DISSERTATION

ADJUSTING FOR CAPTURE, RECAPTURE, AND IDENTITY UNCERTAINTY WHEN ESTIMATING DETECTION PROBABILITY FROM CAPTURE-RECAPTURE SURVEYS

Submitted by

Stacy L. Edmondson

Department of Statistics

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, CO

Summer 2015

Doctoral Committee:

Advisor: Geof Givens

Jean Opsomer
Piotr Kokoszka
Barry Noon

ABSTRACT

ADJUSTING FOR CAPTURE, RECAPTURE, AND IDENTITY UNCERTAINTY WHEN ESTIMATING DETECTION PROBABILITY FROM CAPTURE-RECAPTURE SURVEYS

When applying capture-recapture analysis methods, estimates of detection probability, and hence abundance estimates, can be biased if individuals of a population are not correctly identified (Creel et al., 2003). My research, motivated by the 2010 and 2011 surveys of Western Arctic bowhead whales conducted off the shores of Barrow, Alaska, offers two methods for addressing the complex scenario where an individual may be mistaken as another individual from that population, thus creating erroneous recaptures.

The first method uses a likelihood weighted capture recapture method to account for three sources of uncertainty in the matching process. I illustrate this approach with a detailed application to the whale data.

The second method develops an explicit model for match errors and uses MCMC methods to estimate model parameters. Implementation of this approach must overcome significant hurdles dealing with the enormous number and complexity of potential catch history configurations when matches are uncertain. The performance of this approach is evaluated using a large set of Monte Carlo simulation tests. Results of these test vary from good performance to weak performance, depending on factors including detection probability, number of sightings, and error rates. Finally, this model is applied to a portion of the bowhead survey data and found to produce plausible and scientifically informative results as long as the MCMC algorithm is started at a reasonable point in the space of possible catch history configurations.

ACKNOWLEDGEMENTS

This work was supported and funding by the North Slope Borough (Alaska) and the National Oceanic and Atmospheric Administration (through the Alaska Eskimo Whaling Commission). This research also utilized the CSU ISTeC Cray HPC System supported by NSF Grant CNS-0923386

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Capture Recapture Surveys

Capture-recapture surveys have a long history (Cormack, 1968, Otis et al., 1978, White, 1982, Pollock et al., 1991, Seber, 1982, Schwarz, 1978 and Williams et al., 2002), with one of the main benefits and goals of such surveys being the ability to estimate a population's abundance. In classical capture-recapture experiments there are c independent capture opportunities, where at each capture opportunity animals are captured, tagged with unique identifying tags, and released. During each capture opportunity the researchers record which animals were captured for the first time, and which were recaptured, i.e., which had been captured at a previous capture opportunity. The data collected from capture-recapture experiments are the collection of each individual's capture history. Capture histories can be expressed as sequences of 0's and 1's, where a 0 indicates that the individual was not captured at that specific capture opportunity and a 1 indicates that the animal was captured at that specific capture opportunity. For example, if the number of capture opportunities is $c = 2$, a 10 capture history represents an individual captured at the first opportunity but not recaptured at the second. A 01 history represents an individual not captured at the first opportunity but captured at the second. A 11 capture history represents an individual being captured at the first opportunity and recaptured at the second.

After the experiment is complete capture histories for each individual are used for estimation of detection probability and abundance, where detection probability is defined to be the probability that an individual is captured given it was available to be captured (Amstrup et al., 2005). Many models (Amstrup et al., 2005) include abundance in the likelihood as a parameter to be estimated. Others remove abundance from the likelihood by conditioning on only those individuals captured during the survey. The most notable of these is the Hug-

gins model (1989). For models where abundance is conditioned out of the likelihood, one can use a Horvitz-Thompson type estimator (Särndal et al., 2003) to estimate population abundance where the detection probability estimates are part of the inclusion probabilities. These models are based on the assumption that capturing and physically tagging individuals in the population is possible.

1.1.1 Uncertain captures and recaptures and the goal of this dissertation

More recently individual identification based on naturally existing tags such as DNA (Taberlet et al., 1997) or scar patterns (Langtimm et al., 2004) has been used in capture-recapture studies. While the use of naturally existing tags have broadened the range of ecological studies (Waits, 2004), it has the potential to introduce errors into the data set. In particular, the use of naturally existing tags introduces the potential for mis-identification of individuals (Hammond, 1986; Taberlet et al., 1999; and Wright et al., 2009). Mis-identification occurs when it appears that several individuals were captured when in fact only just one individual was captured. This potential for mis-identification, if unaccounted for, can create bias in estimation (Creel et al., 2003). In some situations, the other type of error can occur: two individuals may be mistaken as a single one. This type of error is seen often in the data I examine in this dissertation.

Some researchers have examined ways to account for mis-identification errors, whether through improved methodology, or through the model used in estimation (Paetkau, 2003, Waits and Paetkau, 2005, Stevick et al., 2001, Lukacs and Burnham, 2005, Yoshizaki, 2007, and Link et al., 2010). Stevick et al. (2001) assumes known error rates. Lukacs and Burnham (2005) extend the full likelihood models of Otis et al. (1978) and the conditional likelihood of Huggins (1989) focusing specifically on an animal's genotype being used as a unique identifying tag when determining model assumptions. Link et al. (2010) relaxes some of the

assumptions made by Lukacs and Burnham (2005) and proceedes by connecting the observed, or recorded frequencies, f, to the latent frequencies, x (with cell probabilities $\pi(\theta)$), via a linear transformation, i.e., $f = A'x$. They then employ a MCMC strategy via the Gibbs and Metropolis-Hasting samplers, treating the latent frequencies as parameters needing to be estimated.

There has been little or no work on methods to cope with other sorts of errors such as falsely declaring a recapture (i.e., when two individuals are mistaken as one). In such a case the data appear as if there are fewer individuals than were actually captured. The large complex whale data set that serves as motivation for this dissertation presents both types of potential errors - failures to match (i.e., mis-identification yielding two purported individual captures instead of one recapture) and false matches (i.e., mistakenly identifying two individuals as a recapture event). My dissertation provides statistical methods for estimating detection probabilities (and hence abundance) in the face of both types of capture and recapture uncertainty. This is challenging, novel work that draws upon a variety of statistical concepts and is strongly grounded in the whale application motivating my work.

1.2 Whale Dataset

In subsequent chapters I will describe the whale dataset motivating my statistical methodology research. Initially, however, it may be useful to provide a quick overview in order to illustrate the presence of false recaptures (i.e., false matches) and failed recognition of recaptures (mis-identification).

The North Slope Borough Department of Wildlife Management in Barrow, Alaska has conducted ice-based visual cenuses of the Bering-Chukchi-Beaufort Seas (BCBS) stock of bowhead whales, *Balaena mysticetus*, for over 30 years. Conducted annually from 1979 to 1988 and then again in 1993, 2001, 2010 and 2011, the main purpose of these censuses is to pro-

vide the International Whaling Commission with population abundance estimates. Further discussion of the history of the surveys and analytical procedures for the years preceding 2010 can be found in George et al. (2004). In 2010 and in 2011 the visual census switched from a single perch design to a two perch independent observer system, creating the need for an updated statistical analysis.

Every spring the BCBS stock of bowhead whales migrates from the Bering Sea into the Beaufort Sea, traveling north-eastward through the Chukchi Sea. From April to June, 2010, and again from April to May, 2011, a two-perch independent observer system was implemented to record visual sightings of passing bowhead whales. The two-perch independent observer system establishes two stationary locations, or perches, constructed on the ice near the water's edge. The two perches are a sufficient distance from one another to ensure independent sightings. When a sighting is made at either perch, the whale location (based on theodolite readings), time, group size, visual sighting conditions, and other covariates are recorded. Note that each perch creates its own independent dataset.

Due to the fact that whales cannot be physically marked or captured, recapturing or 'matching' was done post hoc by a team of experts comparing the sightings from both perches using special data plotting software to display the time, location, and other relevant factors to for each sighting. The researchers replicated the matching process three times for the 2010 survey data over the course of 4 months. Matching was not replicated in 2011. Researchers spent hundreds of hours making these decisions, however it would be a mistake to assume that all of these decisions are correct. Therefore, these surveys produce capture-recapture data with uncertain recaptures.

1.3 Overview

In this dissertation I will provide two new novel approaches that incorporate both types of errors described above in the estimation of detection probabilities. The first method uses a weighted likelihood approach for modeling capture-recapture histories, with weights to account for three sources of identification or match uncertainty, namely conditional sightings, availability and match uncertainty. This approach is described in Chapter 2, with a detailed application to the whale survey data. The second method builds an explicit model for the error processes inherent in the uncertain matching. This approach is developed in Chapter 3. In Chapter 4, the performance of the latter method is examined using a broad suite of simulation testing. Finally, in Chapter 5, application of this approach to a portion of the bowhead data allows an examination of how my methods might be used in practice.

CHAPTER 2

WEIGHTED LIKELIHOOD APPROACH

2.1 Introduction

This chapter describes estimation of detection probabilities from the 2010 and 2011 survey data using a weighted likelihood approach. A revised and extended version of this chapter has been published in *Environmetrics* (Givens et al., 2014).

In April and May of 2010 and 2011, scientists from the North Slope Borough Department of Wildlife Management attempted ice-based counts of bowhead whales (Balaena mysticetus) from the Bering-Chukchi-Beaufort Seas population as the animals migrate northward past Barrow, Alaska. Descriptions of the surveys are given by George et al. (2011).

These surveys employ a two-perch independent observer protocol. Specifically, two teams of observers stand at fixed survey sites (perches) situated atop pressure ridges near leads and open water. The two perches are sufficiently distant that the teams cannot hear each other or incidentally cue each other about their sightings. All sightings and other data are recorded independently at each perch. During some periods (independent observer, or 'IO' periods), both perches were operating. At other times, only one perch, or neither, was used. Typical IO shift periods were 4 hours.

The ultimate purpose of the surveys is to obtain a population abundance estimate. Most previous abundance estimates have been based on ice-based surveys (Braham et al., 1979; Zeh et al., 1986, 1991; George et al., 2004; Zeh and Punt, 2005) The most recent abundance estimate, derived from an aerial photo-identification survey, is 12,631 (95% confidence interval 7,900 to 19,700) by (Koski et al., 2010). The key requirements for estimating abundance from the ice-based survey data are the count of whales seen, the proportion of the population

available to be seen as whales migrate past the perches, and an estimate of the detection probability, i.e., the probability of detecting a whale given that it passes the perch within viewing range. This chapter is focused solely on detection probabilities, which may depend on covariates such as visibility conditions and distance of the whale from the perch.

In 2010, the survey was partially successful. There were 399 hours when at least one perch was operating, during which 306 hours of two-perch independent observer effort was maintained. The total numbers of New and Conditional whales (respectively, first or likely first sightings; see Section 2.2.1 for precise definitions) seen at the primary perch were 1332 and 242, respectively. Of these 1332 New whales, 1216 were seen during 306 hours of IO. Note that many sightings were repeat sightings of the same whale or group made from the same perch or the other one. Despite the partial success of the 2010 survey, it would be difficult to produce a reliable abundance estimate from these data. Ice and weather conditions prevented all survey effort for much of April and May 4-6, and survey effort ended on May 28 before the migration ended. As many as a third to one half of the bowhead population probably passed Barrow during those unmonitored times. However, the 2010 data are ample for detection probability estimation.

The 2011 visual survey was extremely successful. A total of 3379 New and 632 Conditional whales were seen in 859 hours of effort at the primary perch, including 1230 New whales during 180 hours of IO. These counts are within a few whales of the all-time record since surveys began in 1979. Although the total survey effort (859 hours) was much greater than in the previous year due to survey design and weather, the IO effort in 2011 was much less. This change in emphasis from IO and detection probabilities in 2010 to counts in 2011 was planned to optimally use resources and yield the best multi-year dataset for abundance estimation.

This chapter describes estimation of detection probabilities from the available survey data. First I describe how the raw survey data are converted into capture/recapture data appropriate for detection probability estimation. After a description of the estimation approach, I present the results and sensitivity analyses. The chapter ends with a discussion of my findings and future implications.

2.2 Data

In Section 2.3 I describe the statistical model used to estimate detection probabilities (Huggins, 1989). The basis for the analysis is the capture-recapture principle. In a generic case, sighted groups are classified into three categories: those seen by observers in team 1, those seen by team 2, and those seen by both teams. The detection probability (for whale groups) for a team is estimated as the ratio of the number of groups seen by both observers to the total number of animals seen by the other observer team.

However, ordinary capture-recapture methods cannot be directly applied to the bowhead data. Usually, the counts of animal captures and recaptures are known exactly. For example, animals may be marked with bands or matched via photo-identification. In the bowhead case, the best way to count captures and recaptures is less clear because no matching can be done at the moment of sighting or re-sighting (especially at high whale passage rates), and no identifying marks or bands can be placed on the whales. Moreover, even the sightings themselves are subject to confusion since the same whale may be sighted multiple times from the same perch. The sighting and matching methods used in the survey yield very complicated data structures that we call 'chains', which include a time- and perch-varying array of covariates, too. In the following subsections we describe the dataset and how it is organized for analysis.

