noting the location of observed individuals. It is usually assumed individuals can only be captured once per sampling period although unlike single or multi-catch experiments there is no 'trap' to fill up in this situation so this restriction does not impose any additional dependence between traps or individuals captures.

As there are no trapping locations there is an additional component to area search experiments in that each capture can take place at a unique location. As a result, individual's locations for each day will usually need to be specified in the model.

Royle and Young [2008] introduced a model which utilized the data augmentation method for parameter estimation discussed in section 1.3.4. The authors assume individuals have activity centers, \mathbf{s}_i , uniformly distributed throughout a sampling area, say A . It is assumed that the individuals move about their activity centers according to some probability distribution, $g(\mathbf{s}_i, \theta)$ (for example the observed data likelihood model discussed below uses a bi-variate normal distribution centered at each activity center), resulting in a coordinate for individual i on sampling occasion t of \mathbf{u}_{it} .

The indicator variable of individual i being captured on sampling occasion t is then modeled as

$$y_{it} \sim \text{Bernoulli}(p(i, t) * \mathbb{1}(\mathbf{u}_{it} \in A))$$

Where $p(i, t)$ gives the probability of observing an individual which is currently within

the sampling area A and $\mathbb{1}(\mathbf{u}_{it} \in A)$ indicates the individual is in fact currently within the sampling area.

As was noted in section 1.3.4, a possible drawback of the model outlined above is the fact that data augmentation may show increased bias for smaller sampling areas if individuals are not uniformly distributed. As the entire area will be sampled it may often be the case in area search experiments that the sampling area is relatively small. The model described below avoids this issue and can be used with either a homogeneous or inhomogeneous spatial Poisson point process.

Observed data likelihood approach

Dorazio [2013] derives the marginal distribution for observed individuals (with the activity centers averaged out) and then utilizes this likelihood to find MLE estimates for model parameters. With these estimates Dorazio then outlines procedures for both classical and Bayesian estimates of the total abundance.

Dorazio's model can be implemented with either a homogeneous or inhomogeneous spatial process. For the inhomogeneous case activity centers, \mathbf{s}_i , are considered a realization of a inhomogeneous spatial Poisson process with an intensity function $\lambda(\mathbf{v}(\mathbf{s})) = \exp(\beta'\mathbf{v}(\mathbf{s}))$ where $\mathbf{v}(\mathbf{s})$ is a vector of regressors at locations \mathbf{s} and β is a vector of parameters. Depending upon the choice of Poisson process a conditional distribution of the activity centers given N , $[s_1, \dots, s_N|N]$ will be inferred.

Dorazio defines the conditional distribution of individual i 's location during the t^{th}

sampling period given their activity center as $[\mathbf{u}_{it}|\mathbf{s}_i] \sim N_2(\mathbf{s}_i, \sigma^2\mathbb{I}_2)$ and the conditional distribution of whether or not the i^{th} individual was observed during the t^{th} sampling occasion given their location as $[y_{it}|\mathbf{u}_{it}] \sim \text{Bernoulli}(p_0\mathbb{1}(\mathbf{u}_{it} \in A))$ (so that the capture probability is given by the constant parameter p_0 if the individual is within the trapping area A). Dorazio then splits the conditional joint distribution for \mathbf{u}_{it} and y_{it} given \mathbf{s}_i into separate components for observed

$$[\mathbf{u}_{it}, y_{it} = 1|\mathbf{s}_i] = N(\mathbf{u}_{it}|\mathbf{s}_i, \sigma^2, p_0)p_0,$$

and unobserved individuals

$$[y_{it} = 0|\mathbf{s}_i] = 1 - \int_A N(\mathbf{u}_{it}|\mathbf{s}_i, \sigma^2, p_0)p_0 d\mathbf{u}_{it},$$

where $N(\mathbf{u}_{it}|\mathbf{s}_i, \sigma^2, p_0)$ denotes a bivariate normal probability distribution function evaluated at \mathbf{u}_{it} . For the unobserved individuals it is necessary to integrate out all possible locations as the individuals true location was unobserved for at least sampling period t .

Then assuming all captures are independent across individuals and sampling periods are independent Dorazio obtains the joint distribution of the y_{it} , \mathbf{u}_{it} , n and n_0 with \mathbf{s}_i

marginalized out as

$$\begin{aligned}
 [\mathbf{y}, \mathbf{u}, n, n_0 | \theta] = & [N] \frac{N!}{n_0! n!} \left(\int_{\mathcal{S}} [s_i | N] [y = 0 | s] ds \right)^{n_0} \\
 & \times \prod_{i=1}^n \left\{ \int_{\mathcal{S}} [s | N] \left(\prod_{t: y_{it}=1} [\mathbf{u}_{it}] \right) \left([y_{it} = 0 | \mathbf{s}] \right)^{T-y_{it}} ds \right\}
 \end{aligned} \tag{1.12}$$

where $\theta = (\beta, \sigma^2, p_0)$, \mathcal{S} is the geographic region encompassing the entire area individuals may move about (in general $\mathcal{S} > A$ as some individuals will have activity centers near the border of A), y_i denotes the number of $y_{it} = 1$ for individual i and $[N]$ is given by a Poisson distribution with the rate parameter being determined as a result of the chosen spatial Poisson point process. Summing over all possible values of n_0 then gives the marginal distribution of \mathbf{y}, \mathbf{u} and n for the observed individuals

$$[\mathbf{y}, \mathbf{u}, n | \theta] = \sum_{n_0=0}^{\infty} [\mathbf{y}, \mathbf{u}, n, n_0 | \theta]. \tag{1.13}$$

MLE estimates can then be found for the model parameters numerically from Equation 1.13 in addition inference can be performed using the conditional distribution $[n_0 | \mathbf{y}, \mathbf{u}, n]$ which can be derived from Equations 1.12 and 1.13. This approach is another example of a creative way of avoiding indexing likelihood functions in terms of the unknown N and having a closed form likelihood involving n_0 allows for many different estimation possibilities.

1.3.6 Individual movement

The last topic briefly covered here address recent work in individual movement. In Borchers [2012] review of SCR he recommended incorporating non-stationary activity centers into models could lead to fruitful work. Gardner et al. [2010a], in which Gardner et. al. extended a JS model (discussed in section 1.2.4) to handle spatial data using an augmented data approach, the authors suggested the possibility of modeling a time varying activity center such as

$$\mathbf{s}_i(t) \sim N_2(\mathbf{s}_i(t-1), \sigma^2 \mathbb{I}_2). \quad (1.14)$$

Ergon and Gardner [2014] combined a hierarchical model similar to the model from Royle and Young [2008] and a CJS model and included time varying activity centers to account for dispersion. This model did not include abundance estimates but instead aimed to use spatial data to help distinguish between migration and death. The model included 'dispersion' of individuals by updating time varying activity centers (x, y) coordinates according to the following rule

$$\mathbf{s}_i(t) = (s_{x,i}(t-1) + \cos(\theta_{it}), s_{y,i}(t-1) + \sin(\theta_{ik})).$$

In the recent papaer Royle et al. [2016] a simulation study was performed to determine how robust SCR models such as Royle and Young [2008] would be to time varying

activity centers. The authors consider activity center movement according to Equation 1.14 above and assuming a subset of individuals will have one large random shift to a new activity center coordinate throughout the experiment.

While it was found abundance estimators were surprisingly robust to even permanent activity center movement the authors still recommend fitting a model which contains a component to account for individual movement. The main reason for this recommendation was due to the fact that the σ parameter (as defined in Equation 1.10) appears to increase to incorporate the movement in activity centers which leads to a difficult ecological interpretation of the parameter as compared to when centers are stationary.

A model more relevant from the ecological viewpoint was introduced by Royle et al. [2013b]. authors attempt to utilize data from both telemetry (such as color tracking studies) and SCR studies. It is noted that assuming individuals move around their activity centers according to a bivariate normal distribution makes the simplifying assumption that individuals move around their activity centers in a symmetric fashion, while in reality individual movement is most likely motivated by spatial distributed resources.

Royle et al. [2013b] assume the trapping area is separated into a finite number of cells with each cell containing a measurable amount of a certain resource of interest. A likelihood is then introduced which can be separated into two components: (1) space usage (2) a traditional SCR likelihood. For the first component of space usage $M(\mathbf{x})$ is defined as the number of times an individual visits cell \mathbf{x} during R observations of the

trapping area. It is then assumed the vector of cell use frequencies can be modeled as

$$\mathbf{M} \sim \text{Multinomial}\left(R, \pi[\mathbf{s}_i, z(\mathbf{X})]\right),$$

so that the probability of an individual visiting cell \mathbf{x} , is dependent on the individual's activity center \mathbf{s}_i and covariates associated with the cell $z(\mathbf{x})$. For the second component of the likelihood Royle et. al. chose a SCR likelihood similar to that used by Dorazio [2013] discussed above in section 1.3.5.

A possible area of concern for the model presented by Royle et al. [2013b] is in their assumption that the telemetry data is completely independent from the SCR data. This assumption allows the authors to easily combine the two components of the likelihood but may be an unlikely scenario for most datasets resulting from these forms of study.

1.3.7 Discussion

If the rapid developments of SCR models over the past decade are any indication, it would seem many of the future developments in SCR modeling will come from advances in new technology to improve sampling methods. As sampling technology such as GPS tracking chips and cameras become cheaper and more accessible for instance, one could imagine a greater need for models where data from multiple sources are combined to estimate shared parameters of interest (similar to that presented by Royle et al. [2013b]).

One area of research which has not been touched upon so far (as it currently makes up

a fairly small sub field of SCR) which could become more prevalent given an increase in camera trapping studies is continuous time SCR. While it is often the case that multiple captures of a single individual during a single sampling period from proximity trapping studies are reduced down to one capture per trap per trapping period, this uncompressed (or rather continuous time) data could hold a wealth of information.

In a recent paper Distiller and Borchers Distiller and Borchers [2015] were able to develop a model for single-catch experiments using the data augmentation approach assuming they knew the time and order in which individuals were captured. In another recent paper Borchers et. al. Borchers et al. [2014] developed a continuous time model for camera trapping experiments which they found to display less bias than a discrete-time counterpart and proved to be more robust to individual movement as well.

Another topic which may prove fruitful is applying reversible jump MCMC (RJMCMC) to spatial models. In Schofield and Barker [2014] Schofield and Barker used RJMCMC to estimate a closed population model assuming heterogeneity. As RJMCMC allows for simulation even when the number of parameters is unknown, Schofield and Barker were able to construct a likelihood containing N , hence their criticism of the augmented data approach which does not allow for this (noted in section 1.3.4). This would allow for more flexible choices of priors on N , but as of yet it does not seem this method has been applied to any SCR models.

Perhaps the largest area of untapped research seems to be in incorporating a dependence structure between individual's captures. Almost all work to date has simply

assumed independence of captures across individuals and multiple captures of the same individual, even in cases where these assumptions would seem obviously violated from an ecological perspective. In at least one exception Reich and Gardner [2014] used a Strauss process, which (unlike the Binomial and Poisson point processes which generally used for SCR) allows for dependence to include possible avoidance or territorial behavior of individuals. Modeling these types of dependence via capture function covariates or individual movement schemes could allow for modeling real time dependence in future research.

Chapter 2

A Spatial Capture-Recapture Model with Dependent Animal Movement

2.1 Introduction

A critical component for the conservation and management of any animal species is the ability to accurately estimate that species' abundance. Since early on in the development of modern capture recapture (CR) data analysis for this purpose, models have taken into account the possibility of varying capture probability across individuals (Otis et al. [1978], White [1982], Pollock and Otto [1983]). Spatial Capture Recapture (SCR) models, first presented by Efford [2004], hope to explain capture probability heterogeneity as a function of an individual's proximity to a trapping device.

Efford [2004] laid a framework followed by most SCR models to this day, having two major components: (1) a spatial point process for distribution of activity centers (a coordinate an individual will center it's movement around) and (2) the probability of capturing an individual (generally a decreasing function of the distance between the

individual's activity center and a given trapping device).

For nearly two decades there has been a great deal of work put into advancing both the estimation and scope of these models. Researchers now have many tools at their disposal, with the inverse prediction method (Efford [2004]) numerous Bayesian approaches (King et al. [2016], Dorazio [2013], Royle et al. [2009a]) and maximum likelihood models (Efford et al. [2009b], Borchers and Efford [2008]). There are models to handle different types of trapping methods such as single-capture traps (Distiller and Borchers [2015], Efford [2004]), multi-capture traps (Borchers et al. [2015], Borchers and Efford [2008]), proximity detectors (Royle et al. [2009a], Efford et al. [2009a], Royle et al. [2009b]) and area searches (Dorazio [2013], Efford [2011], Royle and Young [2008]). In recent years SCR models have continued to develop to overcome more specific challenges such as missing classification data (Ketz et al. [2019], Royle et al. [2015]), incorporating telemetry data (Royle et al. [2013b], Sollmann et al. [2013a]), including individual's time of arrival (Schofield et al. [2018], Distiller and Borchers [2015], Borchers et al. [2014]), open population modeling (Chandler et al. [2018], Glennie et al. [2017], Gardner et al. [2010a]) and partially marked individuals (Augustine et al. [2018], Chandler et al. [2013]) to name a few.

Another recent area of interest in SCR has been to incorporate realistic animal movement. Royle et al. [2016] and Ergon and Gardner [2014] both allowed for time varying activity centers due to animal dispersion, while Royle et al. [2013b] incorporated individual movement in terms of resource selection. Borchers and Fewster have also

noted this as a promising area for future research in a recent review of SCR (Borchers et al. [2016]).

Despite all of these advancements in SCR there has been little investigation into the robustness of the assumption of independence between individual capture probabilities. While this could introduce bias for certain animal species which exhibit either communal or territorial behavior, to date nearly all SCR models have assumed the location and detection of individuals to be completely independent from one another. In a notable exception, Reich and Gardner [2014] modeled activity centers according to a Strauss process (Zimmerman et al. [2010], Strauss [1975]) to account for avoidance between individuals due to territorial behavior. Additionally, Reich and Gardner demonstrated through simulations that ignoring this dependence introduced greater bias for their population size estimate. Royle et. al. have also noted modeling these types of interactions as a possible future direction for SCR modeling (Royle et al. [2018]).

In this paper we introduce a SCR model which incorporates attractions between individuals through their movement. By allowing for an individual's daily location to vary between it's own activity center and that of nearby activity centers, our model includes dependency among capture probabilities. Our method varies from Reich and Gardner [2014] not only in that we are considering attraction as apposed to avoidance but also that we are considering movement and attractions which can update daily as apposed to including the dependence through fixed activity centers. This model could be more appropriate for certain species with complex territorial and behavioral

dynamics such as tigers, which despite being a mostly solitary species have been known to have overlapping territories in which individuals are aware of one another's movement (McDougal [1977]) and even share kills on occasion (Schaller [1984]).

We introduce the model and discuss parameter estimation in Section 2.2. In Section 2.3 we present a simulation study to demonstrate that accounting for dependence between animal movement can improve parameter estimation. Additionally in this same Section we discuss possible outcomes of applying the model to data when attractions are not present (i.e. individuals behave and are captured independently from one another). In Section 2.4 we present analysis of a real world dataset using the newly developed attraction model. We conclude with a discussion in section 2.5.

2.2 A Spatial Capture-Recapture Model with Attractions

The model presented here, which we will hence for refer to as the categorical attraction (CA) model, is an extension (to include our individual movement scheme) of the model referred to in Royle et al. [2009a] as the 'Bernoulli encounter' model, hence forth referred to as the independent Bernoulli (IB) model. This particular model was chosen as it is an iconic model in SCR, being one of the earliest SCR models to analyze camera trapping data; and it presents a general and easily extendable framework for parameter estimation. Additionally, there have been advancements in technology and modeling

techniques which continue to make proximity trapping a popular choice for SCR studies; such as hair snares (Gardner et al. [2010b]), scat surveys (Morin et al. [2016], Sollmann et al. [2013b]) and acoustic surveys (Stevenson et al. [2015], Dawson and Efford [2009]).