2.2.1 Linking Sightings

The first stage of my data treatment is to identify captures, in other words whales seen for the first time. One of the most important variables recorded for a sighting is a link code, which indicates whether the observer team believes that it has previously detected the sighted whale(s). The link codes and their meanings are given in Table 2.1. Each sighting from a single perch is labeled as New (N), Duplicate (R, X, Y, or Z) or Conditional (C), where the latter two categories represent a possible re-sighting of a previously seen whale or group.

These link codes are assigned independently at each perch and pertain only to the sightings from that perch. For example, a New whale seen at one perch might never be seen at the other; or perhaps this New whale is a Y Duplicate within a sequence of previous/future sightings at the other perch, such as N-X-Y-R-Z. Duplicate and Conditional sightings are explicitly perch-specific (as are New sightings) and hence are not recaptures. Due to the presence of Duplicate and Conditional sightings, the total number of sightings reported by a perch exceeds the total number of distinct whale groups sighted.

Connections between related sightings from a single perch are called 'links'. Links are established by the observer team at the time of sighting using the link codes, and links can only refer backwards in time to previously seen whales. It is required that a New or Conditional whale be identified as the originating sighting for any subsequent sighting coded as a R, X, or Y Duplicate. By definition, C sightings are not linked back to a previous sighting and Z sightings rarely are, but both types may be connected to future sightings via those future sightings' backward links. Critically, sightings and links at one perch are totally independent

Table 2.1: Link codes for sightings of whales or groups. Every sighting is classified with one of these codes. The three fundamental classifications are New, Duplicate, and Conditional.

Link Code	**Meaning**
N	New whale or group. Observer team is confident that it is seen for the first time.
R	Duplicate. Roll. The sighting is part of a sequence of surface dives or 'roll series' of a previously sighted whale or group. A link is assigned to indicate the associated previous sighting.
X	Duplicate. The observer team is 100% confident that the whale or group can be linked to a specific previous sighting. A link is assigned to indicate the associated previous sighting.
Y	Duplicate. The observer team is about 90% confident that the whale or group can be linked to a specific previous sighting. A link is assigned to indicate the associated previous sighting.
Z	Duplicate. The observer team is quite sure that the whale or group has been previously sighted but the team cannot link it back to a specific previous sighting with 90% confidence. Rarely a link is assigned.
C	Conditional. The observer team cannot determine whether this whale or group is New or a Duplicate of some previous sighting. Links to earlier sightings are forbidden.

of those from the other perch. If a group was seen by both perches, the links (if any) assigned by each perch are unrelated.

Since time is recorded with each sighting, a sequence of sightings connected by links has a unique link ordering. I call a sequence of linked sightings a 'link chain'. By definition, a link chain must begin with a N or C (or rarely but improperly a Z) and contain no subsequent N or C. Since the perches operate independently, link chains are associated with a single perch. Let an arrow denote a link, with superscripts representing the number of whales reported in the group and subscripts representing perch. It will be implicit that time flows from left to right. Then the link chain $N_1^1 \to X_1^1 \to Y_1^2$ represents a link chain created by perch 1, where a new sighting of one whale was linked to an X Duplicate of a single whale, and then to a Y Duplicate where two whales were seen in the group. Note that group size may not be consistent within a chain. The direction of our arrows is, technically, inappropriate because

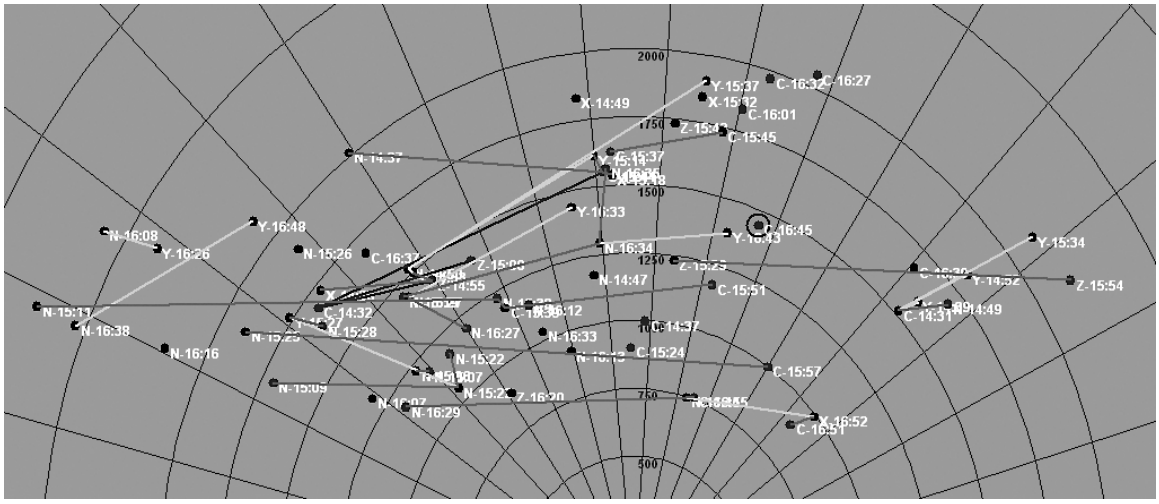

Figure 2.1: Example of the software display for matching sightings during about two hours on May 8, 2010. Sightings (dots) are labeled by link code and time, and associated by perch according to dot shade. Links and matches are indicated by line segments of different shades. The software displays data but does not attempt any automatic matching.

Duplicates are linked backwards in time to previous sightings. However, our notation is more natural to read because it is consistent with time flow. Similarly, we may say that a sighting 'links onward' to another sighting when, more precisely, the future sighting is linked backward in time. Also for simplicity I will omit superscripts or subscripts when they are irrelevant.

We will see later that certain other types of chains may involve loops or other features that preclude the simple notation adopted here. Finally, for brevity later I often adopt the convention that link chains are permitted to be length one, i.e., a single New, Conditional, or Z Duplicate whale, unless the distinction between single- and multi-sighting link chains is important.

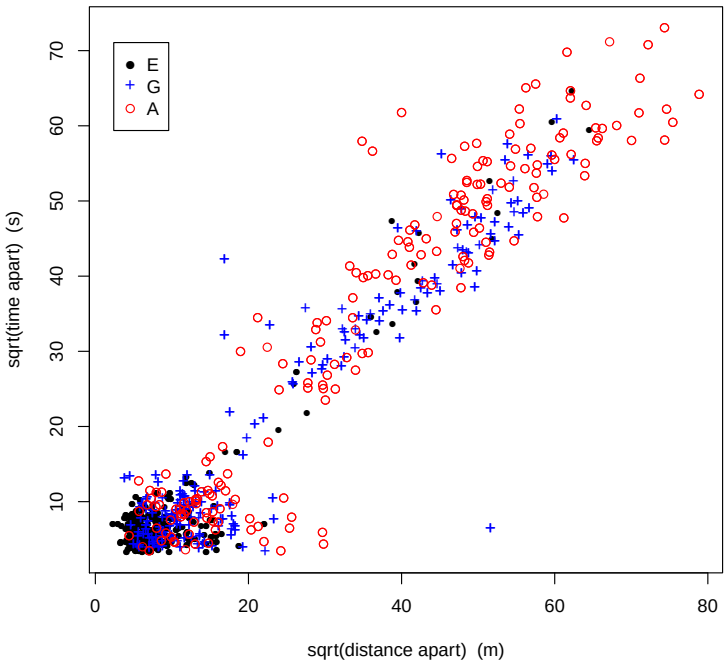

Figure 2.2: Separation of paired matched sightings in space and time. Each dot represents a match pair. The data have been transformed by taking square roots to better separate the dots. The E/G/A categories are explained in Section 2.2.2. E matches (black dots) are the most certain, G matches (plus symbols) are intermediate, and A (open circles)matches are the least certain.

2.2.2 Match Data

Matching sightings

Next I consider the identification of recaptures: whales that were seen from both perches. It is critical to understand that a match connects sightings between perches rather than within a perch. Links connect sightings within a single perch. The observer teams do not communicate, so the process of matching a link chain from one perch to a link chain from the other perch must be done by a third party.

In 2010, matching was attempted both in real time and retrospectively. For the period 1-14 May, observer teams radioed the sighting time, location, swim direction and speed to a com-

mand center. The two teams used different radio frequencies. Master matchers in the command center plotted these data using software adapted specifically for this task and tracked sightings approximately in real time, trying to identify between-perch matches. The software was for data display only–matches were determined using human judgment integrating all relevant information. An example of the match display is shown in Figure 2.1. Sightings are labeled with link code and time. Links and matches are shown with line segments in three different colors. One motivation to attempt real time matching was that the master matchers could radio the perch for additional information if the matchers had a question or needed clarification about a sighting. George et al. (2011) provide much more detail about matching.

At high whale passage rates, it became evident that the real time protocol was not practical. Observers on the perches did not have enough time to record sightings, make links, and also radio back the data to the command center. For the rest of the season, sightings data were collected without real time matching. Several months later, the entire dataset was scrutinized to identify matches. The master matchers used the same software and essentially simulated what would have happened during real time matching, except that the software allowed time to be slowed as much as needed to allow full and careful identification of matches. Match decisions were twice re-assessed and validated through comprehensive reviews later in the year.

In 2011, only post hoc matching was used. Applying the same methods as in 2010, the master matchers produced a matched dataset that they considered to be equivalent to the 2010 match data in terms of data usage and and decision thresholds. The total time spent for the matching effort was about 50 person-weeks for the two surveys.

Figure 2.2 shows how the sightings comprising a matched pair are separated in space and time, for the 2011 data. Most matched sightings are very close spatially and temporally. The gap in the middle of the plot is a result of whale diving behavior. The sighting pairs in

the bottom left portion of the plot are likely to have been sighted by both perches during the same surfacing sequence. The pairs at the top right portion of the plot are likely matches where the sightings were separated by one or more dive periods.

Match chains

Matches explicitly connect sightings, not link chains. However, since every sighting is a member of exactly one perch-specific link chain, a match to any sighting implicitly connects a link chain from one perch to a link chain from the other perch. The entire set of sightings connected by a sequence of matches and links is called a match chain. For 2011, the numbers of match chains, link chains and single sightings were 704, 199, and 3443 respectively.

Match chains may connect a variety of link chains from the two perches, thereby containing multiple matches and multiple links. Note that the matches within a match chain often do not directly connect the earliest sightings within each constituent link chain. Supplementing our previous notation with double arrows to indicate matches, an example match chain might be $N_1^1 \rightarrow X_1^1 \Leftrightarrow N_2^1 \rightarrow Y_2^1 \Leftrightarrow C_1^1$. This represents a link chain from perch 1 (a New whale linked onward to an X duplicate) matched to a link chain from perch 2 (a New whale linked onward to a Y duplicate)–where the match is made between the X duplicate and the New whale at perch 2–and then the Y duplicate is matched to a Conditional sighting at perch 1. The Conditional is not linked to the original link chain at perch 1 even though the master matchers have implicitly asserted that it was an additional sighting of the original chain at perch 1.

The majority of match chains in our dataset are $N_1^1 \Leftrightarrow N_1^1$. With the diversity of cases in the dataset, however, my notation is woefully insufficient. If the first link chain in the previous paragraph was extended so that its X Duplicate was linked onward to an unmatched Y Duplicate at the same perch then the overall match chain would have an additional loose end.

If the New whale from perch 1 was matched to the Y Duplicate from perch 2 then the match chain would include a loop. Forks occur when one sighting is matched to several others. In principle a match chain may have quite a few loose ends, loops, and/or forks. Figure 2.3 shows a match chain of moderately high complexity from the 2010 data.

For the 704 match chains I analyze, the median length is 2, the mean is 2.2, and the longest is 14. To illustrate the potential complexity of chains, consider a whopper from 2010 which has 9 links, 11 matches, 4 forks (one of which is 3-way), 7 loose ends and several loops.

It is impossible for the master matchers to have equal confidence in all match decisions. Thus, to each match they assigned a quality or 'confidence rating'. Excellent (E) matches were considered to have at least 90% certainty. Good (G) matches were believed to be between 66% and 90% certain. Adequate (A) matches were considered to have between 50% and 66% certainty. In addition to these numerical bounds, the matchers were also given plain language interpretations of the categories: 'nearly certain', 'at least twice as likely as not', and 'barely more likely than not', respectively.

2.2.3 Covariate Data

When a sighting is recorded, a variety of covariate data were collected in addition to the time, sighting location, and link code. Additional covariates such as whale passage rate and chain length (hours) can be derived from the observed information. I list the main covariates in the dataset in Table 2.2.

During exploratory data analysis, I investigated a variety of ways to express certain covariates. The most relevant ones for my analyses are as follows. Distance from the perch is

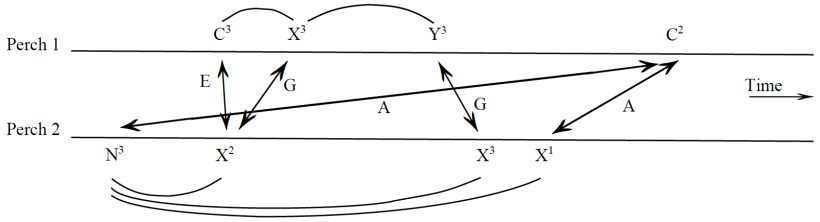

Figure 2.3: Example match chain structure from 2010. Sightings are ordered in time (flowing from left to right). The vertical axis is irrelevant except to separate the two perches. Arcs represent links and arrows are matches. The sighting notation (N/C/X/Z) is explained in Section 2.2.1; the match notation (E/G/A) is explained in Section 2.2.2.

treated as a continuous variable. Visibility is consolidated into three groups: EVG (Excellent or Very Good), GO (Good), and FA (Fair). Although clear, partly cloudy, and overcast weather categories are retained, the remaining categories of heavy fog, heavy rain, heavy snow, light fog, light rain, light snow, and snow/rain mix are consolidated into the category of precipitation. Behavior categories are also consolidated into two groups: easier (breach, spy hop, interaction, tail lobbing, under-ice feeding, and exceptions), and standard (migrating, lingering, resting, trawling, flukes, southbound, unknown, ordinary surfacing, pushing head through hummock, and heard only). Group sizes are consolidated into 1 and > 1, with calves counting towards the total. The rationale for this choice (aside from a good model fit and very few group sizes exceeding two) is that observers felt strongly that the presence of a calf increased detectability at least as much as a second adult whale due to the surface behavior of the cow-calf pair.

Since link and match chains contain several sightings, there is ambiguity about how to assign covariate values to chains. For link chains, the assigned covariate value is taken to be the last value observed within the chain. For match chains, the assigned covariate value is taken to be the value recorded at the first recapture event within the chain. The covariates used in my preferred models vary quite slowly compared to sighting events and are usually assessed consistently by the two perches, so we consider this approach to have at most minor impact compared to other decisions discussed later. For group sizes, the covariate used for model fitting is also taken to be the value observed at the first recapture. However, group sizes introduce an important additional complication in my analysis and are discussed in detail later; see Section 2.3.2.

Table 2.2: Main covariates recorded or derived from sightings data in addition to the time, location, and link code.

Visibility	Unacceptable, Poor, Fair, Good, Very Good, Excellent
Weather	Ten categ. describing cloud cover, precipitation, and fog
Wind Speed and Direction	Miles per hour, degrees clockwise from magnetic north
Ice Coverage	Percent
Lead Condition	Unkn., Closed, Not Contin. (Patchy), Contin., Wide Open
Lead Width	Meters
Swim Direction	North, South, Lingering
Group Size	Number of whales
Calf Presence	Zero or one
Behavior	Fifteen categories, e.g., migrating, trawling, breaching
Number of Observers	Usually 3 or 4 people on a perch
Observer Name	Only for the theodolite operator
Passage Rate	Derived as chains per hour
Chain Length	Derived as total time from first to last sighting
Number of Sightings	Derived as total number of sightings (not whales) in chain

2.2.4 Data Exclusions

Some data were discarded before estimating detection probabilities:

- Closed or indeterminate lead conditions.
- Distances from perch exceeding 4 km, or missing.
- Visibility rated as poor or unacceptable.
- Single sightings occurring at times when matching would have been impossible. See Section 2.3.1.
- Unlinked X, R, Y, and Z Duplicates. All of these are re-sightings of previously seen groups or single animals. Therefore, these exclusions do not reduce the number of captures from the perch.
- Chains not starting with a New or Conditional sighting.

- A handful of sightings where on-perch calculations of distance (using theodolite data and a hand-held calculator) were wildly inconsistent with post-hoc calculations during matching. In half these cases, an obvious single-digit typographical error was corrected; the other cases were discarded.

All results in this paper pertain to the clean dataset. Note that excluding cases while estimating detection probabilities does not have any more direct effect on abundance estimation. For the latter task, all the sightings can be used.

2.3 Analytical Methods

2.3.1 Chain Weighting

My analysis weights chains to account for differences in the opportunity to match sightings and the confidence in sightings and matches.

Availability Windows

I begin our discussion of weighting by introducing the concept of IO windows. A period of IO effort usually lasted about 4 hours, although longer and shorter times sometimes occurred. We call these periods 'IO windows'. For a four hour IO window, we define the first and last hours of the period to be the 'fringe'. The middle two hours are the window 'core'. For shorter windows, the core will be smaller or even zero. For longer windows, the core will be greater than two hours but the fringes are always one hour.

Based on average swim rates and the experience of observers and matchers, the opportunity to sight a whale that is matched to another sighting is nearly zero before and after one hour from that sighting. Notwithstanding this, the master matchers examined a much longer time frame (2.5 hours 'lookback'), yet very few sightings were matched beyond one hour apart.

Note that any unmatched sighting within the IO window core is available for matching during the entire time when it could potentially be matched. Sightings in the fringes of the IO window have less opportunity to be matched. A correction should be applied to prevent downward bias in the number of recaptures and corresponding downward bias in estimated detection probabilities.

Call the 2-hour period centered at a sighting (i.e., ± 1 hour) the sighting's 'availability window'. Then unmatched chains whose availability window is not wholly contained in an IO window might have been matched if IO effort had been expanded. Thus, the evidence such a sighting provides in favor of a non-match is weaker than an unmatched sighting with complete match availability. Weights are assigned to single sightings to reflect this. For a sighting at one perch, define its 'match availability weight' to equal the percent of the sighting's availability window that overlaps with any IO window(s). The simplest example is a case where one perch operates continuously and sights a whale at the exact instant that the other perch ends IO effort. Then this sighting has weight 0.5 because it can be matched backward to the last hour of the IO period but cannot be matched forward to the subsequent hour when the second perch is inoperative. Counterintuitively, a sighting from one perch that occurred outside an IO window may still have some opportunity to be matched: it may be matched to a sighting within an IO window whose boundary is within 1 hour of the outside sighting. Match chains are assigned availability of 1.

Conditional sightings

Another complexity concerns the treatment of chains (including single sightings) involving Conditional sightings. Zeh et al.(1991) and subsequent analyses have treated a Conditional sighting as half a New sighting when estimating detection probabilities and abundance. We too assign a 'Conditional weight' of 0.5 to any Conditional single.

Conditional sightings also occur in link chains. For example, consider a sequence of sightings like $N_1 \quad C_1 \rightarrow X_1$ where, importantly, the gap indicates that there is no link between the initial New whale and the subsequent Conditional sighting. The absence of a preceding link is compelled by the definition of the term Conditional, and indeed the earlier New whale is merely implicit in the sense that every Conditional whale has at least one New or Duplicate whale before it. Assigning equal odds for a Conditional being New or Duplicate, the above sequence of sightings can be interpreted in two ways. With probability 0.5 the Conditional has not been previously seen so the sequence is equivalent to $N_1 \quad N_1 \rightarrow X_1$. On the other hand, with probability 0.5 the Conditional is actually a re-sighting of some previous whale, in which case the sequence is equivalent to $N_1 \rightarrow X_1 \rightarrow X_1$ for some preceding N_1. Thus a Conditional link chain is also half a sighting.

Treating match chains with leading Conditional sightings at one or both perches is more complex–particularly because the chain must be weighted as a single unit rather than weighting the sightings from each perch separately. Nevertheless, we can apply analogous reasoning to show that we should use Conditional weights of 0.75 if one of the matched sightings is Conditional and 0.5 if both are.

Match uncertainty

Section 2.2.2 explained the Excellent, Good, and Average ratings the master matchers used to rate their confidence in each match. I will associate probabilities p_E, p_G, and p_A to the three types of matches to indicate the probability that the declared match is correct. In our baseline analysis, we will let $p_E = 1$, $p_G = 0.78$, and $p_A = 0.58$, with the latter two values chosen to be the midpoints of the respective confidence ranges used by the matchers. Sensitivity to these choices is explored in Section 2.4.2.

These match confidence probabilities pertain to matched sightings, not complete match chains. When one considers that each match connection in a match chain may or may not be correct, it is clear that any observed match chain could be a flawed observation of an underlying truth. Many possible changes to the match calls, in isolation or combination, create alternative configurations of the relevant sighting data, and each configuration can be associated with a corresponding probability weight. Specifically, the other configurations can be enumerated by considering all possible combinations of 'retaining' and 'breaking' declared

matches, and the probability of each resulting configuration can be derived using p_E, p_G, and p_A.

Consider the match chain $N_1 \overset{A}{\Leftrightarrow} N_2 \overset{G}{\Leftrightarrow} C_1$ where the letters above the match arrows indicate match quality ratings. Table 2.3 lists the possible alternative configurations for this match chain, and the corresponding probabilities. Note that when $p_E = 1$, as for our recommended analysis, Excellent matches are never broken.

For any chain or alternative configuration, define its 'match confidence weight' to be the probability of that configuration using the probabilities discussed here, such as the example in Table 2.3. For statistical modeling and analysis, each match chain is replaced by the set of possible configurations, each assigned its corresponding match confidence weight. Finally, I note that the original chain always receives the highest probability since all the match confidence probabilities exceed 0.5.

Replacing match chains with sets of chain configurations weighted by their probabilities may seem to be a bold departure from the raw data until one can witness the actual matching process. To a master matcher, such an approach seems not only sensible but essential. The

Table 2.3: Possible alternative configurations and probabilities for the match chain $N_1 \overset{A}{\Leftrightarrow} N_2 \overset{G}{\Leftrightarrow} C_1$ where the letters above the match arrows indicate match quality ratings. The original chain is most likely and is listed first.

$N_1 \Leftrightarrow N_2 \Leftrightarrow C_1$	$p_A p_G$
$N_1 \quad N_2 \Leftrightarrow C_1$	$(1-p_A)p_G$
$N_1 \Leftrightarrow N_2 \quad C_1$	$p_A(1-p_G)$
$N_1 \quad N_2 \quad C_1$	$p_A p_G$

master matchers can spend up to 1 hour agonizing over just one potential match decision and debating match confidence ratings as they pore over the data. When they decide to declare a match and rate it as Good, they are relying on that confidence rating and its quantitative description (e.g., 'at least 66% confidence') to express their very real uncertainty about the call.

Summary of weighting

In summary, there are three factors that can affect how a chain is weighted. First, the concept of availability is used to adjust sightings made when there was a reduced recapture opportunity due to the timing of IO and single perch effort. This must increase detection probability estimates. Second, the traditional half-weighting of Conditional sightings is applied to single and link chains, with analogous extensions to match chains involving one or more Conditional sightings in a match. This weighting should be roughly neutral with respect to detection probabilities. Third, match chains are weighted by their match confidence. This has a negative effect on detection probabilities. The overall weight for a chain is defined to be the product of these three weights, with the match confidence weighting implemented using the alternative configuration method in Section 2.3.1.

2.3.2 Group Size Consistency

Each recorded sighting may be of a single animal or a group. These groups are not whale pods in the conventional sense. Except for cow-calf pairs, migrating bowheads appear to

have only weak and probably brief allegiance to any aggregation but their associations are not purely random (Zeh et al., 1993). When the number of whales in a sighted group ('group size') is seen to vary along a chain, this may be attributable to whales joining or leaving a chain, and/or to variation in detection. A poll of observer opinions after the survey finds divided views, with some observers tending to attribute group size inconsistencies to a failure to detect group members, and others tending to attribute it to whales ('transients') joining or leaving groups. Group size inconsistencies are relatively common among match chains. Indeed, 574, 109, 16 and 5 match chains had (maximal) group size inconsistencies of 0, 1, 2, and 3 whales, respectively. This corresponds to an inconsistency rate of 18.5% for match chains. The most common inconsistency is a group size 1 matched to a group size 2.

Along with distance, group size is a dominant factor influencing detection probabilities. The fact that animals may join or leave a chain during the period that the chain is (repeatedly) sighted is of paramount importance because estimation of detection probabilities fundamentally relies on counts of captures and recaptures. When group sizes in a chain vary, these counts are uncertain. Thus, resolving group size inconsistency is a major dilemma that must be addressed in the analysis. It is clear that analysis methods that recognize these potential fleeting group allegiances better reflect the behavioral processes that generate the observed data.

We must also recognize that whales are not individually sighted. For the bowhead survey, capture and recapture events pertain to groups, not individuals. Groups are sighted and recorded on the perches, and groups are plotted and matched during the post-hoc matching phase. All the approaches I will describe below embrace this fact: I will estimate detection probabilities of groups. However, this does not require me to ignore group size inconsistency, nor should I. Nevertheless, the first approach I apply for dealing with group size inconsistency is to ignore it. I label this method 'Inconsistent'. This approach amounts to assuming

that there are never transient departures or joinings to a group. Next, I introduce another strategy I call 'deconstruction' to model the possibility of transient group affiliation. With this approach, I may isolate certain possible transients, thereby decomposing a match chain into several parts.

To explain deconstruction, I first define a sub-chain of an original chain to be any chain of connected sightings (links and/or matches) from the original chain with consistent group size that can be obtained by reducing the group sizes at any chosen sightings in the original chain. A chain can be decomposed into a unique set of group size consistent sub-chains by sequentially removing the longest chains of each possible group size, starting with the largest group size and working downward. For example, the sub-chains of a chain having group sizes (2,1,1,2,2) in that order are the chains (1,1,1,1,1), (1,0,0,0,0), and (0,0,0,1,1) where '0' indicates absence of chain membership.

Deconstruction identifies sub-chains of link and match chains that can be reasonably attributed to transients. The following is a deconstruction method based on mode chain group size.

1. Calculate the mode m of group sizes for sightings comprising the chain. (Suppose here that we are beginning with a match chain.) The mode is chosen to be a reasonable estimate of the true group size.