2.2.1 Model formulation

We will assume N individuals are exposed to a trapping array of K traps located at \mathbf{x}_k during a period of T days. Each individual has a respective activity center $\{\mathbf{s}_i, i = 1, \dots, N\}$ which we consider to be uniformly distributed over the trapping area containing the K traps (with a small buffer area to ensure each individual and its movement is contained within this area). The i^{th} individual's location for day t , denoted \mathbf{u}_{it} , will follow

$$\mathbf{u}_{it} \sim \text{Categorical}(\pi_{i1t}, \pi_{i2t}, \dots, \pi_{iNt}),$$

where π_{ijt} denotes the probability tiger i 's daily location is centered at j^{th} activity center $\{\mathbf{s}_j; i = 1, \dots, N\}$ on day t (i.e. tiger i is attracted to tiger j on day t) and is defined

$$\pi_{ijt} = \begin{cases} 1 & \alpha_{it} = 1, j = i \\ \frac{\exp(-d(\mathbf{s}_i, \mathbf{s}_j)^2/2\sigma_\pi^2)}{\sum_{l \neq i} \exp(-d(\mathbf{s}_i, \mathbf{s}_l)^2/2\sigma_\pi^2)} & \alpha_{it} = 0, j \neq i \\ 0 & o.w. \end{cases}$$

with $d(\mathbf{s}_i, \mathbf{s}_j)$ denoting the Euclidean distance between the activity centers of individual i and individual j , and σ_π being a scale parameter dictating the range of attractions. We assume $\alpha_{it} \sim \text{Bernoulli}(\psi_\alpha)$ where ψ_α represents the probability an individual remains at its own activity center on a given day.

Assuming Y_{itk} denotes an indicator of observing the i^{th} individual on the t^{th} day at the k^{th} trap we let

$$Y_{itk} \sim \text{Bernoulli}(\mu_{itk})$$

with

$$\mu_{itk} = P(Y_{itk} = 1) = 1 - \exp \left\{ -\lambda \cdot \exp \left(-\frac{d(\mathbf{u}_{it}, \mathbf{x}_k)^2}{2\sigma^2} \right) \right\}$$

and using the complimentary log-log link we have the following relationship

$$\text{cloglog}(\mu_{itk}) = \log(\lambda) - \frac{d(\mathbf{u}_{it}, \mathbf{x}_k)^2}{2\sigma^2}.$$

2.2.2 Estimation

To combat the issue of N being unknown we will utilize a data augmentation approach (Royle et al. [2007], Royle et al. [2009a]). This approach involves introducing $M - n$ individuals to our dataset, all with zero capture histories, where n is the number of individuals which have been observed and M is chosen large enough so as not to underestimate N . We then introduce a set of latent indicator variables $\{w_i, i = 1, \dots, M\}$ where $w_i = 1$ if the i^{th} individual is a member of the true population of size N and $w_i = 0$

otherwise. In this case, where N is assumed unknown, we will now be modeling

$$Y_{itk} \sim \text{Bernoulli}(w_i \cdot \mu_{itk})$$

assuming

$$w_i \sim \text{Bernoulli}(\psi)$$

$$\psi \sim \text{Uniform}(n/M, 1)$$

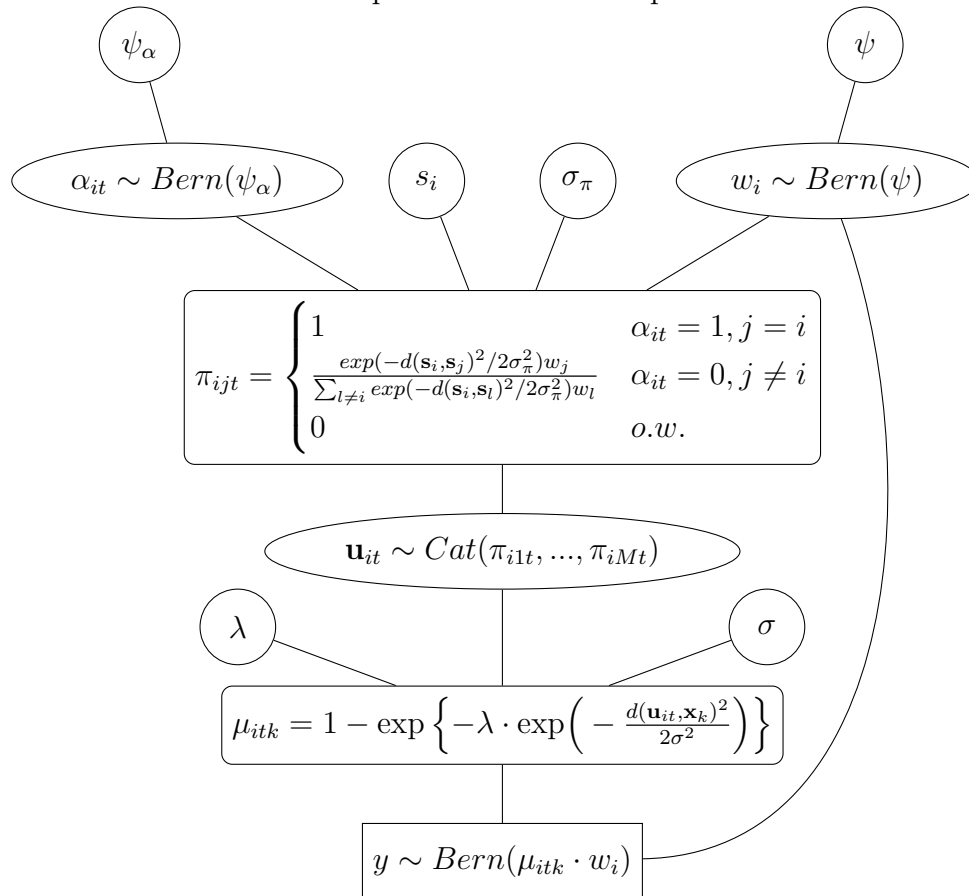
where ψ represents the probability an individual from the augmented population of size M belongs to the true population of size N . We then estimate N as $\hat{N} = \sum_{i=1}^M \hat{w}_i$. Additionally, we will also redefine $\{\pi_{ijt}; i = 1, \dots, M\}$ as follows

$$\pi_{ijt} = \begin{cases} 1 & \alpha_{it} = 1, j = i \\ \frac{\exp(-d(\mathbf{s}_i, \mathbf{s}_j)^2 / 2\sigma_\pi^2) w_j}{\sum_{l \neq i} \exp(-d(\mathbf{s}_i, \mathbf{s}_l)^2 / 2\sigma_\pi^2) w_l} & \alpha_{it} = 0, j \neq i \\ 0 & o.w. \end{cases} \quad (2.1)$$

The inclusion of the w_j term ensures the probability of attraction to individuals with $w_j = 0$ (i.e. augmented individuals which are assumed to not exist for the current chain step) will always be zero. A graphical representation of the model is given in Figure 1.

For parameter estimation we have opted for the Metropolis-Hastings algorithm as opposed to Gibbs sampling for efficiency reasons. As the CA model is now indexed in

Figure 1: Graphical representation of estimation model, ellipses represent parameters, rectangles with rounded corners represent functions of parameters



terms of days, vectorizing code in R helps reduce run times to speeds which can not be achieved with the simpler coding languages available in packages such as OpenBUGS or JAGS.

In practice we also assume $\alpha_{it} = 1$ for unobserved individuals and fix their activity centers. As unobserved individuals are interchangeable this will have little to no effect on parameter estimates even for the IB model assuming M is sufficiently large (this fact was validated with simulations) and will improve the estimate of σ_π for the CA model. An alternative would be to propose the unobserved centers according to a random walk with a small scale factor, however simulations suggest fixed centers produced better parameter estimates. For a more detailed description of the estimation and R code see Appendix B. Additionally we assume $\sigma_\pi > \sigma$ as σ_π and σ would most likely not be identifiable otherwise.

2.3 Simulation Study

A simulation study was conducted to evaluate the performance of the CA model and also compare this to the performance of the IB model. While the IB model is fairly robust to animal movement and activity center dispersion (as discussed in Royle et al. [2016]) the results of this simulation study suggest accounting for the dependence between animal movement can reduce bias and mean squared errors (MSE) for the estimation of N and to a much greater extent for the estimation of σ and λ . The model performance was

also evaluated when being applied to data simulated under the assumptions of the IB model (*i.e.* $\sigma_\pi = 0$ or $\psi_\alpha = 1$) to determine the effects of this misspecification on the model parameters.

2.3.1 CA vs. IB model comparison

Both models were run on data simulated under the assumptions of the CA model for varying $\sigma_\pi = \{6, 9, 12\}$ and with parameter values ($N = 100, M = 300, \sigma = 2, \lambda = 0.1, \psi_\alpha = 0.75$) in a 40x40km trapping area with 100 traps. The models were run on 100 data sets for each value of σ_π . The averages of the MSE's, biases, 90% interval width and the 90% coverages for the two model's estimates of N are given in Table 1. For the CA model the median of N 's posterior distribution was chosen as the estimator for N as it displayed marginally better MSE and bias compared to the mean. The results suggest the CA model's estimate of N exhibits lower MSE's and biases relative to the IB model as σ_π increases. Side-by-side boxplot comparisons of the two models' MSE and bias for the estimate N are given in Figure 2 and Figure 3 respectively.

The improvement in MSE and bias for the CA model can be credited to how well it is estimating σ and λ . Figures 4 and 5 illustrate that the IB model's estimate of σ becomes inflated for increasing σ_π , resulting in an underestimation of λ which in general leads to an over estimate of N .

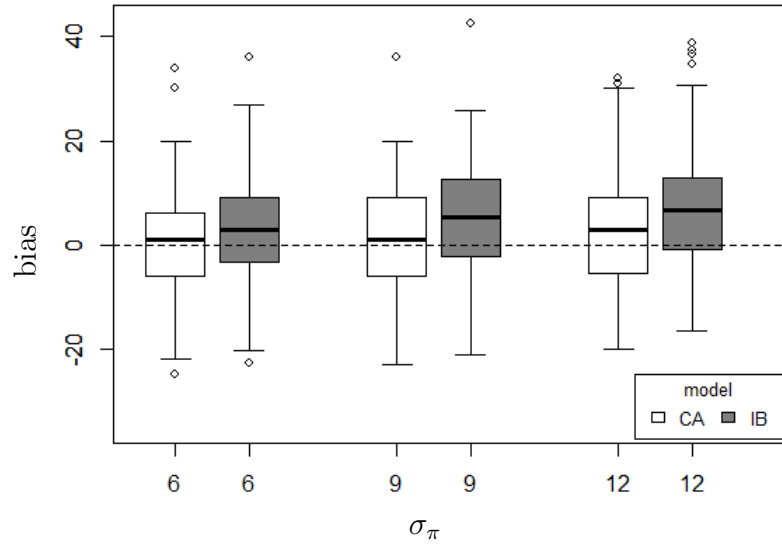
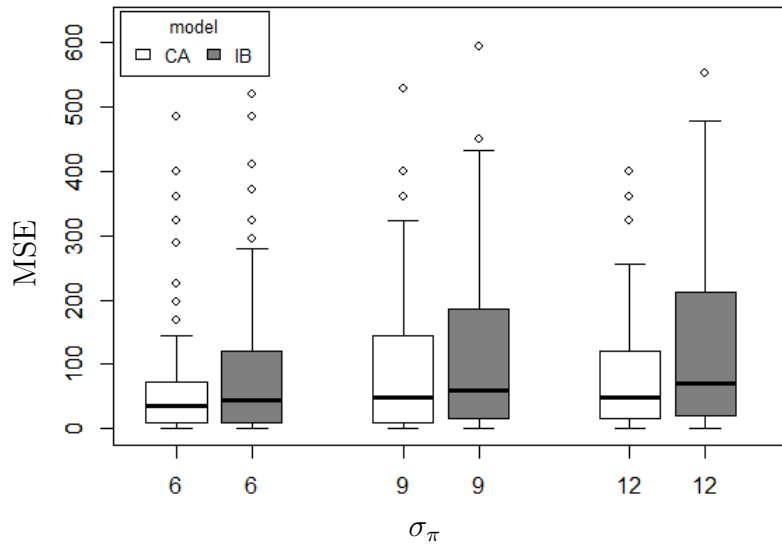
Figure 2: Side-by-side boxplots of the bias for the estimate of N by σ_π Figure 3: Side-by-side boxplots of the MSE for the estimate of N by σ_π 

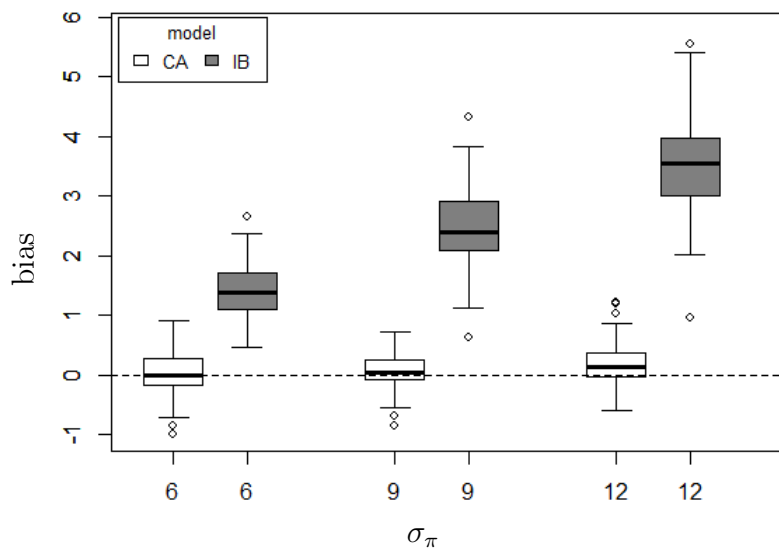
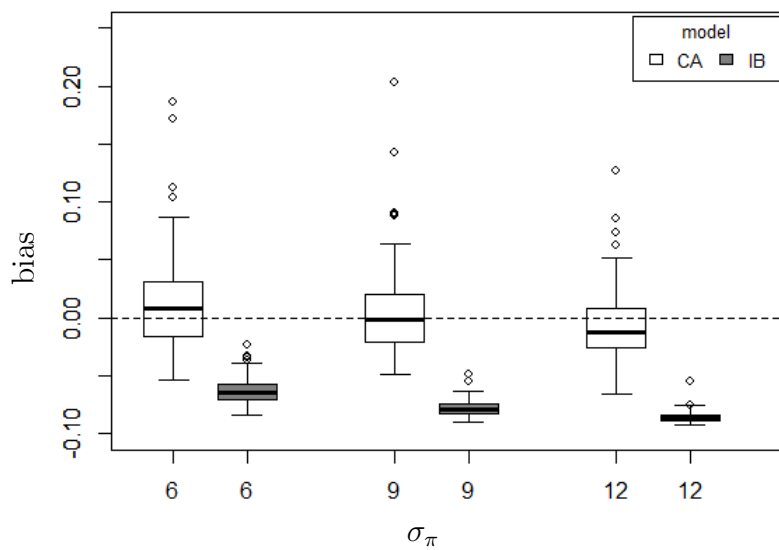
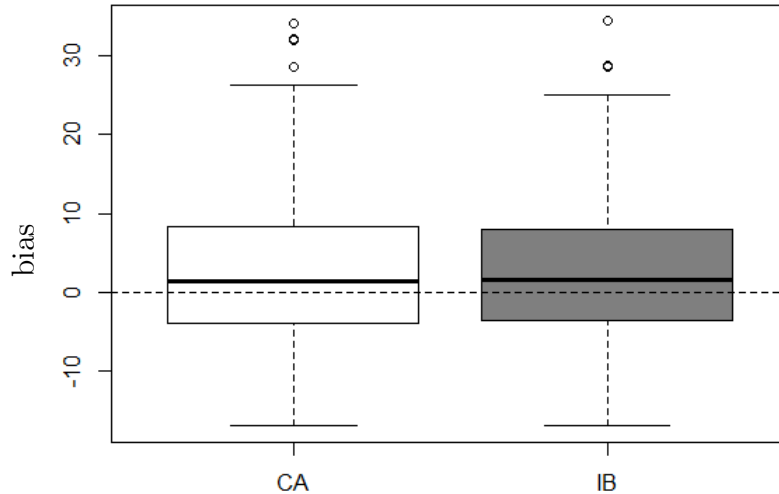
Figure 4: Side-by-side boxplots of σ estimate bias by σ_π Figure 5: Side-by-side boxplots of λ estimate bias by σ_π 

Table 1: Model comparison of \hat{N} while varying σ_π

Model	σ_π	MSE	Bias	Coverage	Width
CA	6	86.73	0.37	0.92	30.34
IB		97.88	2.99	0.92	32.06
CA	9	101.05	1.17	0.84	39.28
IB		130.26	4.92	0.91	32.90
CA	12	119.34	2.38	0.85	29.76
IB		178.82	6.94	0.85	34.65

2.3.2 CA model performance when $\psi_\alpha = 1$

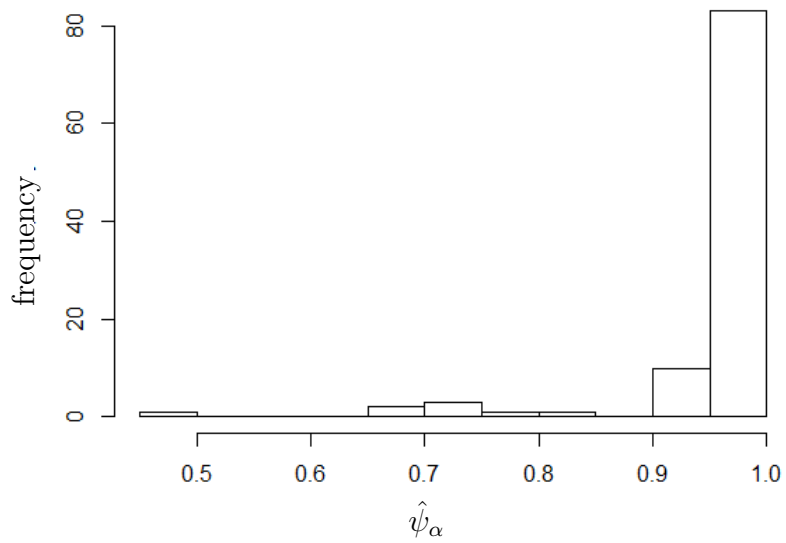
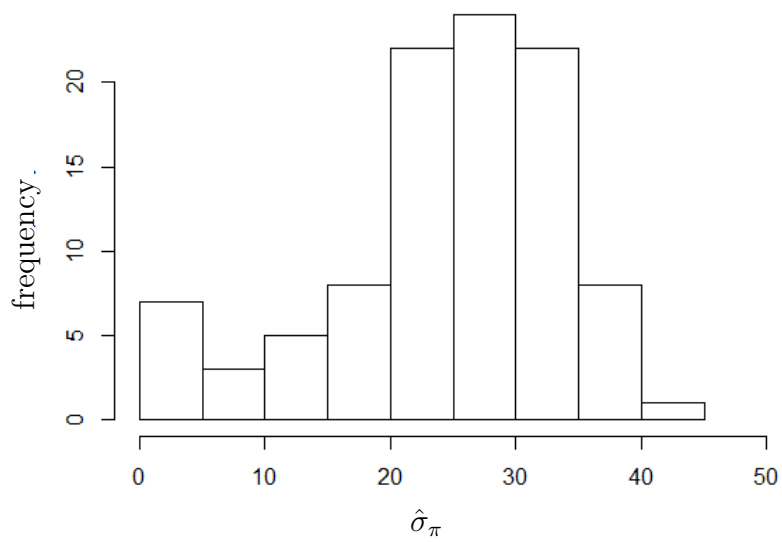
Under the assumption that $\psi_\alpha = 1$ (i.e. all individuals will remain at their own centers with probability 1) the CA model will reduce to the IB model. To test the CA model under this misspecification it was fit to 100 data sets simulated under the assumptions of the IB model (i.e. without daily location changes or attractions among individuals). The results suggest the CA model’s estimate of N is very robust in this situation with little to no reduction in accuracy compared to the IB model, as illustrated by Figure 6. In general, under these conditions ψ_α was estimated to be very close to 1 (as seen in Figure 7), which leads to α_{it} nearly always being 1 (under the assumption $\alpha_{it} \sim \text{Bernoulli}(\psi)$). Since σ_π only enters into the model through Equation 2.1, when nearly all of the $\alpha_{it} = 1$ this will essentially remove σ_π from the model. This leads to inaccurate estimates of σ_π as is illustrated by Figure 8 where σ_π ’s estimated posterior distribution is centered halfway between the minimum and maximum values allowed for σ_π ’s estimate (in this case these values were 1.1 and 56.6, Section B.2 of Appendix B details how these endpoints were

Figure 6: Side-by-side boxplots of N estimate bias by model with $\psi_\alpha = 1$ 

chosen). As a rule of thumb it may be safer to ignore the estimate of σ_π when the estimate of ψ_α exceeds 0.9, or consider using the results from the IB model instead. The results may also be less reliable if $\hat{\psi}_\alpha$ were smaller than around 0.5 (as this may suggest the concept of an activity center is unrealistic).

2.4 Analysis of Tiger Data

The CA model was fit to camera trapping data of a Bengal tiger (*Panthera tigris tigris*) population located in the Nagarahole reserve in the state of Karnataka in southwestern India. This data has previously been modeled with several different SCR approaches (Royle et al. [2009a], Royle et al. [2009b], Royle and Dorazio [2012], Dorazio [2013]). 120 sites were sampled by transitioning 30 cameras over four trapping areas with each area being active for 12 consecutive days. This resulted in the capture of 44 unique

Figure 7: Histogram of ψ_α estimate with $\psi_\alpha = 1$ Figure 8: Histogram of σ_π estimate with $\psi_\alpha = 1$ 

individuals, 16 of which were recaptured at least once. The data analyzed here had been collapsed down to binary data allowing for only one capture at each location per day for a particular individual (regardless of whether that individual was caught multiple times in a single day at that location).

Table 2 and Figure 10 give population size estimates for both models. The estimate of σ_π seems significantly greater than σ and the estimate of ψ_α seems enough below 1 to conclude there may be attraction among tigers (density estimates for these parameters are shown in Figures 11 and 12 respectively). The CA model has produced smaller estimates of σ and N as compared to the IB model as can be seen in Figure 9 and Figure 10 respectively. These results are consistent with the simulation study discussed in Section 2.3.1, which demonstrated the IB model tended to over estimate N in the presence of attraction between individuals. It is also worth noting that the 95% interval estimate of N from the CA model is contained within the IB model's interval estimate (see Table 2).

For the scale parameter σ and σ_π a Highest Posterior Density (HPD) interval may be preferred over the equal tailed intervals presented in table 2. The HPD intervals for σ were given by (0.92,1.65) and (1.52,2.41) for the CA and IB models respectively. The HPD interval for the CA model's σ_π parameter was essentially identical to the symmetric interval and was given by (4.28,9.74). These intervals were calculated using the `HPDinterval` function in the `coda` R package.

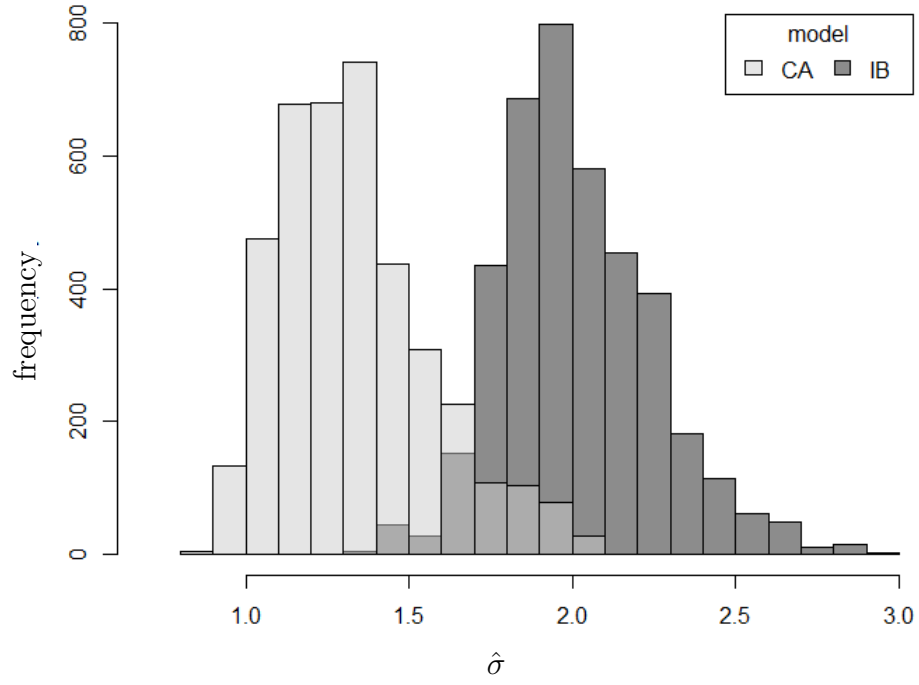
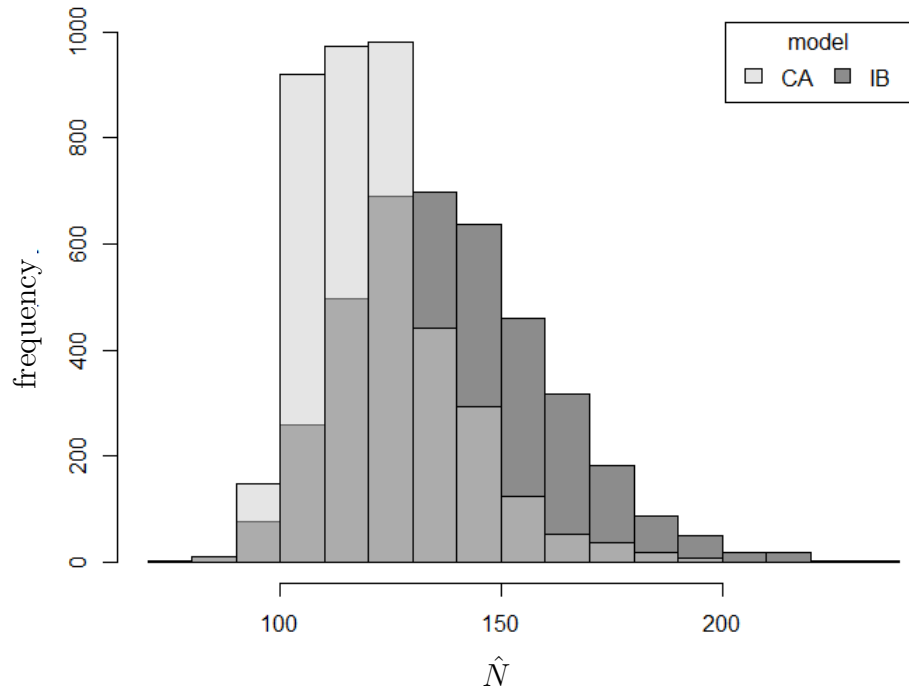
Figure 9: Posterior densities of σ from CA and IB modelsFigure 10: Posterior densities of N from CA and IB models

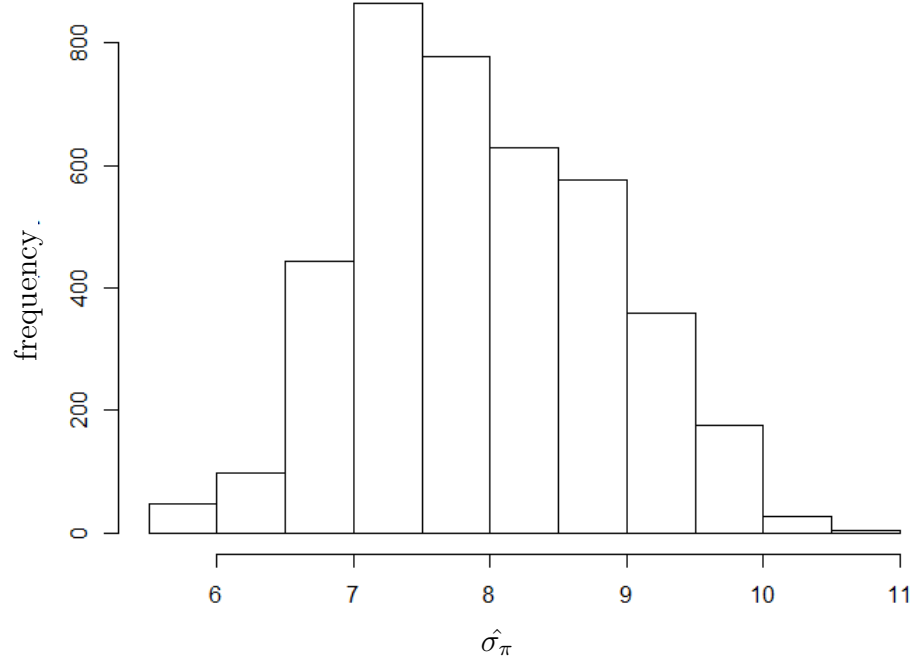
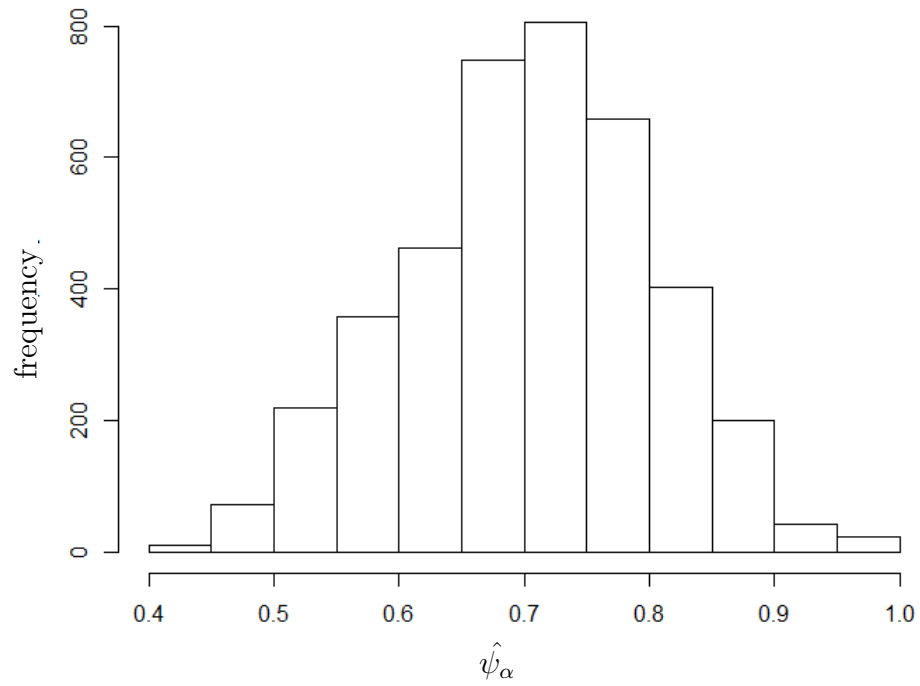
Figure 11: Posterior density of σ_π from CA modelFigure 12: Posterior densities of ψ_α from CA model

Table 2: Parameter estimates from tiger data for CA and IB models

Parameter	mean	sd	2.5%	50%	97.5%
CA model					
σ_π	7.93	0.91	4.30	7.047	9.75
ψ_α	0.703	0.099	0.44	0.66	0.85
σ	1.33	0.23	0.96	1.21	1.77
λ	0.044	0.018	0.02	0.042	0.078
ψ	0.41	0.061	0.33	0.44	0.6
N	122.17	16.38	102	132	175
IB model					
σ	1.98	0.24	1.58	1.93	2.48
λ	0.016	0.0045	0.009	0.016	0.026
ψ	0.46	0.081	0.33	0.47	0.67
N	139.38	23.16	100	139	199

2.5 Discussion

Here we have presented a SCR model which allows for attractions between individuals. A simulation study has demonstrated our model produces estimates with smaller bias and MSE than the independent model in the presence of dependence between animal movement. Applying this now model in a real world example to camera trapping data from a tiger population gave results consistent with the simulation study. These results suggest not only that there may be attraction between the tigers, but additionally that ignoring these attractions may be resulting in an over estimation of the population size.