2. Replace all group sizes less than m in the chain with m.

3. Set aside the sub-chain having consistent group size m.

4. Subtract m from the group size of each sighting in the original chain. Remove from the chain any sightings with group size 0.

5. Break all remaining matches in the original chain, if any. Although we break these matches, the remnant group sizes are unchanged.

6. What remains, if anything, will be a collection of isolated (i.e., unlinked) groups and link chains with sightings, possibly with inconsistent sizes.

7. Repeat the above steps separately for each link chain. No repetition is needed for any single group because it already has consistent group size. Continue iterating this process within each sub-chain, and for every sub-chain, until what remains is a set of chains all having internally consistent group sizes, including the original chain with the mode group size.

An elaborate example of deconstruction is presented by the chain having the sequence of group sizes (2,3,1,1,1,3,3,4,4). The longest sub-chain we can extract with consistent group size is (1,1,1,1,1,1,1,1,1), assuming we define the mode to be 1. Subtracting these whales leaves (1,2,0,0,0,2,2,3,3), which must be further deconstructed. The next sub-chain is (0,0,0,0,0,2,2,2,2) since we attack the right hand remnant first. Continuing in this fashion yields the remaining components: (1,1,0,0,0,0,0,0,0), (0,1,0,0,0,0,0,0,0), and (0,0,0,0,0,0,0,1,1). This example is for illustration; most real chains are far simpler.

It is possible that the subtraction step #4 may leave match connections between sightings of non-zero group sizes. The reason that these matches are broken in step #5 is because the original chain represents one recapture event. Failing to break remnant match sub-chains would count the match event more than once.

The mode group size may be a tie between two values. Accordingly, I define two alternative methods: 'Decon↓' rounds modes downward, and 'Decon↑' rounds upward.

Deconstruction both subsumes and creates transients. In step #2, whales are added to match chains to represent true group members that were undetected. In steps #4-6, the extra chains (if any) produced represent whales having transient memberships to the true group recaptured. This approach is a compromise between assuming that every chain represents

the largest possible group with no transients and assuming that every chain corresponds to the smallest group size seen with all remaining whales being transients. Finally, recall that group size is a direct observation, not a derived variable. The observers are making an explicit statement about what they see. An appealing aspect of deconstruction is that field observations are taken at face value rather than second-guessing the trained observers after the fact.

The choice between the three methods (Decon↑, Decon↓, and Inconsistent) is a choice about how to count captures and recaptures. Each of the three methods makes a different assumption about transients–from none to many–and each of these assumptions leads to the addition of some number of single chains (or none) to represent transient whales. This is unrelated to how whales are counted for the purpose of making an abundance estimate. It is also unrelated to how group size is assigned to a chain as its covariate value for estimation of detection probabilities. As described in Section 2.2.3, the group size covariate value is defined to be the size observed at the first recapture event. The same group size covariate value is assigned to any sub-chains of the original chain. This is the most appropriate choice because detection probabilities relate to detection of groups.

2.3.3 Detection Probability Estimation

I adopt the model of Huggins (1989) for capture-recapture estimation for closed populations. This model assumes that captures at each perch are independent and that the catch history is therefore multinomial for each individual. To form the likelihood, the model conditions on the total number of groups detected. We constrain our analysis to assume that probabilities are equal among the perches. The detection probability for a group is allowed to depend on covariate observations for that group. The effect of covariates is the primary focus of our analysis.

Models are fit using the MARK software (White and Burnham, 1999), using the RMark interface (Laake, 2011). However, all of the analyses described here include consideration of weighted observations. A method for weighted fitting of the Huggins model using MARK/RMark is given next.

Fitting weighted capture - recapture models with MARK

Models are fit using the MARK software (White and Burnham, 1999), using the RMark interface (Laake, 2011). However, all of the analyses described here (except G_0) include consideration of fractional whales, which is equivalent to weighted observations. This section provides more details and describes a method for obtaining the results of the weighted analysis using the MARK program, thereby circumventing the need to develop a customized estimation procedure. The key results are given in equations (1), (2), and (3). Although I refer to MARK results below, it is easiest to implement these ideas using RMark, from which the necessary quantities can easily be extracted and manipulated.

Weighted observations

Suppose we wish to assign the ith outcome (i.e., the catch history for a single, link, or match chain) a weight w_i for $i = 1, \cdots, n$ where n is the sample size and, without loss of generality, $0 < w_i \leq 1$. Our method temporarily replaces the original, unweighted dataset with a larger dataset that replicates the ith observation r_i times, where $r_i = Rw_i$ and R is chosen to be the smallest integer such that r_i is an integer for all i . Thus the total sample size in the expanded dataset is $n_R = \sum_{i=1}^{n} r_i = R \sum_{i=1}^{n} w_i$. If the desired weights are rounded to the first decimal place,

then $R \leq 10$. In an analysis where the presence of conditional whales requires $w_i = 1, 0.5, 0.75$ for all i (see below) then $R = 2$ or $R = 4$ depending on whether any $w_i = 0.75$.

There are several instances where weighting may be used. Consider Conditional sightings, i.e., sightings where the observer cannot determine whether the whale is New or a Duplicate. Zeh et al. (1991) and subsequent analyses have treated a single Conditional whale as half a New whale when estimating detection probabilities and abundance. Thus we assign weights of 0.5 to conditional sightings here.

Conditional whales in link and match chains must also be addressed. For example, consider a sequence of sightings like $N_1^1 \quad C_1^1 \rightarrow X_1^1$ where, importantly, there is no link between the initial New whale and the subsequent Conditional sighting. The absence of a preceding link is compelled by the definition of the term Conditional, and indeed the earlier New whale is merely implicit in the sense that every Conditional whale has at least one New whale at some earlier time. Assigning equal odds for a Conditional being new or duplicate, the above sequence of sightings can be interpreted in two ways. With probability 0.5 the Conditional has not been previously seen so the sequence is equivalent to $N_1^1 \quad N_1^1 \rightarrow X_1^1$. On the other

hand, with probability 0.5 the Conditional is actually a resighting of some previous whale, in which case the sequence is equivalent to $N_1^1 \rightarrow N_1^1 \rightarrow X_1^1$ for some preceding N_1^1.

Assume the simplest possible capture-recapture model with the probability of capture at each perch and the probability of recapture all equal to p. Then the first possibility represents two whales, thereby contributing p^2 to the likelihood. The second possibility represents a single whale contributing p to the likelihood. In the next section we develop a method allowing one to assign a weight of 0.5 to a Conditional whale by letting it contribute $p \times p^{\frac{1}{2}} = p^{\frac{3}{2}}$ to the likelihood. This approach is also consistent with the standard principle of weighting observations in statistical models: that a random variable having half the weight is equivalent to having twice the variance.

The weight of 0.5 is used for Conditional sightings that are unconnected or which lead a link chain. Treating match chains with leading Conditional sightings at one or both perches is more complex - particularly because the chain must be weighted as a single unit rather than weighting the sightings from each perch separately. Nevertheless, we can apply analogous reasoning to show that we should assign such chains a weight of 0.75 if one perch reported Conditional and 0.5 if both perches reported Conditional.

Weighting and MLEs for exponential families

I now discuss my weighting approach in the context of exponential families. For exponential family distributions with weighted observations, there is a useful relationship between certain likelihoods. In the simplest and most generic case, a density in the exponential family can be written as

$$f(x_i|\theta) = \exp\{\frac{x_i\theta - b(\theta)}{\phi/w_i} + c(x_i, \phi/w_i)\}$$

where ϕ is a fixed dispersion parameter and w_i is a known weight. The exponential family includes many familiar distributions including the Gaussian, binomial, and multinomial distributions; the latter two are directly relevant for capture-recapture models.

Consider the simplest (i.e., null) Huggins Huggins (1989) model with one recapture opportunity. Each weighted observation in the dataset (e.g., each catch history) contributes a term to the likelihood function used to estimate the model. Denote the corresponding log likelihood contribution as

$$l_i(\theta|x_i) = w_i A_i + c(x_i, \phi/w_i)$$

$A_i = (x_i\theta - b(\theta))/\phi$, $\theta = \log(p/(1-p))$, $b(\theta) = -\log(2 + \exp\{\theta\})$, $\phi = 1$ and x_i equals 1 if the catch history does not include a recapture and 0 otherwise. For an i.i.d. sample, the overall log likelihood is the sum of such contributions, namely

$$l(\theta|X) = \sum_{i=1}^{n} l_i(\theta|x_i) = \sum_{i=1}^{n} w_i A_i + q$$

where $X = x_i, \cdots, x_n$ represents the entire dataset and q is a constant that does not depend on θ.

Let X_R represent the dataset where each x_i is replicated r_i times and *there is no weighting* of the observations within X_R. Let x_{i,j_i} for $j_i = 1, \cdots, r_i$ denote these replicates of the ith case in X. Then the contribution to the overall *unweighted* log likelihood $l_R(\theta|X_R)$ for x_{i,j_i} is

$$l_{R,i,j_i}(\theta|x_{i,j_i}) = A_i + c(x_{i,j_i}, \phi)$$

and the total contribution associated with the ith catch history is

$$\sum_{j_i=1}^{r_i} (A_i + c(x_{i,j_i}, \phi)) = r_i A_i + q_{R,i}$$

where $q_{R,i}$ is a constant that doesn't depend on θ. It follows that the overall log likelihood given by the unweighted, replicated dataset is

$$l_R(\theta|X_R) = R \sum_{i=1}^{n} w_i A_i + q_R^*$$

where q_R^* doesnt depend on θ.

Our key results follows from the fact that terms not involving θ are irrelevant when computing the score functions from $l(\theta|X)$ and $l_R(\theta|X_R)$. Specifically,

$$\frac{dl(\theta|X)}{d\theta} = \frac{dl_R(\theta|X_R)/R}{d\theta}$$

Thus the MLE for the weighted likelihood equals the maximizer of the replicated likelihood, i.e.,

$$\hat{\theta}_{MLE} = \hat{\theta}_R \tag{2.1}$$

Although this discussion is presented for unidimensional θ, the analogous results for vector parameters are obvious. Because it is based on the multinomial distribution, the results above pertain to the general Huggins (1989) model, including the case when detection probabilities are modelled to depend on a collection of covariates. The covariates inherent in the model lead one to express the parameter vector in the conditional likelihood as a function of the coefficients in the linear predictor portion of the model. The standard maximum likelihood assumptions are also implicit above and in what follows.

Computing AIC_c differences

Consider the comparison of two fitted models, which we will represent as $\hat{\theta}_1$ and $\hat{\theta}_2$, recognizing that the two models may have different parameter sets. Define

$$\Delta = l(\hat{\theta}_1|X) - l(\hat{\theta}_2|X) \qquad \Delta_R = l_R(\hat{\theta}_1|X_R) - l_R(\hat{\theta}_2|X_R)$$

then

$$\Delta = \Delta_{R/R}$$

In other words, the log likelihood difference between the two models fit to the weighted data can be calculated from the log likelihood difference between the two models fit to the replicated data. Therefore, the AIC_c difference (Akaike (1974); Burnham and Anderson (1998)) between two weighted models can be expressed as

$$AIC_c(\hat{\theta}_1, \hat{\theta}_2) = -2\Delta_R/R + K_1 - K_2 \tag{2.2}$$

where $K_i = 2k_i n/(n - k_i - 1)$ for $i = 1, 2$ indexing the two models and k_i denoting the numbers of parameters therein. Thus, $AIC_c(\hat{\theta}_1, \hat{\theta}_2)$ for the weighted analysis of the original data X can be calculated using the results of applying MARK to the unweighted replicated dataset X_R, and then adjusting the results using a collection of known constants.

Estimated standard errors and confidence intervals

The asymptotic variance of a maximum likelihood estimator is $\text{var}\{\theta\} = I^{-1}(\theta_0)/n$ where

$$I(\theta_0) = E\left(\frac{dl(\theta)}{d\theta}\right)^2\Bigg|_{\theta=\theta_0} = -E\left(\frac{d^2 l(\theta)}{d\theta^2}\right)\Bigg|_{\theta=\theta_0}$$

and θ_0 denotes the true value of θ. Consider the likelihood corresponding to X, namely $l(\theta)$. In this case,

$$I(\theta) = -E\sum_{i=1}^{n} w_i A_i''(\theta)$$

Next consider $l_R(\theta)$ originating from the unweighted, replicated dataset X_R. Then

$$I_R(\theta) = -E\left(\frac{d^2 l_R(\theta)}{d\theta^2}\right) = -E\sum_{i=1}^{n} R w_i A_i''(\theta)$$

The sample sizes for X and X_R are n and $n_R = \sum_{i=1}^{n} r_i$ respectively. Thus it is straightforward to show that

$$\text{var}\{\hat{\theta}\} = R\ \text{var}\{\hat{\theta}_R\}$$

Conveniently, $\text{var}\{\hat{\theta}_R\}$ is output from RMark. An intuitive interpretation of this result follows from noting that $R = \frac{n_R}{n\bar{w}}$ where $\bar{w}$ is the mean weight. To obtain $\text{var}\{\hat{\theta}\}$ from $\text{var}\{\hat{\theta}_R\}$ we must scale up $\text{var}\{\hat{\theta}_R\}$ by n_R/n to compensate for the large size of X_R, then scale this up further by $1/\bar{w}$ to adjust for the decreased information contained in X due to the down-weighting in some cases.