As noted by Reich and Gardner [2014], another advantage to these types of models is the ability to investigate components of the animal's behavior without collecting more data. Where as more time intensive or invasive procedures may have been required

previously, our model may now be able to aid researchers in determining if there is attraction among individuals when only CR data is available. Additionally, the hierarchical design of our model (in that the u_{it} 's will be conditionally independent given the activity centers across individuals) should make extending this movement scheme to other SCR models fairly straightforward (in particular models which also rely upon data augmentation). Already with just the flexibility of the IB model to work from a parameter such as an animal's sex could easily be added to the model, which could be of particular interest as tiger's behavior and movement vary between the sexes.

Modeling these types of dependencies between individuals is still in it's infancy and the proposed model could suggest multiple future lines of research. Considering the individual's daily locations as a weighted function of it's own activity center, geographical points of interest and neighbor's activity centers could be an interesting extension to the CA model. Extending this model to handle data which includes individual's time of arrival (similar to Schofield et al. [2018], Distiller and Borchers [2015], Borchers et al. [2014]) could also prove fruitful. The requirement that σ_π be larger than σ may be seen as a drawback of the CA model by some researchers, determining a method to avoid this assumption could open up this type of attraction model to more animal species.

Chapter 3

Spatial Capture-Recapture via Reversible Jump Markov Chain Monte Carlo

3.1 Introduction

One of the most difficult challenges in the analysis of capture-recapture (CA) data is modeling individual heterogeneity. While covariates and random effects associated with individuals are often of great interest to researchers, including such parameters in the model can lead to a number of intractable problems as they are indexed in terms of the unknown population size. To overcome these estimation challenges researchers have proposed many approaches including jackknife procedures (Burnham and Overton [1978], Pollock and Otto [1983]), bootstrap estimators (Smith and van Belle [1984]), maximum likelihood estimators (MLE) (Pledger [2000], Efford et al. [2009b], Borchers and Efford [2008]), non-parametric MLE (Norris and Pollock [1995], Norris and Pollock

[1996]) and numerous Bayesian methods (such as Bonner and Schofield [2014], Royle [2009], Fienberg et al. [1999]).

Reversible jump Markov chain Monte Carlo (RJMCMC) methods, which allow for transitions between models with parameter vectors of varying dimension during Markov chain Monte Carlo estimation (MCMC), have also been considered for fitting these heterogeneous models. Durban and Elston [2005] applied the RJMCMC algorithm of Carlin and Chib [1995] to obtain abundance estimates assuming capture probability could vary across time and individuals. Schofield and Barker [2014] have extended the approach of Durban and Elston [2005] to be implementable in popular Bayesian software such as OpenBUGS or JAGS. King and Brooks [2008] applied the RJMCMC algorithm of Green [1995] to obtain abundance estimates and to perform model selection between the classical closed population CR models proposed by Otis et al. [1978].

Spatial capture-recapture (SCR) models, which consider capture probability as a function of an individual's distance from a given trap, fall within the category of individual heterogeneity models. In general, SCR models assume each individual possesses an activity center (a coordinate the individual will center their activity around during the experiment) and heterogeneity of capture probabilities arise as a juxtaposition of these centers and trapping devices.

While there is the MLE approach of Borchers and Efford [2008] and the more recent semi-complete likelihood approach by King et al. [2016], seemingly the most widely

utilized estimation method for SCR models is the data augmentation technique introduced by Royle and Young [2008]. The DA approach augments a list of n observed individuals' capture histories with $M - n$ unobserved individuals (where M is chosen to be larger than N is believed to be) and then introduces latent indicator variables w_1, \dots, w_M , where $w_i = 1$ if individual i belongs to the true population of size N and $w_i = 0$ otherwise. Assuming $w_i \sim \text{Bernoulli}(\psi)$, MCMC is utilized for parameter estimation and the goal of the inference is to estimate ψ instead of estimating N directly. The DA approach is straightforward to extend to nearly any CR model since unlike many other approaches (which require averaging out parameters or completion of intractable maximization problems) DA avoids working directly with a model's likelihood.

While DA offers an easy to use and general framework it still has drawbacks. As noted by King et al. [2016], there can be a difficult balance between choosing an M which is small enough for computational efficiency yet large enough to ensure N is not underestimated. Royle et al. [2013a] have suggested a more formal analysis for the choice of M is needed, as this is currently done by trial and error. Schofield and Barker [2014] have also taken issue with DA for estimating N as a derived parameter. The DA approach assumes $N = \sum_{i=1}^M w_i$, which induces a discrete uniform prior over $[0, M]$ on N and does not allow for any alternative hierarchical interpretations (King and Brooks [2008] have referenced occasions where informative priors can be constructed for the population size and suggested the use of a Negative-Binomial distribution).

In a recent review of SCR Borchers et al. [2016] have noted that there has yet to

be a formal extension of a RJMCMC method for SCR modeling. To fill this gap, in what follows we extend the RJMCMC approach presented by King and Brooks [2008] to a SCR model. This approach should be easily extendable to any model where DA has been applied and an explicit likelihood containing N can be specified, yet is more efficient as it does not require an upper bound to be set for N . Additionally, unlike DA, this approach includes N within the model's likelihood and thus allows for hierarchical modeling of N .

We introduce the model and discuss the RJMCMC approach taken for parameter estimation in Section 3.2. In Section 3.3 we present a simulation study to demonstrate the RJMCMC model and DA model produce similar parameter estimates and argue the RJMCMC model is considerable more efficient than the DA model. In Section 3.4 we present analysis of a real world dataset using the RJMCMC model. We conclude with a discussion in section 3.5.

3.2 Model

The model presented here (hence forth referred to as the RJMCMC model) is a generalization of the CR model and RJMCMC estimation method given by King and Brooks [2008]. This current model aims to explain heterogeneity of capture probabilities through individual activity centers and thus the likelihood presented by King and Brooks [2008]

must be formulated to include this component. The model component relating to capture probability has been developed to match that of the DA model presented by Royle et al. [2009a] to facilitate a comparison between the DA method and the RJMCMC method considered here.

3.2.1 Model formulation

We will assume N individuals are exposed to an array of K traps located at \mathbf{x}_k during a period of T days. Each individual has a respective activity center $\{\mathbf{s}_i, i = 1, \dots, N\}$ which we consider to be uniformly distributed over the trapping area containing the K traps (with a small buffer area to ensure all catchable activity centers are included).

Assuming Y_{itk} denotes an indicator of observing the i^{th} individual on the t^{th} day at the k^{th} trap, we let

$$Y_{itk} \sim \text{Bernoulli}(\mu_{ik})$$

with

$$\mu_{ik} = P(Y_{itk} = 1) = 1 - \exp \left\{ -\lambda \cdot \exp \left(-\frac{d(\mathbf{s}_i, \mathbf{x}_k)^2}{2\sigma^2} \right) \right\} \quad (3.1)$$

where $d(\mathbf{s}_i, \mathbf{x}_k)$ denotes the euclidean distance between the i^{th} individual's activity center (\mathbf{s}_i) and the k^{th} trap located at \mathbf{x}_k . Using the complementary log-log link in keeping with Royle et al. [2009a] we have the following relationship

$$\text{cloglog}(\mu_{ik}) = \log(\lambda) - \frac{d(\mathbf{s}_i, \mathbf{x}_k)^2}{2\sigma^2}.$$

When N is assumed unknown this model formulation leads to the following likelihood

$$L(N, \lambda, \sigma, \mathbf{S}|\mathbf{y}) = \binom{N}{n} \prod_{i=1}^N \prod_{t=1}^T \prod_{k=1}^K \mu_{ik}^{y_{itk}} (1 - \mu_{ik})^{1-y_{itk}}$$

where n is the number of observed individuals, \mathbf{S} is a matrix of individual activity centers and \mathbf{y} is an $(N \times T \times K)$ matrix containing capture histories for each individual on each day at each trap.

3.2.2 Estimation

The model parameters will be estimated using the RJMCMC method presented by King and Brooks [2008] which is a generalization of the RJMCMC approach introduced by Green [1995]. The parameters σ, λ and \mathbf{s}_i are updated according to the standard Metropolis-Hastings algorithm with new values being proposed via uniform random walks. In keeping with King and Brooks [2008] N is updated as follows

1. A new N (N') is proposed according to a discrete uniform distribution on the interval $[\max(N - \delta_N, n), N + \delta_N]$ where δ_N is a scale tuning parameter chosen to ensure proper mixing.
2. If $N' > N$ then $N' - N$ new centers ($\mathbf{s}'_{N+1}, \dots, \mathbf{s}'_{N'}$) are drawn uniformly over the trapping area and a proposal list of centers is defined $\mathbf{S}' = (\mathbf{s}_1, \dots, \mathbf{s}_N, \mathbf{s}'_{N+1}, \dots, \mathbf{s}'_{N'})^1$.

If $N' < N$ then a new truncated proposal list of centers is defined as $\mathbf{S}' =$

¹ \mathbf{s}_i denotes a previously accepted activity center

$(\mathbf{s}_1, \dots, \mathbf{s}_{N'})$.

3. The general Metropolis-Hastings update is performed with (N, \mathbf{S}) vs. (N', \mathbf{S}') .

This step is simplified compared to the more general case presented in the supplementary material for King and Brooks [2008] due to the fact that the individual effects (the \mathbf{s}_i for this model) are proposed according to a uniform distribution.

For this study a Jeffreys prior proportional to N^{-1} was considered for N (in keeping with King and Brooks [2008]), and a uniform prior over the trapping area for the \mathbf{s}_i and improper uniform priors over the positive real line for σ and λ (in keeping with Royle et al. [2009a]).

3.3 Simulation Study

A simulation study was performed to evaluate the performance of the RJMCMC model and compare this to the performance of the DA model. 200 datasets were simulated with a 40km by 40km trapping area containing 100 traps which were assumed active for 10 trapping periods. Activity centers were drawn uniformly for $N = 80$ individuals over the trapping area plus a buffer area of 4km (10% of the trapping area width). Captures for each individual (i) at each trap (j) were drawn for each day according to a Bernoulli(μ_{ij}) distribution, where μ_{ij} was defined as in Equation 3.1 with parameter values set at ($\lambda = 0.1$ and $\sigma = 2$). The RJMCMC model and equivalent DA SCR model (the Bernoulli model introduced by Royle et al. [2009a] which is implemented in the

`scrbook` R package) were then fit to this simulated data.

3.3.1 Efficiency of RJMCMC vs. data augmentation

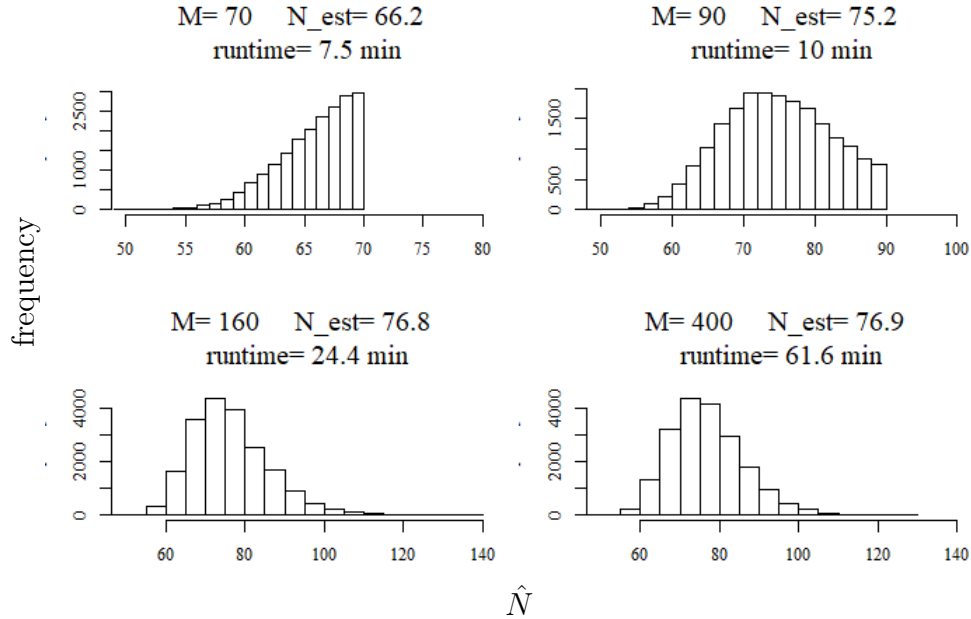
To compare the efficiency of the RJMCMC model to that of the DA model, a method was needed for determining the optimal value of M (the smallest M which does not lead to N being underestimated) for DA for a given dataset. In practice, M is usually chosen to be substantially larger than a researcher may believe the true population size to be (for example see Royle and Dorazio [2012]) since if M is chosen to be too small this can truncate the estimated posterior distribution of N , leading to N being underestimated. Choosing too large of an M in general should not effect parameter estimation but will decrease computational efficiency.

An illustration of the effects M has on the estimation of N and run times is given in Figure 13 which shows histograms of N 's estimated posterior distribution for the same dataset for increasing values of M . Truncation can be seen for the smaller values of M (70 and 90) and runtimes are increasing with M (in general runtimes increase linearly with M). Also note that the estimate of N is essentially unaffected by even a substantial increase in M (here from 160 to 400) after a sufficiently large M has been reached.

As the mode of N 's posterior distribution will vary among datasets² this means the minimum value of M needed for the estimate of N to converge will vary among datasets as well. This is illustrated in Figure 14 where $M = 110$ is a reasonable choice for the

²the mode of N 's estimated posterior distribution will vary among datasets as it depends upon n and the number of recaptures

Figure 13: Histograms of N 's estimated posterior distribution using data augmentation with increasing M

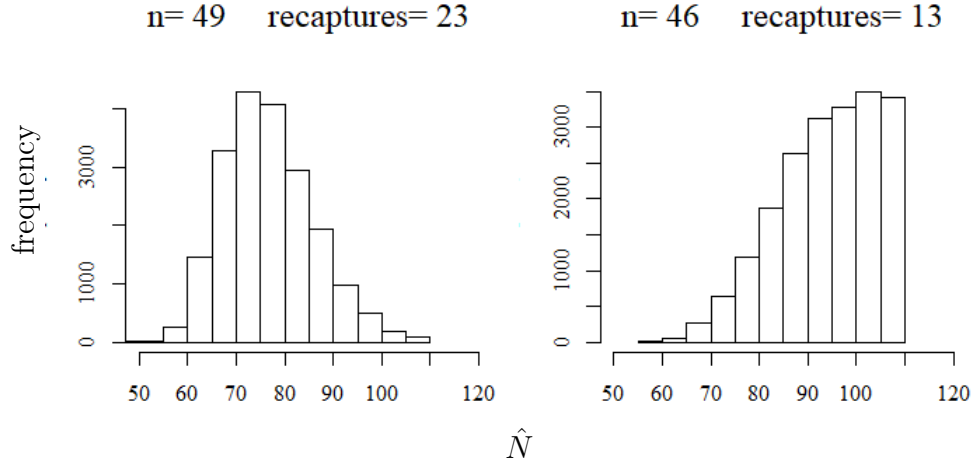


first dataset (which has a large number of recaptures relative to n) but N 's estimated posterior is heavily truncated for the second dataset (which has a small number of recaptures relative to n) for this same value of M .

A method was established for picking M for a given dataset based around the first value of M to result in N 's estimated posterior distribution having positive skew³, denoted as $M^{(+skew)}$. The DA model was fit to the 200 simulated datasets for values of M between 80 to 400 increasing M by increments of 10 and the adjusted Fisher-Pearson coefficient of skewness was then calculated for each of the resulting estimated posteriors for N . Figure 15 shows histograms of N 's posterior distribution for increasing values

³Skewness is a relevant diagnostic tool here as N 's estimated posterior distribution will always display truncation when it is negatively skewed

Figure 14: Histograms of N 's estimated posterior for two datasets with the same $M(=110)$

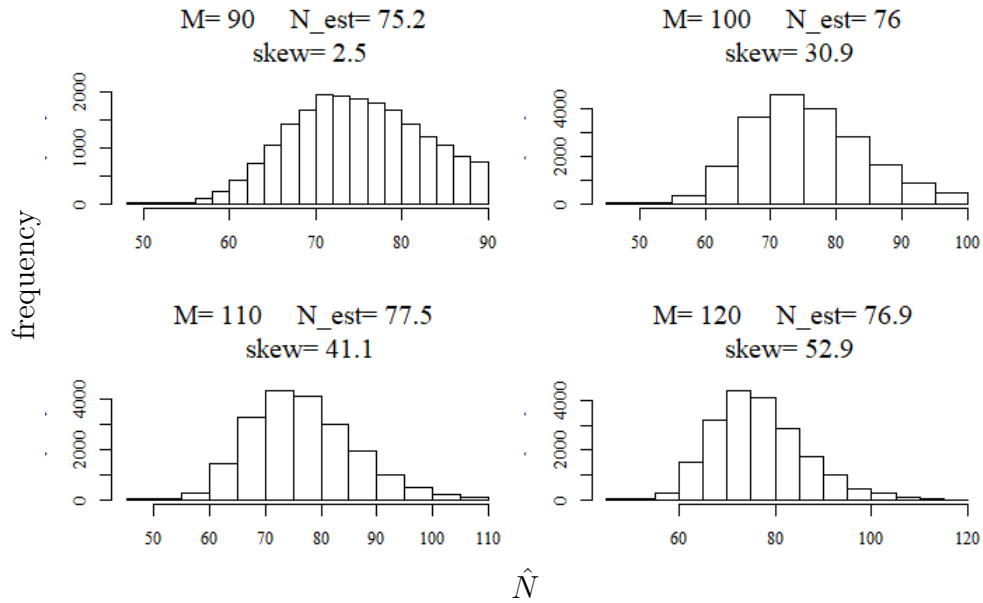


of M for a single dataset. For this particular dataset, $M^{(+skew)} = 90$, however there is still clearly truncation being induced. Also note that $M = 120$ (which corresponds to $M^{(+skew)} + 30$ in this case) appears to be the first value of M where N 's posterior has not reached M .

$M = M^{(+skew)} + 30$ was chosen as the most efficient value of M since in general the estimate of N seems to converge at or around this point. This is illustrated in Figure 16 which shows side-by-side boxplots of the difference between estimates of N using $M = 400^4$ vs. estimates of N for increasing values of M starting from $M^{(+skew)}$. These \hat{N} differences, which seem to have converged to zero for $M \geq M^{(+skew)} + 30$ in this plot, can work as a measure of the convergence of the estimate of N (with values around zero indicating the estimate of N will not continue to increase with M , i.e. the estimate of

⁴ $M = 400$ was chosen as it is a large enough value that N 's estimated posterior distribution did not display clipping for any of the 200 simulated datasets

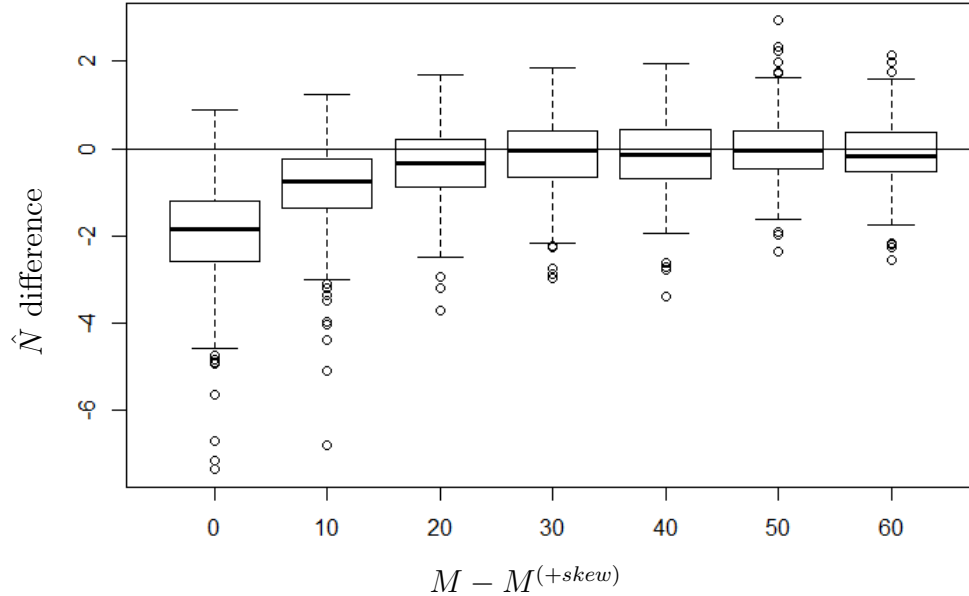
Figure 15: Histograms of N 's estimated posterior for increasing M and Fisher-Pearson coefficient of skewness



N is no longer being effected by the truncation of its estimated posterior distribution).

In practice, unless perhaps chains are run in parallel, using a large M will usually be a more efficient approach than spending the time and resources needed to determine $M^{(+skew)}$. The concept of using $M = M^{(+skew)} + 30$ is presented here to demonstrate the theoretically most efficient value of M for the purpose of comparison with the RJMCMC model. For the 200 simulated datasets, using $M = M^{(+skew)} + 30$ resulted in a mean M of 133.7. In terms of run times the RJMCMC model shows considerable improvement over the DA model. Figure 17 shows side-by-side boxplots of run times for analysis of the 200 simulated datasets using the RJMCMC model, the DA model with $M = M^{(+skew)} + 30$ and the DA model with $M = 400$ for equal length chains. On average the RJMCMC model ran in around half the time of the DA model utilizing the best theoretical values

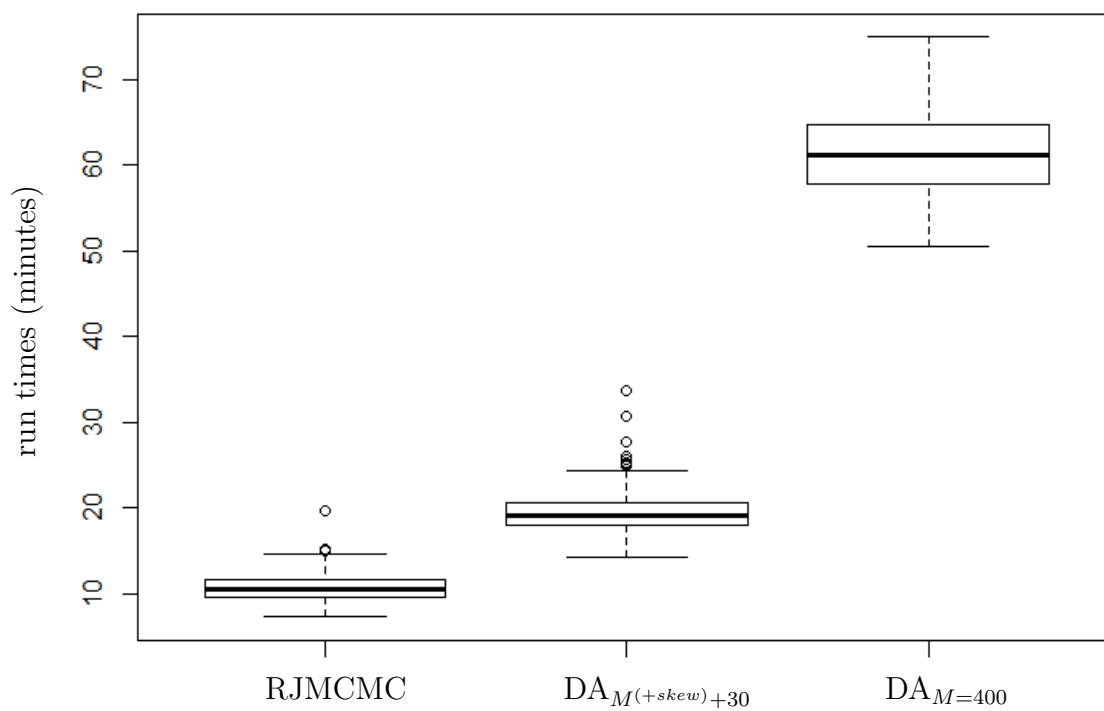
Figure 16: Side-by-side boxplots of the difference between \hat{N} with $M = 400$ and \hat{N} with increasing values of M starting from $M^{(+skew)}$ for 200 simulated datasets



of M and around six times faster than the DA model utilizing the more practical value of M ($M = 400$). The R code to fit these models was run on machines with two 4-core AMD Opteron 2350 processors (2 GHz) and 8 gigabytes of RAM.

It is also worth noting for the DA model the average largest value obtained by N 's estimated posterior distribution came out to be 132.3 while for the RJMCMC the average was 137.4. Despite the fact that no user input is needed to set an upper bound limit on the estimate of N when using the RJMCMC model, the largest values reached by N 's estimated posterior distribution have still remained only slightly higher as compared to those from the DA model utilizing the most efficient values of M .

Figure 17: Side-by-side boxplots of RJMCMC runtimes vs. DA runtimes using best M (i.e. $M = M^{(+skew)} + 30$) and DA using $M = 400$



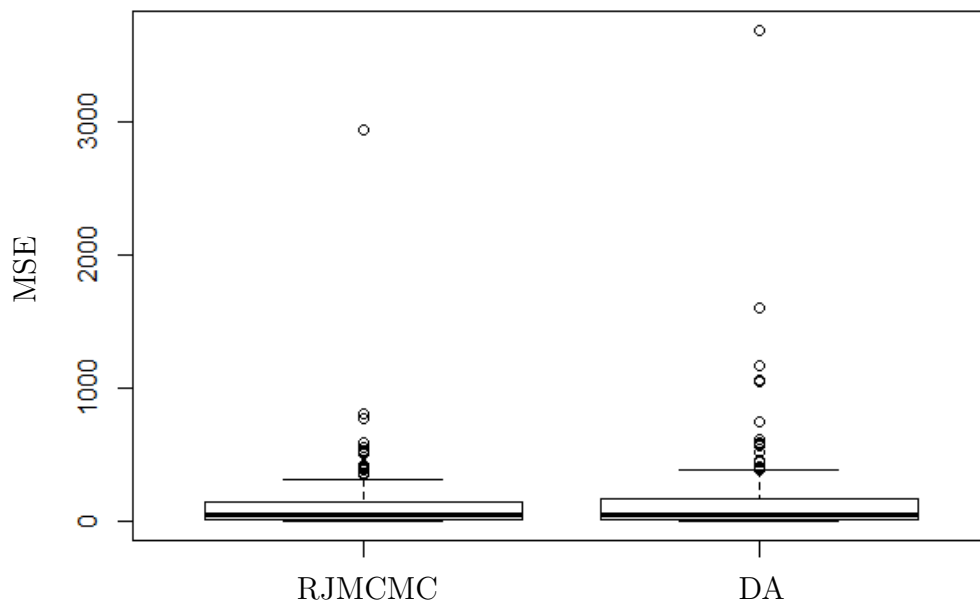
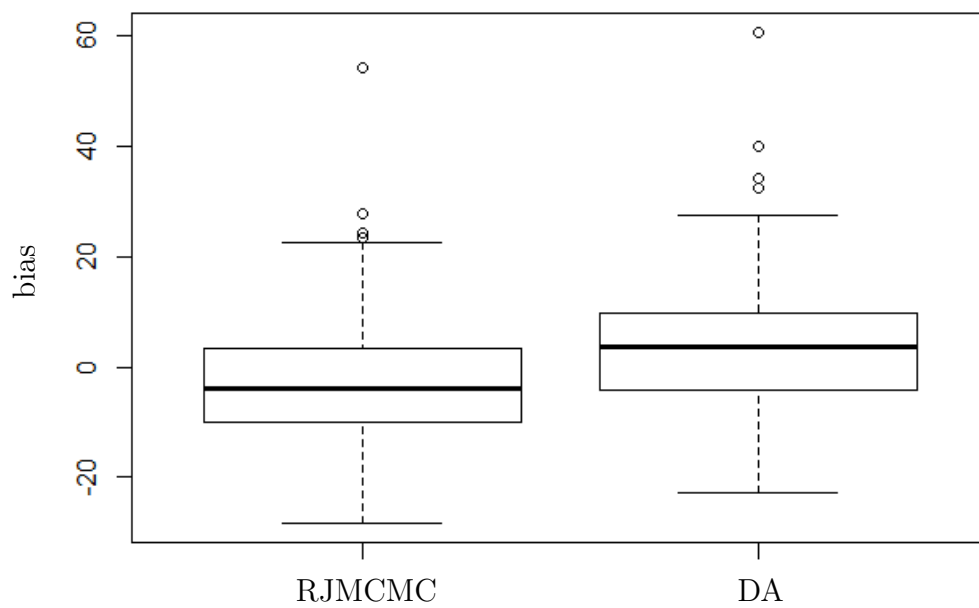
3.3.2 RJMCMC vs. data augmentation estimation

Both models were fit to the 200 simulated datasets with M set at 400 for the DA model. The mean of N 's estimated posterior distribution was used as an estimator of the population size. A comparison of the two models is given in Table 3 which contains the model's average mean squared error (MSE), average bias, 90% coverage and average 90% interval width for the estimation of N . These results suggest the models perform fairly similarly with the RJMCMC model performing marginally better in this example.

Table 3: Model comparison of \hat{N} for 200 simulated datasets

Model	Parameter	MSE	Bias	Coverage	Width
RJMCMC	N	122.5	-2.78	0.86	33.39
DA		151.9	4.17	0.9	38.23

These results are also illustrated in Figure 18 and Figure 19 which are side-by-side boxplots of the two models' MSE and biases respectively. Neither model has displayed a great deal of bias, the most noteworthy difference is the DA model's outlying MSE values are slightly more extreme than those of the RJMCMC model. The main take away is the fact that both models produce comparable results, while we have seen in Section 3.3.1 that the RJMCMC model has done so in considerable less time than the DA model.