Model selection is carried out using AIC_c as an objective function (Akaike, 1974; Hurvich and Tsai, 1989). The simplest model within 2.0 units from the minimum is selected (Burnham and Anderson, 2002). For our data, this was essentially equivalent to choosing the model with the lowest AIC_c. I use a stepwise approach (generally forward selection) considering additive and two-way interactive effects, with backward elimination and tangential explorations conducted when it appears they might be helpful. Model selection was especially focused on the variables generally found to be important during our exploratory data analysis with the 2010 dataset: visibility, distance, lead condition and group size.

Automated model selection methods and model averaging are not used. An alternative model selection criterion like BIC (Schwarz, 1978) could be used to penalize model complexity more severely, but in my case the selected models are so simple and the retained effects so clearly scientifically justifiable that such an option is not pursued. The same reasons eliminate the need for automated model selection methods. There are several reasons not to average

Table 2.4: Detection probability estimates using the Decon↓ method for the simple and complex models from analysis of the 2011 survey data. The top portion of the table shows sample-weighted means of detection probability estimates from each model, and the mean that would be obtained if the parameter estimates from our preferred simple model were applied to the non-deconstructed dataset. Predictions for single whales and groups at several distances are also shown in the bottom half of the table. These estimates are also based on the simple model and *Decon* ↓.

Quantity	Estimate	Std. Error	Confidence Interval
Sample Mean, Simple Model	0.495		
...to orig. data	0.492		
Sample Mean, Complex Model	0.494		
Fit for single whale, 3000 m	0.377	0.023	(0.332, 0.423)
Fit for single whale, 2000 m	0.475	0.018	(0.440, 0.510)
Fit for group, 1000 m	0.645	0.032	(0.583, 0.709)

models here in the manner of Burnham and Anderson (2002). Most importantly, the model fitting process shows that there is very little model uncertainty in the senses that (i) there were few selection choices presenting ambiguous decrements near 2.0, (ii) the same model was selected for each approach, and (iii) fits to alternative models yielded extremely similar parameter estimates and predictions in most cases. Finally, note that the three methods for addressing group size inconsistency use different datasets of different sizes, thereby yielding incomparable AIC_c values.

2.4 Results

2.4.1 Main Findings

All three approaches favor the same model: additive effects for distance and group size. There is no evidence of an interaction. This model will be referred to as the 'simple model' hereafter and it is the one I recommend. I also fit a larger model (the 'complex model') which includes additive effects for visibility, distance, lead condition and group size. These are variables that observers consider likely to affect detection probabilities, and these pre-

dictors have often been found important in preliminary research with the 2010 data. Again, no interactions were necessary.

The sample-weighted mean detection probabilities for the two fitted models are shown in Table 2.4. These employed the Decon↓ method, which is the one I recommend to address group size inconsistency. My recommendation of Decon↓ is motivated by my belief that this approach best represents bowhead behavioral patterns and likely observer tendencies.

Table 2.4 also shows the sample-weighted mean the would be obtained when the parameter estimates from the Decon↓ simple model are applied to the non-deconstructed data. This is what would be done when forming an abundance estimate. The result differs very little.

Although useful as overall summaries, the sample-weighted means cannot reveal whether the methods identify any differing covariate effects. Consequently, I also show in Table 2.4 the predicted detection probabilities for groups at three distances and two group size categories (group size = 1 and > 1). Since large whale groups near the perch are the easiest to detect, the estimated detection probabilities for groups > 1 at 1000 m is largest. Single whales are more difficult to detect, especially at greater distances, so the corresponding estimate at 3000 m is smallest. Group sizes of 1 predominate in the dataset and 2000 m is very close to the dataset average, so the detection probability estimate shown for single whales at 2000 m is essentially the central value. The standard error for this case is also smallest because of the greater frequency and centrality of such cases.

General detection probability estimates can be expressed via the following equations. Let d and s denote the distance and group size of a sighting, respectively. Then the detection probability p can be calculated as

Table 2.5: Sample-weighted mean estimated detection probabilities for the three methods of addressing (or ignoring) group size inconsistency. Results are shown for both fitted models (simple and complex).

Method	Simple Model	Complex Model
Decon↓	0.495	0.492
Inconsistent	0.509	0.509
Decon↑	0.506	0.504

$$p = \frac{\exp\{\beta_0 + \beta_1 d + \beta_2 I(s=1)\}}{1 + \exp\{\beta_0 + \beta_1 d + \beta_2 I(s=1)\}} \tag{2.3}$$

where $I(s = 1) = 1$ for group size = 1 a single whale and 0 for group size > 1. For our recommended model (simple) and group size approach (Decon↓), the parameter estimates are $\widehat{\beta}_0 = 1.0139$, $\widehat{\beta}_1 = -4.0251 \times 10^{-4}$, and $\widehat{\beta}_2 = -0.3104$.

Our results confirm that increasing distance from the perch has a very strong negative effect on detection probabilities, which (for singles) are as high as 67% near the perch and 29% at 4000 m. The results also indicate that, compared to a single, the odds of detecting a group are increased by a multiplicative factor of $\exp\{.3014\} = 1.35$.

2.4.2 Other Results

My analysis also finds that detection probabilities vary significantly between observers. Although it is common to find observer effects in studies like mine, there is one surprising aspect here. On the perch, search success is a team effort, yet it is only the operator of the theodolite whose identity is recorded for the sighting. The theodolite operator is not the only–nor the primary–person discovering whales, particularly considering the narrow field of view of the device. Moreover, team memberships were continuously remixed so that any

particular theodolite operator used the device with many different combinations of teammates. For these reasons, it is difficult to explain the observer effects. Despite an extensive analysis not reported here, I have found no correlation between observer effects and the level of observer experience (e.g., total hours of effort during the survey) or mean whale passage rate during the observers effort.

My current plan for the future abundance estimate is to ignore observer effects. Such unmodeled extra heterogeneity will tend to cause a downward bias in abundance estimates using standard capture-recapture abundance models (Carothers, 1973, 1979; Otis et al., 1978; Seber, 1982; Pollock et al., 1990; Hwang and Chao, 1995; Pledger and Efford, 1998; Pledger and Phillpot, 2008). Bias is undesirable, but if it is unavoidable then downward bias in abundance is preferable to upward for the sake of conservative management. Since the observer heterogeneity is substantial in our case, any resultant bias can be viewed as a safety guard against the detection probability bias corrections discussed in this paper. In other words, if one does not accept my bias corrections but cannot supply replacements, they may be considered to be 'canceled out' by uncontrolled observer heterogeneity.

I use sensitivity analyses to investigate some of the choices made in my analysis. First, I can ask whether the methods for addressing group size inconsistencies strongly influence the results. Table 2.6 shows that the three methods produce quite similar sample-weighted mean detection probabilities. This is not the whole story, however. The estimated detection probability for groups > 1 obtained from the Decon$\downarrow$ approach is roughly 0.07 and 0.10 lower than the ones obtained from the Decon$\uparrow$ and Inconsistent methods, respectively. Although this may seem like a large difference, consider that less than 20% of the data have group sizes exceeding 1. The net effect of, say, a 0.085 decrease in group effect is a sample-average

detection probability reduction of only $0.20 \times 0.085 \approx 0.017$, which is consistent with my reported findings.

The second sensitivity analysis investigates the choice for (p_E, p_G, p_A). Recall that the master matchers were instructed to adopt $(0.90, 0.66, 0.50)$ as their decision thresholds. In my trials, I evaluate high, medium and low values. The medium values are those used in the main analyses, namely $(1.00, 0.78, 0.58)$ where the latter two values are the midpoints of the decision boundaries. The high set is $(1.00, 0.89, 0.65)$, and the low set is $(0.90, 0.65, 0.50)$. Table 2.6 shows estimates obtained using each of these choices. Although I recommend the Decon$\downarrow$ approach, results in Table 2.6 employ the Decon$\uparrow$ approach since it is the one that yields the more central marginal mean. I find that varying (p_E, p_G, p_A) operates in the expected direction: higher matching confidence probabilities yield higher detection probability estimates. However, the magnitude of the changes in the estimates is not extraordinary, especially considering that the high and low scenarios are the absolute extremes of what could be chosen. In other words, observers can't be more than 100% certain, nor would they declare a match that they believe is more likely not one. Overall, my results clearly support my medium choice as a reasonable one.

2.5 Discussion

It is worthwhile reviewing the big picture surrounding this work. The 2010 survey was not successful for the purpose of estimating abundance, but it was very successful as an independent observer experiment. The 2011 survey was nearly unprecedentedly successful for obtaining sighting counts, and also produced much additional IO data. Together, the IO data from these two surveys clearly present the best opportunity in the history of bowhead surveys to estimate ice-based detection probabilities.

Table 2.6: Results of sensitivity analysis using high, medium, and low match confidence weights with the Decon↑ method for addressing group size inconsistency. Although Decon↑ is not our recommended approach, it is used here because it produces intermediate results with the medium weights in Table 2.5.

Quantity	Estimate	Std. Error	Confidence Interval
Single, 3000 m			
$(p_E, p_G, p_A) = (0.9, 0.66, 0.5)$	0.323	0.023	(0.280, 0.370)
$(p_E, p_G, p_A) = (1.0, 0.78, 0.58)$	0.377	0.023	(0.327, 0.418)
$(p_E, p_G, p_A) = (1.0, 0.89, 0.65)$	0.391	0.024	(0.346, 0.438)
Single, 2000 m			
$(p_E, p_G, p_A) = (0.9, 0.66, 0.5)$	0.416	0.018	(0.380, 0.452)
$(p_E, p_G, p_A) = (1.0, 0.78, 0.58)$	0.474	0.018	(0.439, 0.509)
$(p_E, p_G, p_A) = (1.0, 0.89, 0.65)$	0.494	0.018	(0.459, 0.529)
Group, 1000 m			
$(p_E, p_G, p_A) = (0.9, 0.66, 0.5)$	0.654	0.034	(0.585, 0.717)
$(p_E, p_G, p_A) = (1.0, 0.78, 0.58)$	0.717	0.031	(0.654, 0.773)
$(p_E, p_G, p_A) = (1.0, 0.89, 0.65)$	0.734	0.029	(0.675, 0.790)

Early during the 2010 survey it became apparent that sufficient coverage of the migration would be unlikely, so effort was overwhelmingly dedicated to ensuring that we obtained a large amount of high quality independent observer data. For the 2011 survey, the critical need became obtaining equally good count data. Thus in 2011 we prioritized continuous effort from one perch instead of more limited effort from both perches.

This strategy is reflected in Figure 2.4 which shows the periods of single-perch and IO effort during the survey season in each year. Heavy lines indicate IO periods, with the boundaries indicated with an 'x'. Light lines indicate periods where only one perch was operating. In 2010, virtually all effort was for IO, and in 2011 coverage with at least one perch was nearly continuous throughout the portion of the season when the largest numbers of whales pass Barrow.

With respect to statistical methodology, perhaps the greatest focus in this chapter has been on bias correction. The challenge presented by the bowhead survey is that the animals,

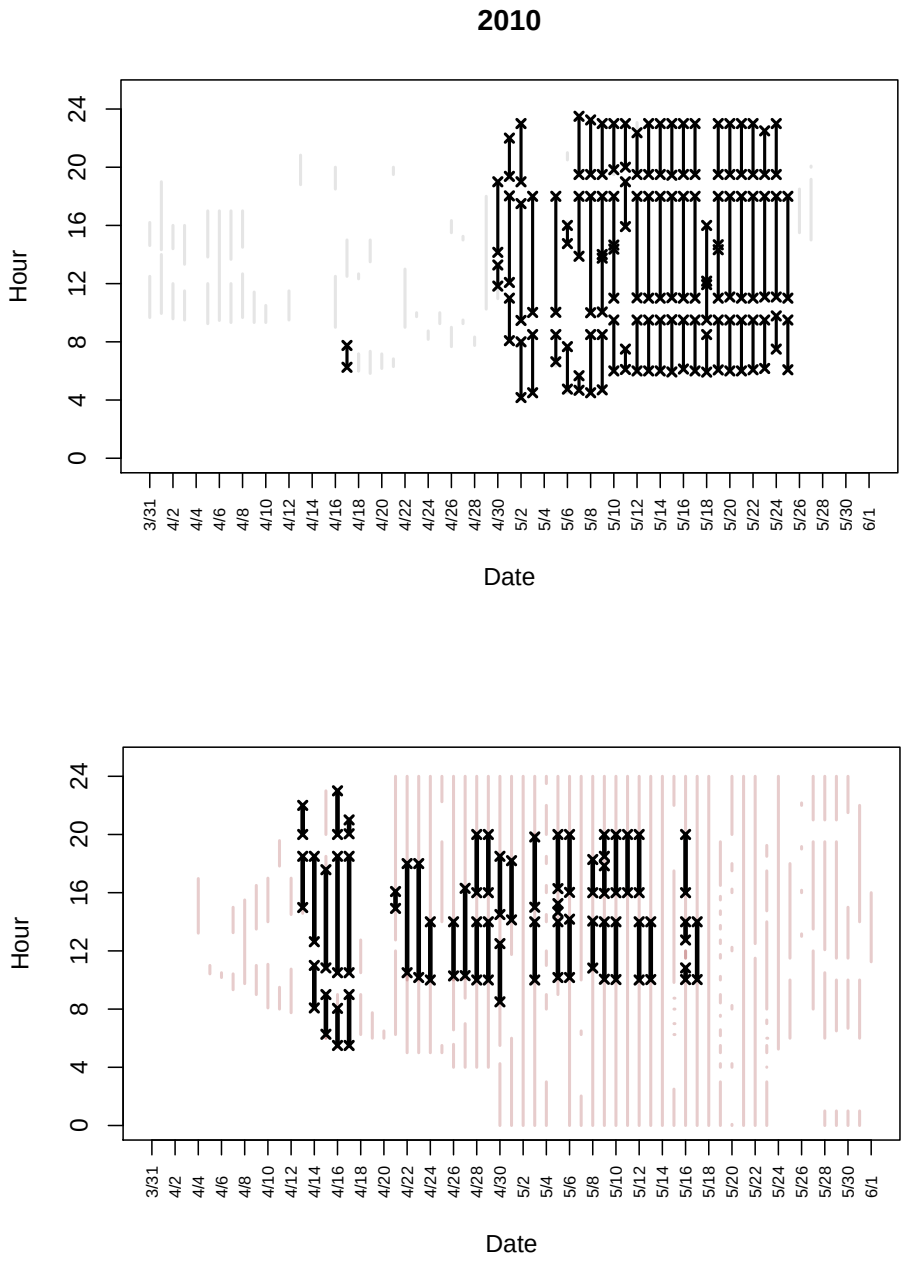

Figure 2.4: IO effort for 2010 and 2011 survey years. Each vertical line represents one day. Black lines represent time with 2-perch IO effort. The boundaries of these periods are indicated with an 'x'. Gray lines indicate periods when only one perch was used.

environment, and survey protocol force us to work with data carrying obvious biases. In this chapter, I have chosen to confront these biases directly and quantitatively. To balance the possibilities of partial group detection and transient group membership, I introduce the Deconstruction methods. To correct for missing opportunity for matches, I weight cases based on match availability. Finally, to reflect the master matchers' doubt when declaring matches, I consider weighted possibilities.

In my view, these three sources of bias are so clearly present and possibly substantial that it would be a mistake to fail to account for them. In practice, I find that the Deconstruction adjustments have limited effect, while the other two corrections have greater impact, but in opposing directions. Overall, the net effect of our corrections is to reduce detection probability estimates compared to a naive uncorrected analysis of the raw data.

CHAPTER 3

EXPLICIT MODEL FOR UNCERTAINTY ABOUT CAPTURE, RECAPTURE AND ANIMAL IDENTITY

3.1 Introduction

This chapter describes my second approach to estimate detection probability in the presence of uncertain, captures, recaptures and animal identity. Before we can move forward, some simplying assumptions need to be made about the data produced by the experimental design.

Let perch 1 denote the location for the first capture opportunity and let perch 2 denote the location for the second capture opportunity. Perch 1 and perch 2 are operational simultaneously at nearby positions along a migratory corridor. We assume that each animal has the same opportunity to be captured at perch 1 and perch 2 at the same time due to the proximity of the perches relative to the visual field and the animal dispersal. Let I represent the number of independent observational time intervals in which perch 1 and perch 2 are simultaneously operational.

The use of unique identifying tags from perch 1 to identify animals recaptured at perch 2 is impossible for a situation like the whale survey where there is no contact with the animals. Instead, perch 1 identifies and records unique identifying labels for each sighted animal as it migrates past the perch. At the same time, perch 2 identifies and records unique identifying labels for each sighted animal as it migrates past the perch. Since capturing at perch 1 and perch 2 occurs simultaneously, it is unknown which animals were recaptured, i.e, seen at both perch 1 and perch 2 so the identification of recaptures must be done post hoc. It is important to note that perch 1 and perch 2 produce their own independent data records and that the researchers at perch 1 and perch 2 may or may not have sighted the same animals.

After the study is complete (i.e., each perch has gathered their own independent data), researchers determine which animals, if any, were sighted at both perches. Define a "match" to be the researcher's decision that a sighting at one perch is the same animal as a sighting at the other perch. A non-match is the researcher's decision that any sighting from either perch was the only time that animal was sighted during the experiment. There are a few important details to note. First, a sighting cannot be matched to more than one other sighting from the other perch. If this occurs, it needs to be revisited by the researchers and fixed. Second, for all sightings at both perches the group size associated with each sighting is assumed to be one. Third, unlike the data described in Chapter 2, the data described here have no link chains, i.e., each sighting at each perch is assumed to have been the only time that animal was sighted at that perch.

The researchers subject the two independent sighting data records to J independent matching decision processes. That is, the researchers make match/non-match decisions on the data J independent times. This may be done by the same set of researchers employing waiting periods sufficiently long to ensure that their repeated efforts are essentially independent, or by J sets of independent researchers using the same standards and procedures. In either case, it is clear that there will be matching decision errors and inconsistencies even though the researchers used their best expert judgement.

3.2 Notation and Definitions

A large set of definitions and notation will be needed to proceed. Suppose that $w_{1,i}$ unique animals were sighted at perch 1 in time interval i and $w_{2,i}$ unique animals were sighted at perch 2 in time interval i. From now on, assume without the loss of generality that $w_{1,i} \leq w_{2,i}$.

Because the experimental design prevents knowledge of the true capture history, i.e., which animals were sighted at both perches and which were sighted only once, define $\mathbb{E}_i$ be the set of all hypothetical recaptured sightings given $w_{1,i}$ and $w_{2,i}$ for time interval i. Each element in $\mathbb{E}_i$ represents one possible underlying, unobservable set of capture histories for the entire sightings dataset. Let e_{i_k} denote an element in $\mathbb{E}_i$ and K_i be the number of elements in $\mathbb{E}_i$. Hereafter I will refer to the elements in $\mathbb{E}_i$ as configurations.

For each configuration in $\mathbb{E}_i$, several quantities can be defined. In the following, it is important to draw a distinction between individual animals and sightings of animals. A duplicate individual is an individual that is sighted at both perch 1 and perch 2. Denote the number of duplicate individuals in configuration e_{i_k} as r_{i_k}. A duplicate sighting is one of the two sightings of a duplicate individual. A complimentary sighting is the other sighting of a duplicate individual. For example, let sighting A be a duplicate sighting at perch 1 from individual AB. Let B be a duplicate sighting at perch 2 from the same duplicate individual AB. Then A and B are each other's complimentary sighting. Next, a single individual is an individual that is sighted at either perch 1 or perch 2 but not both, and a single sighting is the sighting of a single individual. Finally, let d_{i_k} be the number of duplicate sightings in configuration e_{i_k} and let s_{i_k} be the number of single sightings given e_{i_k}. Note that d_{i_k} and s_{i_k} can be different for different configurations.

Everything defined previously pertains to one time interval at a time. More set notation is needed to describe the collection of this information. Let $\mathbb{W}_1$ be the set of all $w_{1,i}$ for all $i \in I$ and let $\mathbb{W}_2$ be the set of all $w_{2,i}$ for all $i \in I$. Let $\mathbb{C}$ be the set of all possible combinations of configurations for all $i \in I$ and for all $k \in K_i$. Let L be the number of sets of combinations in $\mathbb{C}$. Denote c_l as an element of $\mathbb{C}$. That is, c_l represents one set of combinations of configurations for all $i \in I$, $k \in K_i$, and $l \in L$. Additionally, let c_{l_i} be an element in c_l.

Table 3.1: Configurations in $\mathbb{E}_1$ for Example 3.1

Configuration 1 (e_{1_1})		**Configuration 2** (e_{1_2})		**Configuration 3** (e_{1_3})	
Sighting	Complimentary	Sighting	Complimentary	Sighting	Complimentary
A	none	A	B	A	C
B	none	B	A	B	none
C	none	C	none	C	A

The ith element in c_l, c_{l_i}, is the e_{i_k} configuration for time interval i in c_l. Finally, let s_l be the number of single sightings given c_l and let d_l be the number of duplicate sightings given c_l.

Example 3.1

A very simple example is used to help clarify of all this notation. Suppose during time interval $i = 1$, only a single animal was observed at perch 1. Denote this sighting as A. At perch 2 two sightings were observed. Denote these sightings as B and C. It is unknown whether A is B's complimentary sighting, C's complimentary sighting, or neither. Now we consider what configurations are in $\mathbb{E}_1$ given $w_{1,1} = 1$ and $w_{2,1} = 2$. The first possibility is that sightings A, B and C are of all unique animals. In this case, animal A had the opportunity to be seen at both perches but was only seen at perch 1, and animals B and C both had the opportunity to be seen at each perch, however they were only seen at perch 2. The second possibility is that sightings A and B are the same animal, AB, i.e., they are each other's complimentary sighting, while animal C had the opportunity to be spotted at both perches but was only seen at perch 2. The last configuration is the same as the second with sightings B and C switching roles. A summary of $\mathbb{E}_1$ is shown in Table 3.1.

Let's take a closer look at e_{1_2}. In all, there are three sightings: A, B and C. Now A and B are the sightings of same animal and hence represent a duplicate individual, therefore making both A and B duplicate sightings. Animal C was only sighted once and hence is a single

Table 3.2: Configurations in $\mathbb{E}_2$ for Example 3.1

Configuration 1 (e_{2_1})		**Configuration 2** (e_{2_2})		**Configuration 3** (e_{2_3})	
Sighting	Complimentary	Sighting	Complimentary	Sighting	Complimentary
D	none	D	E	D	F
E	none	E	D	E	none
F	none	F	none	F	D

individual, thus making C a single sighting. Thus, $d_{1_2} = 2$ and $s_{1_2} = 1$. By similar reasoning we get $d_{1_1} = 0$, $s_{1_1} = 3$, $d_{1_3} = 2$ and $s_{1_3} = 1$.

Now suppose that researchers collected sightings during another time interval and assume, just as with the first interval, one animal was sighted at perch 1 and two animals were sighted at perch 2. Denote these sightings as D, E and F respectively. Assume that no animals sighted in the first interval were re-sighted in the second interval. Table 3.2 shows a summary of $\mathbb{E}_2$.

Recall that $\mathbb{C}$ is the set of all possible combinations of configurations from $\mathbb{E}_1$ and $\mathbb{E}_2$. Now that $\mathbb{E}_1$ and $\mathbb{E}_2$ have been defined $\mathbb{C}$ can be enumerated. Specifically,

$$\begin{aligned}\mathbb{C} \;=\; & \{\{e_{1_1}, e_{2_1}\}, \{e_{1_1}, e_{2_2}\}, \{e_{1_1}, e_{2_3}\}, \{e_{1_2}, e_{2_1}\}, \{e_{1_2}, e_{2_2}\}, \\ & \{e_{1_2}, e_{2_3}\}, \{e_{1_3}, e_{2_1}\}, \{e_{1_3}, e_{2_2}\}, \{e_{1_3}, e_{2_3}\}\}.\end{aligned}$$

Thus for this example, $c_1 = \{e_{1_1}, e_{2_1}\}$, $c_{1_1} = e_{1_1}$, $d_1 = 0$ and $s_1 = 6$.

3.3 Parameters

The above notation is used to describe potential configurations of sightings. This section parameterizes a model for the sampling distribution of sightings. Recall that recaptures are

inherently uncertain since the researchers determine, post hoc, which animals, if any, were seen at both perches. The two decisions that the researchers can make are (1) a sighting at one perch is determined by the researchers to be of the same animal as a sighting at the other perch, i.e., that a duplicate individual was sighted. Call this decision a match. Or (2) a sighting at a perch is determined to be of a single individual. Call this decision a non-match. For decision (1) three outcomes can occur: (*i*) A *false positive duplicate* is the event of a mis-match, i.e., that a true duplicate sighting is matched to a sighting other than its complimentary one; (*ii*) A *false positive single* is the event that a true single sighting is matched; (*iii*) A *true positive* is the event that a true duplicate sighting is matched to its compliment. For decision (2) two outcomes can occur: (*i*) A *false negative* is the event that a true duplicate sighting is called a non-match; (*ii*) A *true negative* is the event that a true single sighting is called a non-match.

Finally, we define detection probability as the probability of sighting an animal given that it was available to be sighted. Due to the nature of the bowhead study, it is reasonable to assume that the probability of detection at perch 1 is equal to the probability of detection at perch 2 regardless of whether the whale has been previously seen. Denote the probabilities of the above events as follows:

$$\begin{aligned}
\Pr(\text{false positive duplicate}) &= \alpha_d \\
\Pr(\text{false positive single}) &= \alpha_s \\
\Pr(\text{true positive}) &= (1 - \alpha_d - \beta) \\
\Pr(\text{false negative}) &= \beta \\
\Pr(\text{true negative}) &= (1 - \alpha_s) \\
\Pr(\text{detection}) &= p
\end{aligned}$$

3.3.1 Effective Parameters for Error Rates

Without forcing any constraints, α_s, α_d, and β would be the parameters for two independent multinomial distributions. However, the nature of the problem constraints the space of possible data sets. No forks are allowed and illogical connections are forbidden. A fork occurs when the researchers accidentally match multiple sightings from one perch to only one sighting at the other perch. This occurrence is not allowed because of the assumption that every sighting within a perch is a different animal. Illogical connections refers to possible outcomes that can not occur due to the design of the study but are possible if we assume two independent multinomial distributions. For example, an independent multinomial distribution could produce one sighting as a true positive and the rest as true negatives. This possible outcome cannot occur because a positive outcome of any kind implies a duplicate sighting was sighted at each perch. Together, the issues of forks and illogical connections mean that the error probabilities, α_s, α_d and β, must be re-examined under the constraints. A simple example will be used to illustrate certain constraints and the implications for error probabilities.