Figure 18: Side-by-side boxplots of RJMCMC vs. DA MSE's for the estimation of N Figure 19: Side-by-side boxplots of RJMCMC vs. DA Biases for the estimation of N 

3.4 Black bear study example

The RJMCMC model was fit to hair snare SCR data of an American black bear (*Ursus americanus*) population located within the Fort Drum U.S. military reservation in up-state New York, USA. This data was originally published by Wegan [2008] and has been subsequently analyzed for several SCR papers (including Gardner et al. [2009] and Royle et al. [2013a]). Bears were uniquely identified by DNA analysis of their hair caught in baited barbwire traps. The experiment was conducted over an eight week period with 38 traps being checked once per week. In total there were 151 captures of 38 individuals, 28 of which had at least one recapture.

For simplicity the trapping area was considered the minimum rectangle needed to include the 38 trapping locations⁵ plus an extra buffer of 1.8km (10% of the width of the minimum area needed to include the traps). The DA model was also fit to the data using $M = 400$.

Table 4 gives the parameter estimates for both models. Again we note the runtime of the RJMCMC model (3.87 minutes) is considerably lower than that of the DA model (45.48 minutes). While the RJMCMC model has produced a larger estimate for N , the RJMCMC estimate for N is still contained within the 95% interval for the estimate of N given by the DA model and visa-versa. This is also illustrated in Figure 20 which shows overlapping histograms of N 's estimated posterior distributions from both models.

⁵Royle et al. [2013a] considered using a convex hull as a trapping border due to the asymmetric shape of the trapping locations

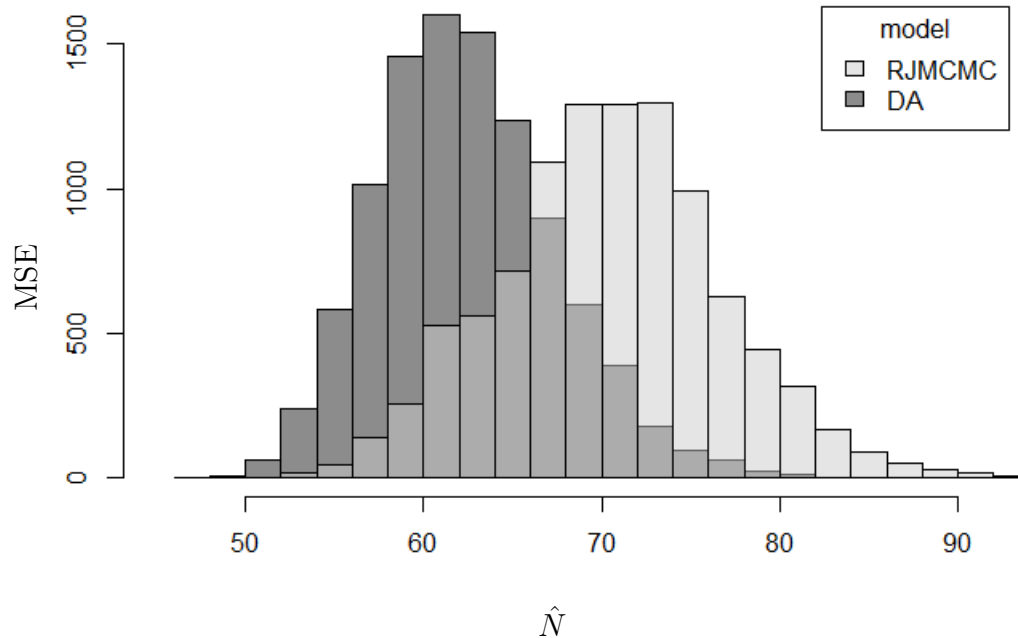
Table 4: Parameter estimates from bear data for RJMCMC and DA models

Parameter	mean	sd	2.5%	50%	97.5%
RJMCMC model					
σ	1.47	0.062	1.35	1.47	1.60
λ	0.13	0.015	0.11	0.13	0.17
N	71.06	6.36	60	72	87
runtime: 3.87 minutes					
DA model					
σ	1.94	0.12	1.74	1.95	2.24
λ	0.11	0.015	0.082	0.11	0.14
ψ	0.16	0.022	0.12	0.16	0.21
N	63.35	5.13	54	63	74
runtime: 45.48 minutes					

For the scale parameter σ a Highest Posterior Density (HPD) interval may be preferred over the equal tailed intervals presented in table 4. These HPD intervals were essentially identical to the symmetric intervals in this case and were given by (1.35,1.59) and (1.73,2.22) for the RJMCMC and DA models respectively. These intervals were calculated using the `HPDinterval` function in the `coda` R package.

Another desirable feature of the RJMCMC model visible here is the fact that it tends to produce more symmetric estimated posterior distributions for N as compared to the DA model utilizing a large M .

Figure 20: Overlapping histograms of estimated posterior distributions of N from the RJMCMC and DA models



3.5 Discussion

Here we have presented a seemingly long overdue approach for analyzing SCR data by way of RJMCMC. We have shown that for equivalent SCR models RJMCMC can be considerably more efficient than the commonly used DA approach, even when DA is performed under the most ideal circumstances (i.e. for the most efficient values of the upper bound parameter M). This increase in efficiency is especially encouraging since we have also demonstrated the RJMCMC and DA models produced similar results with both real and simulated data.

Royle et al. [2013a] argue RJMCMC is unnecessary as it can be more difficult to implement than DA and is thus less accessible especially to researchers from outside of

statistics. Schofield and Barker [2014] contend this is an unfounded characterization of RJMCMC approaches and also argue that the DA approach is simply one generalization of a RJMCMC algorithm⁶. With the approach presented here it could be maintained RJMCMC is in fact more user friendly than DA as it does not require a researcher to specify M and ensure N 's estimated posterior distribution has not been truncated. In addition to the increased efficiency and the ability to impose priors on N (which DA does not allow for) this would seem to more than justify the consideration of SCR RJMCMC models.

Future work could focus on making this approach more accessible to researchers from outside of statistics. As it stands, there is already a good deal of literature and software devoted towards implementing DA for CR analysis. Development of an R package which could fit the model presented here would go a long way towards making this RJMCMC approach more accessible to researchers attempting SCR modeling.

More work could also be put into developing priors for N which are specifically related to SCR data. For instance, Schofield and Barker [2014] have considered a RJMCMC approach for modeling hierarchically modeling N over time for non-spatial CR models. A prior for N in a SCR setting could take into account available area and animal movement to impose an ecologically relevant upper bound on the population size.

⁶This argument is worked out in the supplementary material of Schofield and Barker [2014]

Appendix A

Classical Capture-Recapture

Notation

A.1 Closed population notation

For the classical closed population capture-recapture models keeping inline with Otis et al. [1978] we denote

- N : the true population size
- t : the number of days the experiment was run
- n_i : the number of individuals caught in the i^{th} sample
- n_{\cdot} : the total number of captures during the experiment $n_{\cdot} = \sum_{i=1}^t n_i$
- m_i : the number of marked individuals in the i^{th} sample
- M_i : the number of marked individuals in total by the i^{th} sample
- M_{t+1} : the total number of individuals captured during the experiment

- u_i : the number of unmarked individuals in the i^{th} sample, thus $u_i = n_i - m_i$
- f_i : the capture frequencies, f_i is the number of animals which were caught exactly i times, with f_0 being an unknown quantity we hope to estimate
- X_ω : the number of animals with the specific capture history ω

For X_ω the capture histories ω refer to a specific sequence of $\{0, 1\}$ summarizing the days the individual was (1) or was not (0) captured. For example, given $t = 4$ then the number of individuals captured on day 1,3 and 4 but not on day 2 would be denoted X_{1011} . Another common notation for denoting the number of individuals with a specific capture history is as follows

u_i : the number of individuals caught only in sample i

$u_{i,j}$: the number of individuals caught only in samples i and j

⋮

$u_{1,2,\dots,s}$: the number of individuals caught only in samples 1, 2, ..., t

A.2 Open population notation

For open population capture-recapture models keeping inline with McCrea and Morgan [2014] and Schwarz and Arnason [1996] we denote

- L_1, L_2, L_3 : the J-S likelihood components corresponding to the probability of first

capture, loss on capture and recapture respectively

- t : the total number of trapping occasions (often denoted alternatively as k)
- U_i : number of unmarked individuals in the population at the time of the i^{th} sample
- u_i : the number of unmarked individuals caught in the i^{th} sample
- n_i : the total number of individuals caught in the i^{th} sample
- r_i : the number of individuals lost on capture in the i^{th} sample
- v_i : the probability of loss on capture in the i^{th} sample
- r_i : the number of individuals lost on capture in the i^{th} sample
- m_{ij} : the number of individuals caught at time i and recaptured at time j
- v_{ij} : the probability of m_{ij}
- R_i : the total number of individuals released after sample i
- ϕ_i : the probability an individual that was released after sample i survives until sample $i + 1$

Appendix B

Code for Spatial Capture-Recapture

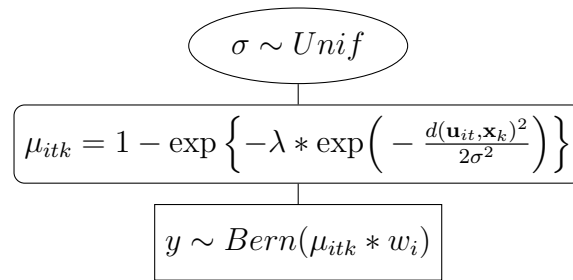
Model with Dependent Animal

Movement Presented in Chapter 2

The following is a breakdown of the code for each parameter of the SCR model with dependent animal movement presented in Chapter 2. Since observed individuals are fixed as $\{w_i = 1; i \leq n\}$ and unobserved individual's centers $\{\mathbf{s}_i; i > n\}$ are fixed (to allow the attractions to unobserved individuals to remain consistent across chain steps) these will not be discussed below.

B.1 σ and λ estimation

The logic and code behind estimating σ and λ are virtually identical so only the code for σ is given below.



pseudo code

1. Propose σ^{prop} with Normal random walk, symmetric proposal function will cancel out in the M-H ratio
2. Calculate μ^{prop} for individual's with $w_i = 1$
3. Calculate log-likelihood for \mathbf{y} given σ (lly) and σ^{prop} (lly_prop)

note: the likelihood for σ 's prior is not included as it will cancel out in the M-H ratio under the assumption $\sigma \sim Unif$

code

```
sigma_prop <- rnorm(1, sigma, sigma_scale)
if (sigma_prop > 0) {

  #proposal mu
  mu.w1_prop <- 1-exp(-lam*exp(-dist2[w1Index]/
    (2*sigma_prop^2)))

  #log-likelihood for data (y)
  lly.vec_prop <- dbinom(y[w1Index],1,mu.w1_prop,
    log = TRUE)

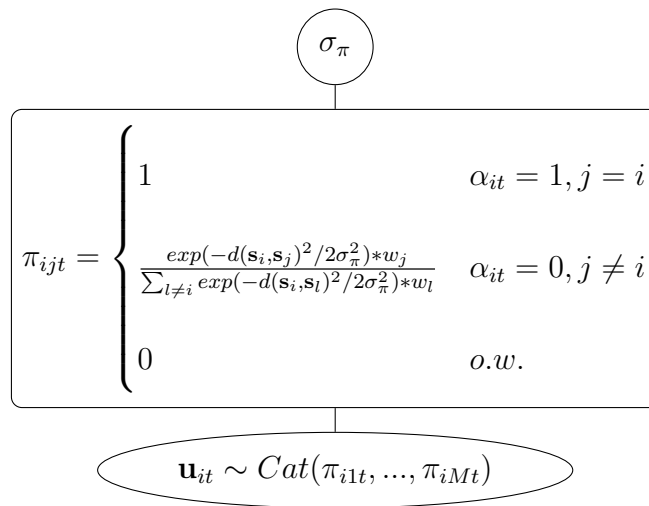
  lly_prop <- sum(lly.vec_prop)
```

```

if (runif(1) < exp(lly_prop - lly)) {
  mu[w1Index] <- mu.w1_prop
  lly.vec[w1Index] <- lly.vec_prop
  lly <- lly_prop
  sigma <- sigma_prop
}
}

```

B.2 σ_π estimation



pseudo code

1. Propose σ_π^{prop} with Normal random walk, symmetric proposal function will cancel out in the M-H ratio, this proposal will be rejected if it is either too large or too small. The minimum (`sigma_pi_min`) is chosen to ensure the probability of an individual being attracted away from it's center is always non-zero, and is a function of whatever the minimum value allowed in R is set to before rounding

to zero (denoted as c^1) and the minimum between the trapping area width or height. Assuming the trapping area width is less than the area height without loss of generality we would have:

$$\begin{aligned} \exp\left(-\frac{areaWidth^2}{2\sigma_\pi^2}\right) &= c \\ \Rightarrow \sigma_\pi &= \sqrt{\frac{areaWidth^2}{-2 * \ln(c)}}. \end{aligned}$$

A maximum (`sigma_pi_max`) is simply chosen as the largest possible distance an individual can travel within the trapping area $\sqrt{areaWidth^2 + areaHeight^2}$.

2. Calculate $\boldsymbol{\pi}^{prop}$, must calculate full matrix
 - (a) calculate `Pi_num_prop`, a matrix of the numerator terms of each π_{ijt}
 - (b) set the diagonal of `Pi_num_prop` to zero to ensure an individual won't be attracted to it's own center (as there is already a term to control this event in the model, namely α_{it})
 - (c) calculate `Pi_num_w_prop`, which is `Pi_num_prop` with the columns corresponding to unobserved individuals with $w_i = 0$ zeroed out to ensure there cannot be attraction to unobserved individuals
 - (d) calculate `Pi_prop` which is `Pi_num_w_prop` scaled by the row sums

¹For all the analysis run for this study we have used $c = 1e - 300$

3. Calculate log-likelihood for the attractions (u_{it} 's for individuals on days where $\alpha_{it} = 0$) given σ_π (11a) and σ_π^{prop} (11a_prop)

note: the likelihood for σ_π 's prior is not included as it will cancel out in the M-H ratio under the assumption $\sigma_\pi \sim Unif(\sigma_{\pi_min}, \sigma_{\pi_max})$

code

```
sigma_pi_prop <- rnorm(1, sigma_pi, sigma_pi_scale)
if (sigma_pi_prop > sigma_pi_min &
    sigma_pi_prop < sigma_pi_max) {

  #proposal pi matrix
  Pi_num_prop <- exp(-S_dist2/(2*sigma_pi_prop^2))
  diag(Pi_num_prop) <- 0
  Pi_num_w_prop <- t(t(Pi_num_prop)*w)
  Pi_prop <- Pi_num_w_prop/rowSums(Pi_num_w_prop)

  #log-likelihoods for attractions
  lla_prop <- 0
  for(i in 1:n){
    for(t in 1:nDays){
      if(Alpha[i,t]==0){
        lla_prop <- lla_prop + log(Pi_prop[i,U[i,t]])
      }
    }
  }

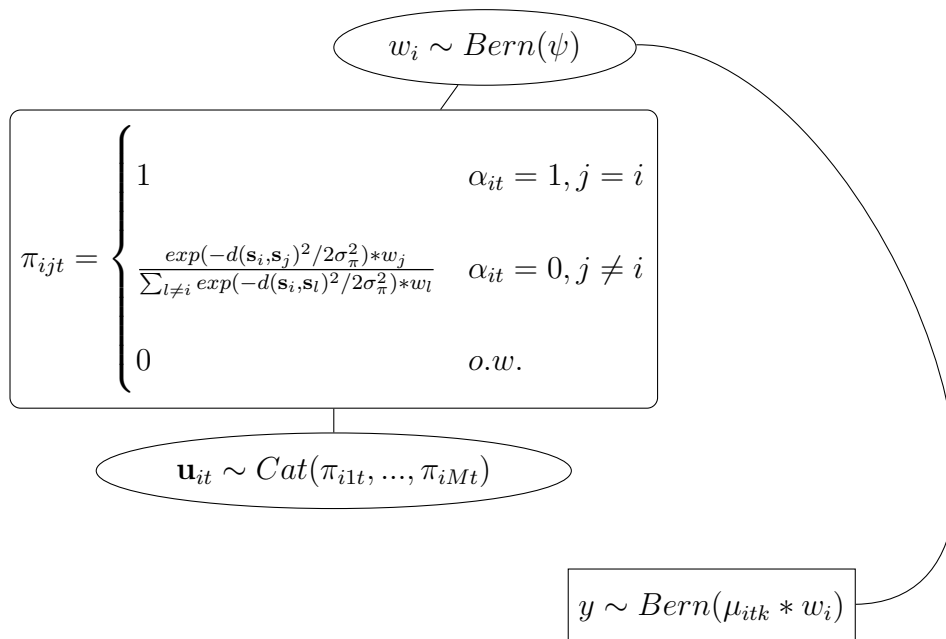
  if (runif(1) < exp(lla_prop-lla)) {
    sigma_pi <- sigma_pi_prop
    Pi_num <- Pi_num_prop
    Pi_num_w <- Pi_num_w_prop
    Pi <- Pi_prop
  }
}
```

```

    lla <- lla_prop
  }
}