Example 3.2

Suppose in one time interval, there is one sighting, denoted as A, at perch 1 and one sighting, denoted B, at perch 2. Further, assume A and B are each other's complimentary sightings, i.e., they are sighting of the same animal. Note that this assumption that A and B are each other's complimentary sighting is the same as assuming that configuration $e_{1_1} \in \mathbb{E}_1$, is true. The whole, unrestricted space of outcomes can be found in Table 3.3. The first two columns represent all possible combinations of errors for sightings A and B. Notice that there can be multiple ways to achieve certain combinations of errors. Hereafter I will refer to a specific combination of errors as an error set type. Denote $T^e_{U_t}$ as error set type t in the unrestricted space. The third column indexes $T^e_{U_t}$.

Table 3.3: Unrestricted space of outcomes for Example 3.2.

A	**B**	T_U^e
TP	TP	1
TP	FP_d	3
TP	FN	4
FP_d	TP	3
FP_d	FP_d	5
FP_d	FN	6
FN	TP	4
FN	FP_d	6
FN	FN	2

Table 3.4: Reduced space of outcomes for Example 3.2.

A	**B**	T_R^e
TP	TP	1
FN	FN	2

Eliminating forks and illogical connections, a restricted space of possible outcomes can be found in Table 3.4. As with Table 3.3, the first two columns of Table 3.4 represent all possible allowable combinations of errors for sighting A and B, but this time limited to the restricted space. Each combination of errors in the restricted space will also be defined as a error set type and be denoted $T_{R_t}^e$.

Each error set type in the restricted space has a corresponding error set type in the whole space. For example, error set type 1 in the reduced space represents the same combination of errors as in error set type 1 of the whole space. While error set types have a counterpart for each space, the number of a particular error set type for the whole space doesn't necessarily equal the number of that particular error set type in the restricted space. For example, there is one error set type 1 ($T_{U_1}^e$)) in the unrestricted space and one of error set type 1 ($T_{R_1}^e$) in the restricted space. This won't always be the case.

Next, false positive single, false positive duplicate, and false negative probabilities can be used to determine the probability of a specific type t, $T^e_{U_t}$. Before that can be done, more notation is needed. Let:

- nFP_{s_t} = the number of false positive singles in error set type $t, T^e_{U_t}$.
- nFP_{d_t} = the number of false positive duplicates in error set type $t, T^e_{U_t}$.
- nFN_t = the number of false negatives in error set type $t, T^e_{U_t}$.
- $nT^e_{U_t}$ = the amount of error set type $t, T^e_{U_t}$.
- $nT^e_{R_t}$ = the amount of error set type $t, T^e_{R_t}$
- $N^e_{T_U}$ = the total amount of error set types in the unrestricted space of possible outcomes.
- $N^e_{T_R}$ = the total amount of error set types in the restricted space of possible outcomes.

In general, the probability mass function for the unrestricted space is:

$$\begin{aligned}\Pr(T^e_{U_t}|\alpha_s, \alpha_d, \beta, r_{i_k}, w_{1,i}, w_{2,i}) &= nT^e_{U_t} \times (1-\alpha_d-\beta)^{2r_{i_k}-nFP_{d_t}-nFN_t} \\ &\times (\alpha_d)^{nFP_{d_t}}(\beta)^{nFN_t}(\alpha_s)^{nFP_{s_t}} \\ &\times (1-\alpha_s)^{w_{1,i}+w_{2,i}-2r_{i_k}-nFP_{s_t}}.\end{aligned}$$

The probability mass function for the restricted space is defined in terms of the unrestricted space:

$$\Pr(T^e_{R_t}|\alpha_s, \alpha_d, \beta, r_{i_k}, w_{1,i}, w_{2,i}) = \left[\frac{nT^e_{R_t}}{nT^e_{U_t}}\right] \frac{\Pr(T^e_{U_t}|\alpha_s, \alpha_d, \beta, r_{i_k}, w_{1,i}, w_{2,i})}{d}$$

where

$$d = \sum_{t=1}^{N^e_{T_R}} \left[\frac{nT^e_{R_t}}{nT^e_{U_t}} \right] \Pr(T^e_{U_t} | \alpha_s, \alpha_d, \beta, r_{i_k}, w_{1,i}, w_{2,i}).$$

Now, recall that α_s, α_d and β are defined as the probabilities of a false positive single, false positive duplicate, and false negative respectively in the whole, unrestricted space. Another way to think of these probabilities is as the weighted sum of the probability (proportion of times) that each event occurs in each error set type. In other words:

$$\begin{aligned}
\alpha_s &= \sum_{t=1}^{N^e_{T_U}} \left(\frac{nFP_{s_t}}{w_{1,i} + w_{2,i} - 2r_{i_k}} \right) \Pr(T^e_{U_t} | \alpha_s, \alpha_d, \beta, r_{i_k}, w_{1,i}, w_{2,i}) \\
\alpha_d &= \sum_{t=1}^{N^e_{T_U}} \left(\frac{nFP_{d_t}}{2r_{i_k}} \right) \Pr(T^e_{U_t} | \alpha_s, \alpha_d, \beta, r_{i_k}, w_{1,i}, w_{2,i}) \\
\beta &= \sum_{t=1}^{N^e_{T_U}} \left(\frac{nFN_t}{2r_{i_k}} \right) \Pr(T^e_{U_t} | \alpha_s, \alpha_d, \beta, r_{i_k}, w_{1,i}, w_{2,i})
\end{aligned}$$

The question is, what are the probabilities of observing a false positive single, false positive duplicate, and false negative sighting in the restricted space? Define the following:

$$\begin{aligned}
\Pr(\text{false positive duplicate in the restricted space}) &= \tilde{\alpha}_d \\
\Pr(\text{false positive single in the restricted space}) &= \tilde{\alpha}_s \\
\Pr(\text{false negative in the restricted space}) &= \tilde{\beta} \\
\Pr(\text{true positive in the restricted space}) &= (1 - \tilde{\alpha}_d - \tilde{\beta}) \\
\Pr(\text{true negative in the restricted space}) &= (1 - \tilde{\alpha}_s)
\end{aligned}$$

where $\tilde{\alpha}_s$, $\tilde{\alpha}_d$ and $\tilde{\beta}$ can be computed as functions of the original parameters. Specifically,

$$\begin{aligned}
\tilde{\alpha}_s &= \sum_{t=1}^{N^e_{T_R}} \left(\frac{nFP_{s_t}}{w_{1,i} + w_{2,i} - 2r_{i_k}} \right) \Pr(T^e_{R_t} | \alpha_s, \alpha_d, \beta, r_{i_k}, w_{1,i}, w_{2,i}) \\
\tilde{\alpha}_d &= \sum_{t=1}^{N^e_{T_R}} \left(\frac{nFP_{d_t}}{2r_{i_k}} \right) \Pr(T^e_{R_t} | \alpha_s, \alpha_d, \beta, r_{i_k}, w_{1,i}, w_{2,i}) \\
\tilde{\beta} &= \sum_{t=1}^{N^e_{T_R}} \left(\frac{nFN_t}{2r_{i_k}} \right) \Pr(T^e_{R_t} | \alpha_s, \alpha_d, \beta, r_{i_k}, w_{1,i}, w_{2,i})
\end{aligned}$$

Due to the nature of the study and the data restrictions, my model will estimate these effective parameters rather than the parameters defined in the unrestricted space.

Example 3.2 continued

Now, suppose that $\alpha_s = 0$, $\alpha_d = .3$ and $\beta = .2$. The probability of observing error set type 1 in the unrestricted space is:

$$\begin{aligned}
\Pr(T^e_{U_1} | \alpha_s, \alpha_d, \beta, r_{1_1}, w_{1,1}, w_{2,1}) &= (1 - \alpha_d - \beta)^2 \\
&= (1 - .3 - .2)^2 \\
&= (.5)^2 \\
&= .25
\end{aligned}$$

The probabilities for the remaining error set types can be found in a similar way and are shown in Table 3.5. Finally, probabilities for each type in the restricted space (i.e., the effective parameters) can be calculated. These are shown in Table 3.6. For example,

Table 3.5: Unrestricted space of outcome probabilities for Example 3.2.

T_U^e	$\Pr(T_U^e)$	nT_U^e
1	.25	1
2	.04	1
3	.30	2
4	.20	2
5	.09	1
6	.12	2

Table 3.6: Restricted space of outcome probabilities for Example 3.2.

T_R^e	$\Pr(T_R^e)$	nT_R^e
1	.862	1
2	.138	1

$$\begin{aligned}
\Pr(T_{R_1}^e|\alpha_s,\alpha_d,\beta,r_{1_1},w_{1,1},w_{2,1}) &= \frac{\Pr(T_{U_1}^e|\alpha_s,\alpha_d,\beta,r_{1_1},w_{1,1},w_{2,1})}{d} \\
&= \frac{.25}{.29} \\
&= .862
\end{aligned}$$

Finally, we can calculate the probabilities of a false positive duplicate and a false negative in the restricted space.

$$\begin{aligned}
\tilde{\alpha}_d &= \sum_{t=1}^{2}\left(\frac{nFP_{d_t}}{2r_{1_1}}\right)\Pr(T_{R_t}^e|\alpha_s,\alpha_d,\beta,r_{1_1},w_{1,1},w_{2,1}) \\
&= \left(\frac{0}{2}\right).862+\left(\frac{0}{2}\right).138 \\
&= 0 \\
\tilde{\beta} &= \sum_{t=1}^{2}\left(\frac{nFN_t}{2r_{1_1}}\right)\Pr(T_{R_t}^e|\alpha_s,\alpha_d,\beta,r_{1_1},w_{1,1},w_{2,1}) \\
&= \left(\frac{0}{2}\right).862+\left(\frac{2}{2}\right).138 \\
&= .136
\end{aligned}$$

Thus, the probability of a false positive duplicate in the restricted space is 0 and the probability of a false negative in the restricted space is .136.

3.3.2 Effective Parameter for Detection Probability

In classical capture recapture experiments, detection probability is the probability that an animal is sighted given it was available to be sighted. In studies where captured animals are tagged, the number of individuals and/or recaptures is known exactly and estimation of detection probability is straight forward. However, due to the nature of the bowhead data (namely that tagging isn't possible), the true number of animals that were captured and recaptured is uncertain. The only information available is the number of unique animals sighted at each perch. Given $w_{1,i}$ and $w_{2,i}$ there are many different possibilities for the number of animals and the number of recaptures, as we saw during the discussion of the definition of $\mathbb{E}_i$. Each possibility will produce a different amount of duplicate individuals. Therefore, we need to examine the probability a sighting is a duplicate sighting given $w_{1,i}$ and $w_{2,i}$.

Example 3.3

Let $w_{1,1} = 1$, with the lone sighting being denoted A, and $w_{2,1} = 2$, with the two sightings being denoted B and C. Table 3.7 shows the distribution of possible duplicate sightings histories if we assume a binomial distribution in determining which of the $w_{1,1}$ and $w_{2,1}$ sightings are duplicate sightings or single sightings. D indicates that a sighting is a duplicate sighting and S that a sighting is a single sighting. Table 3.7 lists different combinations of duplicate and single sightings and the fourth column indexes those different combinations. Hereafter I will refer to these different combinations as duplicate types, denoted at T^d.

Table 3.7: Unrestricted space of duplicate sighting outcomes for Example 3.3.

A	**B**	**C**	T^d
D	D	D	3
D	D	S	1
D	S	D	1
S	D	D	1
D	S	S	4
S	D	S	4
S	S	D	4
S	S	S	2

Table 3.8: Restricted space of duplicate sighting outcomes for Example 3.3.

A	**B**	**C**	T_R^d
D	D	S	1
D	S	D	1
S	S	S	2

Let's take a closer look at duplicate type 3. In this case each sighting is a duplicate sighting. However, since sighting B and C can not be the same animal since they were sighted at perch 2 and therefore distinguished by perch 2 observers and perch 1 has only one sighting, it stands to reason that at most one of B or C are duplicate sightings; both couldn't have been. Thus, type 3 represents an illogical, i.e. impossible, outcome. Removing all illogical outcomes we get Table 3.8. The latter table reads exactly like Table 3.7 except that T_R^d represents the duplicate type combination in the now restricted space. Just like with the effective errors, duplicate types in the restricted space correspond to duplicate types in the unrestricted space, where the number of each type can differ between the two spaces.