```

B.3 Unobserved w_i estimation



pseudo code

1. Propose w_i^{prop} as the opposite of whichever value w_i took the previous chain step, since w_i^{prop} will take this value with probability 1, the proposal function will not effect the M-H ratio
2. Calculate log-likelihoods for w_i 's prior
3. Calculate $\boldsymbol{\pi}^{prop}$, must calculate full matrix, following the same steps as when we

have proposed σ_π in Section B.2 but starting from step (c) in the pseudo code

4. Calculate log-likelihood for the attractions given w_i (11a) and w_i^{prop} (11a_prop)
5. Calculate log-likelihood for y given w_i (11y) and w_i^{prop} (11y_prop)

code

```

for (i in (n+1):M) {
  w_prop.i <- ifelse(w[i] == 0, 1, 0)

  #log-likelihood for w_i's prior
  llPrior <- dbinom(w[i], 1, psi, log = TRUE)
  llPrior_prop <- dbinom(w_prop.i, 1, psi, log = TRUE)

  #proposed pi matrix
  w_prop <- w #vector of w's including proposed w_i
  w_prop[i] <- w_prop.i
  Pi_num_w_prop <- Pi_num_w
  Pi_num_w_prop[,i] <- Pi_num[,i]*w_prop.i
  Pi_prop <- Pi_num_w_prop/rowSums(Pi_num_w_prop)

  #log-likelihood for attractions
  lla_prop <- 0
  for(j in 1:n){
    for(t in 1:nDays){
      if(Alpha[j,t]==0){
        lla_prop <- lla_prop + log(Pi_prop[j,U[j,t]])
      }
    }
  }

  #precalculated indexes for individual i
  index.i <- index.i.mat[i,]

```

```

#log-likelihood for data (y)
if(w_prop.i==1){
  mu.i_prop <- 1-exp(-lam*exp(-dist2[index.i]/
                                (2*sigma^2)))
}else{
  #if w_i=0 don't calculate mu
  mu.i_prop <- 0
}
lly.i.vec_prop <- log(1-mu.i_prop)
lly.i_prop <- sum(lly.i.vec_prop)
lly.i <- sum(lly.vec[index.i])

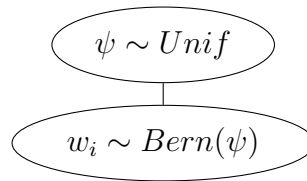
if (runif(1) < exp((lly.i_prop+llPrior_prop+lla_prop)-
                  (lly.i + llPrior + lla))) {
  w[i] <- w_prop.i
  Pi_num_w <- Pi_num_w_prop
  Pi <- Pi_prop
  lla <- lla_prop
  mu[index.i] <- mu.i_prop
  lly.vec[index.i] <- lly.i.vec_prop
  accept_w[i-n] <- accept_w[i-n]+1
}
}

```

B.4 ψ and ψ_α estimation

The logic and code for estimating ψ and ψ_α is virtually identical so only ψ is discussed below. This approach of using Gibbs sampling within the M-H step to sample ψ was

presented in Royle et al. [2013a].



pseudo code As ψ only enters the likelihood through the w_i 's and its own prior we can obtain the full conditional posterior likelihood for ψ as

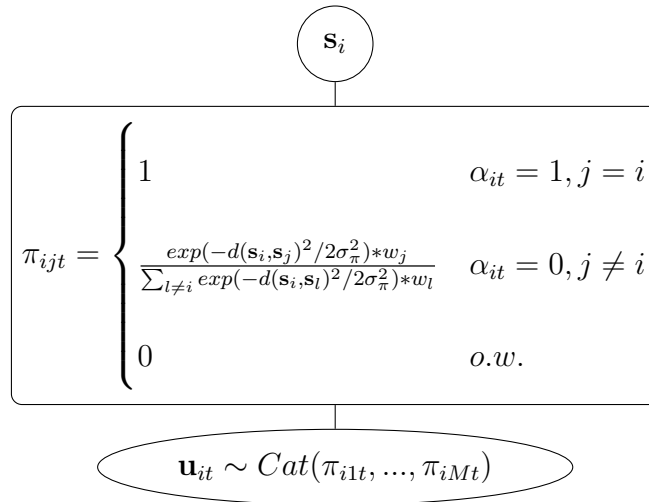
$$\begin{aligned}
 f(\psi|\cdot) &\propto [\psi] \sum_{i=1}^M [w_i|\psi] \\
 &\propto \sum_{i=1}^M [w_i|\psi] && \text{-under } \psi \sim Unif \text{ assumption} \\
 &\propto \psi^{\sum w_i} (1 - \psi)^{M - \sum w_i}.
 \end{aligned}$$

Thus ψ 's full conditional posterior is given by a $Beta(\sum_{i=1}^M w_i + 1, M - \sum w_i + 1)$ and can be sampled using Gibbs.

code

```
psi <- rbeta(1, 1 + sum(w), 1 + M - sum(w))
```

B.5 Observed s_i estimation



pseudo code

1. Propose \mathbf{s}_i^{prop} with Normal random walk. A symmetric proposal function will cancel out in the M-H ratio
2. Calculate squared distances to traps and μ corresponding to proposed centers (`dist2_prop.i1` and `mu_prop.i1`) for a single day
3. Calculate squared distance matrix between centers (`S.dist2_prop`)
4. Calculate $\boldsymbol{\pi}^{prop}$ then divide by row sums
5. Calculate log-likelihood for the attractions given \mathbf{s}_i (`11a`) and \mathbf{s}_i^{prop} (`11a_prop`), need to calculate for all attractions as $\boldsymbol{\pi}^{prop}$ has changed
6. Calculate log-likelihood for \mathbf{y} given \mathbf{s}_i (`11y.n`) and \mathbf{s}_i^{prop} (`11y_prop.n`) only for individuals attracted to \mathbf{s}_i , if there are no individuals with $u_t = \mathbf{s}_i$ for any t then

these values will be set to zero so they are not present in the M-H ratio

note: the likelihood for \mathbf{s}_i 's prior is not included as it will cancel out in the M-H ratio under the assumption $\mathbf{s}_i \sim Unif$. For efficiency we only calculate the squared distances and capture probabilities for a single day in step (1) and then repeat these vectors for parts of the likelihood related to \mathbf{s}_i , additionally we only update cells and rows corresponding to \mathbf{s}_i in steps (3) and (4). Also note we only update μ , the squared distance vector and `lly.vec` for indexes corresponding to individuals with daily locations assigned as \mathbf{s}_i . All the indexes for these vectors will still be updated when other \mathbf{s}_i are proposed which have individuals with these corresponding indexes attracted to them.

code

```
for (i in 1:n) {
  sx_prop.i <- rnorm(1, sx[i], s_scale)
  sy_prop.i <- rnorm(1, sy[i], s_scale)
  if(sx_prop.i>leftBorder & sx_prop.i<rightBorder &
     sy_prop.i>bottomBorder & sy_prop.i<topBorder) {

    #distances to traps from proposal centers
    #for a single day
    dist2_prop.i1 <- (sx_prop.i-traps$x)^2+
                     (sy_prop.i-traps$y)^2
    mu_prop.i1 <- 1-exp(-lam*exp(-dist2_prop.i1/
                                (2*sigma^2)))

    #vector of centers including the proposals
    #sx_prop.i and sy_prop.i
    sx_prop<-sx[1:M]; sx_prop[i]<-sx_prop.i
    sy_prop<-sy[1:M]; sy_prop[i]<-sy_prop.i
  }
}
```

```

#calculate proposal pi matrix
S_dist2_prop.i <- (sx_prop.i-sx_prop)^2+
                  (sy_prop.i-sy_prop)^2
Pi_num_w_prop <- Pi_num_w
Pi_num_w_prop[,i] <- exp(-S_dist2_prop.i/
                        (2*sigma_pi^2))
Pi_num_w_prop[i,] <- Pi_num_w_prop[,i]*w
Pi_num_w_prop[i,i] <- 0
Pi_prop <- Pi_num_w_prop/rowSums(Pi_num_w_prop)

#log-likelihood for attractions
lla_prop <- 0
attracted.indexes <- c()
lly.i.vec_prop <- c()
dist2_prop.i <- c()
nAttractions <- 0
for(j in 1:n){
  for(t in 1:nDays){
    if(Alpha[j,t]==0){
      lla_prop <- lla_prop + log(Pi_prop[j,U[j,t]])
    }

    #keep track of individuals attracted to s_i
     #(including individual i) for calculating
     #lly_prop below
    if(U[j,t]==i){
      attracted.indexes <- c(attracted.indexes,
                            index.it.list[[j]][t,])
      nAttractions <- nAttractions + 1
    }
  }
}

```



```

#log-likelihood for data (y) for individuals
#attracted to s_i (including individual i)
if(nAttractions>0){
  mu_prop.i <- rep(mu_prop.i1,nAttractions)
  lly.i.vec_prop <- dbinom(y[attracted.indexes],
                           1,mu_prop.i,log=TRUE)
  lly.i_prop <- sum(lly.i.vec_prop)
  lly.i <- sum(lly.vec[attracted.indexes])
}else{
  #if no attractions to s_i do not include
  #lly in M-H ratio
  lly.i_prop <- 0
  lly.i <- 0
}

#M-H step
if (runif(1) < exp(la_prop+lly.i_prop-
                  (la+lly.i))) {
  sx[i] <- sx_prop.i
  sy[i] <- sy_prop.i

  #if there were attractions to s_i
  #update mu, dist2 and lly, otherwise
  #these will be updated when another
  #s_i is proposed which the individuals
  #with corresponding indexes are
  #attracted to
  if(nAttractions>0){
    mu[attracted.indexes] <- mu_prop.i
    dist2[attracted.indexes] <- rep(dist2_prop.i1,
                                     nAttractions)
    lly.vec[attracted.indexes] <- lly.i.vec_prop
  }
}

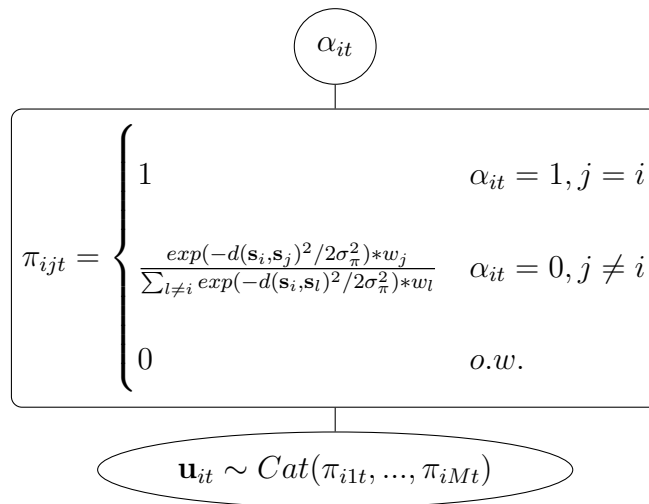
```

```

S_dist2[i,] <- S_dist2_prop.i
S_dist2[,i] <- S_dist2_prop.i
Pi_num[i,] <- Pi_num_w_prop[,i]
Pi_num[,i] <- Pi_num_w_prop[,i]
Pi_num_w <- Pi_num_w_prop
Pi <- Pi_prop
lla <- lla_prop
}
}
}
lly <- sum(lly.vec)

```

B.6 α_{it} estimation



pseudo code

1. By similar logic as when we proposed the w_i 's, propose α_{it}^{prop} as the opposite of whichever value α_{it} took the previous chain step, since α_i^{prop} will take this value with probability 1, the proposal function will not effect the M-H ratio

2. Calculate log-likelihoods for α_{it} 's prior
3. Calculate proposed location for individual i on the current day ($U_{prop.it}$)
4. Calculate squared distances ($dist2_{prop.it}$), μ (mu_{prop}) and log-likelihoods ($lly_{prop.it}$ and lly_{prop}) just for individual i on day t

code

```

for(i in 1:n){
  for(t in 1:nDays){

    #prop for alpha for individual i during
    #a single day
    alpha_prop.it <- ifelse(Alpha[i,t]==1,0,1)

    #log-likelihood for alpha.it's prior
    llAlpha_prior <- dbinom(Alpha[i,t],1,psi_alpha,
                           log=TRUE)
    llAlpha_prior_prop <- dbinom(alpha_prop.it,
                                1,psi_alpha,
                                log=TRUE)

    #propose attraction if alpha.it==0,
    #otherwise set location to home center
    if(alpha_prop.it==0){
      U_prop.it <- rcat(1,Pi[i,])
    }else{
      U_prop.it <- i
    }

    #proposed squared distances for individual i
    #on day t to traps
  }
}

```

```

dist2_prop.it <- (sx[U_prop.it]-traps$x)^2 +
                 (sy[U_prop.it]-traps$y)^2

#proposed mu for individual i on day t
mu.it_prop <- 1 - exp(-lam*exp(-dist2_prop.it/
                              (2*sigma^2)))

#likelihood for observation y for individual i
#on day t
index.it <- index.it.list[[i]][t,]
lly.it.vec_prop <- dbinom(y.list[[i]][t,],1,
                          mu.it_prop,log=TRUE)
lly.it_prop <- sum(lly.it.vec_prop)
lly.it <- sum(lly.vec[index.it])

#M-H step
if (runif(1) < exp((lly.it_prop+llAlpha_prior_prop)-
                  (lly.it + llAlpha_prior))) {
  Alpha[i,t] <- alpha_prop.it
  U[i,t] <- U_prop.it
  dist2[index.it] <- dist2_prop.it
  mu[index.it] <- mu.it_prop
  lly.vec[index.it] <- lly.it.vec_prop
}
}
}

```

Bibliography

Steven C Amstrup, Trent L McDonald, and Bryan FJ Manly. *Handbook of capture-recapture analysis*. Princeton University Press, 2010.

Ben C Augustine, J Andrew Royle, Marcella J Kelly, Christopher B Satter, Robert S Alonso, Erin E Boydston, Kevin R Crooks, et al. Spatial capture–recapture with partial identity: An application to camera traps. *The Annals of Applied Statistics*, 12(1):67–95, 2018.

E Balderamaa, B Gardnera, and B Reichb. Species abundance estimation methods for single-catch trap surveys. *Environmetrics*, 2012.

Simon Bonner and Matthew Schofield. Mc (mc) mc: exploring monte carlo integration within mcmc for mark–recapture models with individual covariates. *Methods in Ecology and Evolution*, 5(12):1305–1315, 2014.

David Borchers. A non-technical overview of spatially explicit capture–recapture models. *Journal of Ornithology*, 152(2):435–444, 2012.

David Borchers, Greg Distiller, Rebecca Foster, Bart Harmsen, and Lorenzo Milazzo. Continuous-time spatially explicit capture–recapture models, with an application to a jaguar camera-trap survey. *Methods in Ecology and Evolution*, 5(7):656–665, 2014.

David Borchers, Rachel Fewster, et al. Spatial capture–recapture models. *Statistical Science*, 31(2):219–232, 2016.

David L Borchers and Murray G Efford. Spatially explicit maximum likelihood methods for capture–recapture studies. *Biometrics*, 64(2):377–385, 2008.

David L Borchers, Walter Zucchini, and Rachel M Fewster. Mark-recapture models for line transect surveys. *Biometrics*, pages 1207–1220, 1998.

David L Borchers, Stephen T Buckland, and Walter Zucchini. *Estimating animal abundance: closed populations*, volume 13. Springer Science & Business Media, 2002.

David L Borchers, BC Stevenson, D Kidney, L Thomas, and Tiago A Marques. A unifying model for capture–recapture and distance sampling surveys of wildlife populations. *Journal of the American Statistical Association*, 110(509):195–204, 2015.

Kenneth P Burnham and W Scott Overton. Robust estimation of population size when capture probabilities vary among animals. *Ecology*, 60(5):927–936, 1979.

Kenneth P Burnham and Walter Scott Overton. Estimation of the size of a closed population when capture probabilities vary among animals. *Biometrika*, 65(3):625–633, 1978.

Bradley P Carlin and Siddhartha Chib. Bayesian model choice via markov chain monte carlo methods. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(3):473–484, 1995.

AD Carothers. Quantifying unequal catchability and its effect on survival estimates in an actual population. *The Journal of Animal Ecology*, pages 863–869, 1979.

Richard B Chandler, J Andrew Royle, et al. Spatially explicit models for inference about density in unmarked or partially marked populations. *The Annals of Applied Statistics*, 7(2):936–954, 2013.

Richard B Chandler, Kristin Engebretsen, Michael J Cherry, Elina P Garrison, and Karl V Miller. Estimating recruitment from capture–recapture data by modelling spatio-temporal variation in birth and age-specific survival rates. *Methods in Ecology and Evolution*, 9(10):2115–2130, 2018.

Anne Chao. An overview of closed capture-recapture models. *Journal of Agricultural, Biological, and Environmental Statistics*, 6(2):158–175, 2001.

Anne Chao, SM Lee, and SL Jeng. Estimating population size for capture-recapture data when capture probabilities vary by time and individual animal. *Biometrics*, pages 201–216, 1992.

Anne Chao, Wenten Chu, and Chiu-Hsieh Hsu. Capture–recapture when time and behavioral response affect capture probabilities. *Biometrics*, 56(2):427–433, 2000.

Anne Chao, HY Pan, and Shu-Chuan Chiang. The petersen-lincoln estimator and its extension to estimate the size of a shared population. *Biometrical Journal*, 50(6):957, 2008.

Douglas George Chapman. Some properties of the hypergeometric distribution with applications to zoological sample censuses. 1(7), 1951.

RM Cormack. Estimates of survival from the sighting of marked animals. *Biometrika*, 51(3/4):429–438, 1964.

John N Darroch. The multiple-recapture census: I. estimation of a closed population. *Biometrika*, 45(3/4):343–359, 1958.

- John N Darroch. The multiple-recapture census: II. estimation when there is immigration or death. *Biometrika*, 46(3/4):336–351, 1959.
- Deanna K Dawson and Murray G Efford. Bird population density estimated from acoustic signals. *Journal of Applied Ecology*, 46(6):1201–1209, 2009.
- Greg Distiller and David L Borchers. A spatially explicit capture–recapture estimator for single-catch traps. *Ecology and evolution*, 5(21):5075–5087, 2015.
- Robert M Dorazio. Bayes and empirical bayes estimators of abundance and density from spatial capture-recapture data. *PloS one*, 8(12):e84017, 2013.
- John W Durban and David A Elston. Mark-recapture with occasion and individual effects: abundance estimation through bayesian model selection in a fixed dimensional parameter space. *Journal of agricultural, biological, and environmental statistics*, 10(3):291, 2005.
- Murray Efford. Density estimation in live-trapping studies. *Oikos*, 106(3):598–610, 2004.
- Murray G Efford. Estimation of population density by spatially explicit capture–recapture analysis of data from area searches. *Ecology*, 92(12):2202–2207, 2011.
- Murray G Efford, David L Borchers, and Andrea E Byrom. Density estimation by spatially explicit capture–recapture: likelihood-based methods. In *Modeling demographic processes in marked populations*, pages 255–269. Springer, 2009a.
- Murray G Efford, Deanna K Dawson, and David L Borchers. Population density estimated from locations of individuals on a passive detector array. *Ecology*, 90(10):2676–2682, 2009b.
- Torbjørn Ergon and Beth Gardner. Separating mortality and emigration: modelling space use, dispersal and survival with robust-design spatial capture–recapture data. *Methods in Ecology and Evolution*, 5(12):1327–1336, 2014.
- Stephen E Fienberg, Matthew S Johnson, and Brian W Junker. Classical multilevel and bayesian approaches to population size estimation using multiple lists. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 162(3):383–405, 1999.
- R.A. Fisher and E.B. Ford. The spread of a gene in natural conditions in a colony of the moth panaxia dominula l. *Heredity*, (1):143–174, 1947.
- Beth Gardner, J Andrew Royle, and Michael T Wegan. Hierarchical models for estimating density from dna mark–recapture studies. *Ecology*, 90(4):1106–1115, 2009.

Beth Gardner, Juan Reppucci, Mauro Lucherini, and J Andrew Royle. Spatially explicit inference for open populations: estimating demographic parameters from camera-trap studies. *Ecology*, 91(11):3376–3383, 2010a.

Beth Gardner, J Andrew Royle, Michael T Wegan, Raymond E Rainbolt, and Paul D Curtis. Estimating black bear density using dna data from hair snares. *The Journal of Wildlife Management*, 74(2):318–325, 2010b.

Richard Glennie, David Louis Borchers, Matthew Murchie, Bart Harmsen, and Rebecca Foster. Open population maximum likelihood spatial capture-recapture. *Biometrics*, 2017.

Peter J Green. Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika*, 82(4):711–732, 1995.

Annegret Grimm, Bernd Gruber, and Klaus Henle. Reliability of different mark-recapture methods for population size estimation tested against reference population sizes constructed from field data. *PloS one*, 9(6):e98840, 2014.

RM Huggins. On the statistical analysis of capture experiments. *Biometrika*, pages 133–140, 1989.

CHN Jackson. The analysis of an animal population. *The Journal of Animal Ecology*, pages 238–246, 1939.

George M Jolly. Explicit estimates from capture-recapture data with both death and immigration-stochastic model. *Biometrika*, 52(1/2):225–247, 1965.

William L Kendall, Kenneth H Pollock, and Cavell Brownie. A likelihood-based approach to capture-recapture estimation of demographic parameters under the robust design. *Biometrics*, pages 293–308, 1995.

William L Kendall, James D Nichols, and James E Hines. Estimating temporary emigration using capture–recapture data with pollocks robust design. *Ecology*, 78(2):563–578, 1997.

Alison C Ketz, Therese L Johnson, Mevin B Hooten, and N Thompson Hobbs. A hierarchical bayesian approach for handling missing classification data. *Ecology and Evolution*, 2019.

R King and SP Brooks. On the bayesian estimation of a closed population size in the presence of heterogeneity and model uncertainty. *Biometrics*, 64(3):816–824, 2008.

Ruth King, Brett T McClintock, Darren Kidney, David Borchers, et al. Capture–recapture abundance estimation using a semi-complete data likelihood approach. *The Annals of Applied Statistics*, 10(1):264–285, 2016.

Michelle L Kissling and Edward O Garton. Estimating detection probability and density from point-count surveys: a combination of distance and double-observer sampling. *The Auk*, 123(3):735–752, 2006.

Shen-Ming Lee and Anne Chao. Estimating population size via sample coverage for closed capture-recapture models. *Biometrics*, pages 88–97, 1994.

F.C. Lincoln. *Calculating Waterfowl Abundance on the Basis of Banding Returns*. Circular (United States. Dept. of Agriculture). U.S. Department of Agriculture, 1930. URL https://books.google.com/books?id=w4n_MQEACAAJ.

Rachel S McCrea and Byron JT Morgan. *Analysis of capture-recapture data*. CRC Press, 2014.

Charles McDougal. *The face of the tiger*. Rivington Books: Distributed by Deutsch, 1977.

P.A.P. Moran. A mathematical theory of animal trapping. *Biometrika*, 38(3/4): 307–311, 1951.

Dana J Morin, Marcella J Kelly, and Lisette P Waits. Monitoring coyote population dynamics with fecal dna and spatial capture–recapture. *The Journal of Wildlife Management*, 80(5):824–836, 2016.

James L Norris and Kenneth H Pollock. A capture-recapture model with heterogeneity and behavioural response. *Environmental and Ecological Statistics*, 2(4): 305–313, 1995.

James L Norris and Kenneth H Pollock. Nonparametric mle under two closed capture-recapture models with heterogeneity. *Biometrics*, pages 639–649, 1996.

Gordon H Orians and PH Leslie. A capture-recapture analysis of a shearwater population: With a statistical appendix. *The Journal of Animal Ecology*, pages 71–86, 1958.

David L Otis, Kenneth P Burnham, Gary C White, and David R Anderson. Statistical inference from capture data on closed animal populations. *Wildlife monographs*, 62:3–135, 1978.

Carl Georg Johannes Petersen. The yearly immigration of young plaice into the limfjord from the german sea. *Report of the Danish Biological Station*, 6:1–48, 1896.

Shirley Pledger. Unified maximum likelihood estimates for closed capture–recapture models using mixtures. *Biometrics*, 56(2):434–442, 2000.

Kenneth H. Pollock. The assumption of equal catchability of animals in tag-recapture experiments. *unpublished Ph.D. dissertation, Cornell University, Biometrics Unit*, 1974.

Kenneth H. Pollock. Review papers: modeling capture, recapture, and removal statistics for estimation of demographic parameters for fish and wildlife populations: past, present, and future. *Journal of the American Statistical Association*, 86(413): 225–238, 1991.

Kenneth H Pollock and Mark C Otto. Robust estimation of population size in closed animal populations from capture-recapture experiments. *Biometrics*, pages 1035–1049, 1983.

Kenneth H Pollock, James D Nichols, Cavell Brownie, and James E Hines. Statistical inference for capture-recapture experiments. *Wildlife monographs*, pages 3–97, 1990.

Brian J Reich and Beth Gardner. A spatial capture-recapture model for territorial species. *Environmetrics*, 25(8):630–637, 2014.

J Andrew Royle. Analysis of capture–recapture models with individual covariates using data augmentation. *Biometrics*, 65(1):267–274, 2009.

J Andrew Royle and Robert M Dorazio. Parameter-expanded data augmentation for bayesian analysis of capture–recapture models. *Journal of Ornithology*, 152(2): 521–537, 2012.

J Andrew Royle and Kevin V Young. A hierarchical model for spatial capture–recapture data. *Ecology*, 89(8):2281–2289, 2008.

J Andrew Royle, Robert M Dorazio, and William A Link. Analysis of multinomial models with unknown index using data augmentation. *Journal of Computational and Graphical Statistics*, 16(1):67–85, 2007.

J Andrew Royle, K Ullas Karanth, Arjun M Gopalaswamy, and N Samba Kumar. Bayesian inference in camera trapping studies for a class of spatial capture–recapture models. *Ecology*, 90(11):3233–3244, 2009a.

J Andrew Royle, James D Nichols, K Ullas Karanth, and Arjun M Gopalaswamy. A hierarchical model for estimating density in camera-trap studies. *Journal of Applied Ecology*, 46(1):118–127, 2009b.

J Andrew Royle, Richard B Chandler, Rahel Sollmann, and Beth Gardner. *Spatial capture-recapture*. Academic Press, 2013a.

J Andrew Royle, Richard B Chandler, Catherine C Sun, and Angela K Fuller. Integrating resource selection information with spatial capture–recapture. *Methods in Ecology and Evolution*, 4(6):520–530, 2013b.

J Andrew Royle, Chris Sutherland, Angela K Fuller, and Catherine C Sun. Likelihood analysis of spatial capture-recapture models for stratified or class structured populations. *Ecosphere*, 6(2):1–11, 2015.

J Andrew Royle, Angela K Fuller, and Chris Sutherland. Spatial capture–recapture models allowing markovian transience or dispersal. *Population ecology*, 58(1):53–62, 2016.

J Andrew Royle, Angela K Fuller, and Christopher Sutherland. Unifying population and landscape ecology with spatial capture–recapture. *Ecography*, 41(3):444–456, 2018.

George B Schaller. *The deer and the tiger*. University of Chicago Press, 1984.

Zoe Emily Schnabel. The estimation of total fish population of a lake. *The American Mathematical Monthly*, 45(6):348–352, 1938.

Matthew Schofield, Richard Barker, et al. 50-year-old curiosities: Ancillarity and inference in capture–recapture models. *Statistical Science*, 31(2):161–174, 2016.

Matthew R Schofield and Richard J Barker. Hierarchical modeling of abundance in closed population capture–recapture models under heterogeneity. *Environmental and ecological statistics*, 21(3):435–451, 2014.

Matthew R Schofield, Richard J Barker, and Nicholas Gelling. Continuous-time capture–recapture in closed populations. *Biometrics*, 74(2):626–635, 2018.

Carl James Schwarz and A Neil Arnason. A general methodology for the analysis of capture-recapture experiments in open populations. *Biometrics*, pages 860–873, 1996.

George AF Seber. A note on the multiple-recapture census. *Biometrika*, 52(1/2): 249–259, 1965.

George Arthur Frederick Seber. *The estimation of animal abundance*. Griffin London, 1982.

Eric P Smith and Gerald van Belle. Nonparametric estimation of species richness. *Biometrics*, pages 119–129, 1984.

Rahel Sollmann, Beth Gardner, Arielle W Parsons, Jessica J Stocking, Brett T McClintock, Theodore R Simons, Kenneth H Pollock, and Allan F O’Connell. A spatial mark–resight model augmented with telemetry data. *Ecology*, 94(3):553–559, 2013a.

Rahel Sollmann, Natália Mundim Tôrres, Mariana Malzoni Furtado, Anah Tereza de Almeida Jácomo, Francisco Palomares, Severine Roques, and Leandro Silveira. Combining camera-trapping and noninvasive genetic data in a spatial capture–recapture framework improves density estimates for the jaguar. *Biological conservation*, 167:242–247, 2013b.

Thomas R Stanley and Kenneth P Burnham. A goodness-of-fit test for capture–recapture model mt under closure. *Biometrics*, 55(2):366–375, 1999.

Ben C Stevenson, David L Borchers, Res Altwegg, René J Swift, Douglas M Gillespie, and G John Measey. A general framework for animal density estimation from acoustic detections across a fixed microphone array. *Methods in Ecology and Evolution*, 6(1):38–48, 2015.

David J Strauss. A model for clustering. *Biometrika*, 62(2):467–475, 1975.

M Wegan. Aversive conditioning, population estimation, and habitat preference of black bears (*ursus americanus*) on fort durm military installation in northern new york. *MSc Thesis, Cornell University*, 2008.

Gary C White. *Capture-recapture and removal methods for sampling closed populations*. Los Alamos National Laboratory, 1982.

DL Zimmerman, M Stein, AE Gelfand, PJ Diggle, M Fuentes, and P Guttorp. *Handbook of spatial statistics*, 2010.

Calvin Zippin. An evaluation of the removal method of estimating animal populations. *Biometrics*, 12(2):163–189, 1956.