Similar to the logic that led to the definition of $\tilde{\alpha}_s$, $\tilde{\alpha}_d$, and $\tilde{\beta}$ parameters, I can calculate the general probability mass functions for duplicate types for unrestricted and restricted spaces. Specifically,

$$\Pr(T_t^d|w_{1,i}, w_{2,i}, p) = nT_t^d p^{nD_t}(1-p)^{w_{1,i}+w_{2,i}-nD_t}$$

and

$$\Pr(T_{R_t}^d|w_{1,i}, w_{2,i}, p) = \left[\frac{nT_{R_t}^d}{nT_t^d}\right] \frac{\Pr(T_t^d|w_{1,i}, w_{2,i}, p)}{d}$$

where

$$\begin{aligned}
d &= \sum_{t=1}^{N_{T_R}^d} \left[\frac{nT_{R_t}^d}{nT_t^d}\right] \Pr(T_t^d|w_{1,i}, w_{2,i}, p) \\
nD_t &= \text{the number of duplicate sightings in duplicate type } t, T_t^d \\
N_{T_R}^d &= \text{the total amount of duplicate types in the restricted space} \\
nT_t^d &= \text{the number of duplicate types } t, T_t^d \\
nT_{R_t}^d &= \text{the number of duplicate types } t, T_{R_t}^d \\
p &= \text{classical definition of detection probability}
\end{aligned}$$

As with α_s, α_d, and β, another way to view the probability of a duplicate sighting is with the following weighted sum:

$$p = \sum_{t=1}^{N_T^d} \left(\frac{nD_t}{w_{1,i} + w_{2,i}}\right) \Pr(T_t^d|w_{1,i}, w_{2,i}, p)$$

where, N_T^d is the number of duplicate types in the unrestricted space. Thus,

$$\tilde{p} = \sum_{t=1}^{N_{R_t}^d} \left(\frac{nD_t}{w_1 + w_2}\right) \Pr(T_{R_t}^d|w_1, w_2, p)$$

Table 3.9: Probabilities for the outcomes in the unrestricted space for Example 3.3.

T^d	$\Pr(T^d)$	nT^d
1	.288	3
2	.216	1
3	.064	1
4	.432	3

Example 3.3 Continued

I will use this example to further illustrate the calculation of $\tilde{p}$. Suppose that $p = .4$. The probability of observing duplicate type 1 in the unrestricted space is:

$$\begin{aligned} \Pr(T_1^d | w_{1,i}, w_{2,i}, p) &= nT_1^d p^2 (1-p)^1 \\ &= 3(.4)^2(.6) \\ &= .288 \end{aligned}$$

The rest of the probabilities are calculated in a similar fashion and can be found in Table 3.9. Using the probabilities of each duplicate type in the unrestricted space, probabilities for each duplicate type in the restricted space can be calculated; see Table 3.10.

Finally, the probability that a sighting is a duplicate sighting given $w_{1,1}$ and $w_{2,1}$ in the restricted space is:

$$\begin{aligned} \tilde{p} &= \sum_{t=1}^{2} \left(\frac{nD_t}{w_{1,1} + w_{2,1}} \right) \Pr(T_{Rt}^d | w_{1,1}, w_{2,1}, p) \\ &= \left(\frac{2}{3} \right) .470 + \left(\frac{0}{3} \right) .530 \\ &= .313 \end{aligned}$$

The value of $\tilde{p}$ is the parameter that can be estimated in the bowhead case due to the constraints imposed on possible outcomes of the sighting process.

Table 3.10: Probabilities for the outcomes in the restricted space for Example 3.3.

T_R^d	$\Pr(T_R^d)$	nT_R^d
1	.470	2
2	.530	1

3.4 Data

An unavoidable aspect of the study design for the bowhead case is that the true number of recaptured animals is unknown. That is, it is unknown whether an animal was seen at both perches or not. Using the data collected from each perch, researchers try to determine which animals, if any, were sighted at both perches. This process entails making a series of match and non-match decisions. In the case we consider, this match process task is repeated J times independently. These match processes, or replications, give statistical evidence about the matching errors.

For each time interval i and match process j let $Y_{i,j}$ be the collection of match/non-match decisions. Each sighting at each perch will have J match/non-match decisions associated with it so let Y_i be considered the collection of independent replications for time interval i. Finally, let Y be the collection of match/non-match decisions for all I time intervals including all J independent match efforts.

Example 3.4

Let $I = 2$, $J = 3$, $w_{1,1} = 1$, $w_{2,1} = 2$, $w_{1,2} = 1$ and $w_{2,2} = 2$. Let A denote the sighting at perch 1 at time interval 1. Let B and C denote the sightings at perch 2 at time interval 1. Let D denote the sighting at perch 1 at time interval 2 and let E and F denote the sightings at perch 2 at time interval 2. The collection of data $Y_{i,j}$, Y_i and Y can be found in Table 3.11, 3.12, and 3.13 respectively.

Table 3.11: Match/Non-Match data, $Y_{1,1}$, for interval 1 and replicate 1 (Rep 1) for Example 3.4.

Sighting	**Rep 1**
A	B
B	A
C	non-match

Table 3.12: Match/Non-Match data, Y_1, for interval 1 and replicates 1-3 (Rep 1, Rep 2 and Rep 3 respectively) for Example 3.4.

Sighting	**Rep 1**	**Rep 2**	**Rep 3**
A	B	B	non-match
B	A	A	non-match
C	non-match	non-match	non-match

Table 3.13: Match/Non-Match data, Y, for intervals 1 and 2 and replicates 1-3 (Rep 1, Rep 2 and Rep 3 respectively) for Example 3.4.

Sighting	**Rep 1**	**Rep 2**	**Rep 3**
A	B	B	non-match
B	A	A	non-match
C	non-match	non-match	non-match
D	non-match	F	non-match
E	non-match	non-match	non-match
F	non-match	D	non-match

3.5 Likelihood

The statistical evidence about error rates is provided by the replication of matching decisions. While the true capture history is unknown, we can enumerate all possibilities given $\mathbb{W}_1$ and $\mathbb{W}_2$ with the set $\mathbb{C}$. Recall that an element of a set in $\mathbb{C}$, i.e., $c_{l_i} = e_{i_k}$, represents one possible underlying true capture history for time interval i. Comparing the data, $Y_{i,j}$, to e_{i_k} allows us to calculate the number of false positive duplicates, false positive singles and false negatives. Define the following:

$$\begin{aligned} x_{ijk}^{TP} &= \text{the number of true positives in } Y_{i,j}|e_{i_k} \\ x_{ijk}^{FP_d} &= \text{the number of false positive duplicates in } Y_{i,j}|e_{i_k} \\ x_{ijk}^{FP_s} &= \text{the number of false positive singles in } Y_{i,j}|e_{i_k} \end{aligned}$$

$$\begin{aligned} x_{ik}^{TP} &= \text{the number of true positives in } Y_i|e_{i_k} \\ x_{ik}^{FP_d} &= \text{the number of false positive duplicates in } Y_i|e_{i_k} \\ x_{ik}^{FP_s} &= \text{the number of false positive singles in } Y_i|e_{i_k} \end{aligned}$$

$$\begin{aligned} x_{l}^{TP} &= \text{the number of true positives in } Y|c_l \\ x_{l}^{FP_d} &= \text{the number of false positive duplicates in } Y|c_l \\ x_{l}^{FP_s} &= \text{the number of false positive singles in } Y|c_l \end{aligned}$$

The likelihood function for $\tilde{\alpha}_s$, $\tilde{\alpha}_d$, $\tilde{\beta}$, $\tilde{p}$, $w_{1,i}$ and $w_{2,i}$ given $Y_{i,j}$ can be derived by applying the law of total probability over the set $\mathbb{E}_i$. It is important to note that we are conditioning on $w_{1,i}$ and $w_{2,i}$ just like one would condition on n in a binomial distribution.

$$\begin{aligned}
\Pr(Y_{i,j}|\tilde{\alpha}_s, \tilde{\alpha}_d, \tilde{\beta}, \tilde{p}, w_{1,i}, w_{2,i}) &= \sum_{k=1}^{K_i} \big[\Pr(Y_{i,j}|\tilde{\alpha}_s, \tilde{\alpha}_d, \tilde{\beta}, \tilde{p}, w_{1,i}, w_{2,i}, e_{i_k}) \\
&\quad \times \Pr(e_{i_k}|\tilde{p}, w_{1,i}, w_{2,i}) \big] \\
\Pr(Y_{i,j}|\tilde{\alpha}_s, \tilde{\alpha}_d, \tilde{\beta}, \tilde{p}, e_{i_k}, w_{1,i}, w_{2,i}) &= (1-\tilde{\alpha}_d-\tilde{\beta})^{x_{ijk}^{TP}} (\tilde{\alpha}_d)^{x_{ijk}^{FP_d}} \\
&\quad \times (\tilde{\beta})^{d_{i_k}-x_{ijk}^{TP}-x_{ijk}^{FP_d}} (\tilde{\alpha}_s)^{x_{ijk}^{FP_s}} \\
&\quad \times (1-\tilde{\alpha}_s)^{s_{i_k}-x_{ijk}^{FP_s}} \\
\Pr(e_{i_k}|\tilde{p}, w_{1,i}, w_{2,i}) &= (\tilde{p})^{d_{i_k}} (1-\tilde{p})^{s_{i_k}}
\end{aligned}$$

Due to the independence of the matching process and the independence of the time intervals, we get the following:

$$\begin{aligned}
\Pr(Y_i|\tilde{\alpha}_s, \tilde{\alpha}_d, \tilde{\beta}, \tilde{p}, w_{1,i}, w_{2,i}) &= \prod_{j=1}^{J} \Pr(Y_{i,j}|\tilde{\alpha}_s, \tilde{\alpha}_d, \tilde{\beta}, \tilde{p}, w_{1,i}, w_{2,i}) \\
\Pr(Y|\tilde{\alpha}_s, \tilde{\alpha}_d, \tilde{\beta}, \tilde{p}, w_{1,i}, w_{2,i}) &= \prod_{i=1}^{I} \Pr(Y_i|\tilde{\alpha}_s, \tilde{\alpha}_d, \tilde{\beta}, \tilde{p}, w_{1,i}, w_{2,i})
\end{aligned}$$

3.6 Estimation Framework

While maximum likelihood estimation is computationally feasible when $w_{1,i}$ and $w_{2,i}$ are small in every time interval it is essentially impossible when at least one set of $w_{1,i}$ and $w_{2,i}$ is of any non-trivial magnitude. To understand why, one needs only to look so far as K_i, the

number of configurations in $\mathbb{E}_i$, because the likelihood requires summing over all elements in $\mathbb{E}_i$. The calculation for K_i is:

$$K_i = \sum_{r=0}^{w_{1,i}} \left(\frac{w_{1,i}!}{(w_{1,i}-r)!(r)!} \right) \left(\frac{w_{2,i}!}{(w_{2,i}-r)!(w_{2,i})!} \right) (r!).$$

In Example 3.1 when $w_{1,1} = 1$ and $w_{2,1} = 2$, it is easy to determine that $K_1 = 3$. If we slightly increase the size of the dataset so $w_{1,1} = 4$ and $w_{2,1} = 5$, K_1 jumps dramatically to 501. Realistic values for $w_{1,i}$ and $w_{2,i}$ yield huge K_i, making maximum likelihood estimation very difficult. Therefore, I am proposing a Bayesian framework for parameter estimation, treating configurations as parameters. Of course $\tilde{\alpha}_s$, $\tilde{\alpha}_d$, $\tilde{\beta}$ and $\tilde{p}$ are also parameters to be estimated. The treatment of configurations as parameters resembles the approach of Link et al. (2010), though my framework is much more general and my estimation strategy is more complex.

There are several levels to my modeling approach. First, the sampling distribution of the data is:

$$Y \sim \Pr(Y|\tilde{\alpha}_s, \tilde{\alpha}_d, \tilde{\beta}, \tilde{p}, \mathbb{W}_1, \mathbb{W}_2, c_l)$$

where

$$\begin{aligned}
\Pr(Y|\tilde{\alpha}_s, \tilde{\alpha}_d, \tilde{\beta}, \tilde{p}, \mathbb{W}_1, \mathbb{W}_2, c_l) &= \prod_{i=1}^{I} \Pr(Y_i|\tilde{\alpha}_s, \tilde{\alpha}_d, \tilde{\beta}, \tilde{p}, w1, i, w_{2,i}, c_{l_i}) \\
&= \prod_{i=1}^{I} \prod_{j=1}^{J} \Pr(Y_{i,j}|\tilde{\alpha}_s, \tilde{\alpha}_d, \tilde{\beta}, \tilde{p}, w_{1,i}, w_{2,i}, c_{l_i}) \\
&= \prod_{i=1}^{I} \prod_{j=1}^{J} \Pr(Y_{i,j}|\tilde{\alpha}_s, \tilde{\alpha}_d, \tilde{\beta}, \tilde{p}, w_{1,i}, w_{2,i}, e_{i_k}).
\end{aligned}$$

Next,

$$c_l \sim \Pr(c_l|\tilde{p}, \mathbb{W}_1, \mathbb{W}_2)$$

where

$$\begin{aligned} \Pr(c_l|p, \mathbb{W}_1, \mathbb{W}_2) &= \prod_{i=1}^{I} \Pr(c_{l_i}|\tilde{p}, w_{1,i}, w_{2,i}) \\ &= \prod_{i=1}^{I} \Pr(e_{i_k}|\tilde{p}, w_{1,i}, w_{2,i}). \end{aligned}$$

Next, I adopt the following priors.

$$\begin{aligned} \tilde{\alpha}_s &\sim \text{Unif(0,0.3)} \\ \tilde{\alpha}_d &\sim \text{Unif(0,0.3)} \\ \tilde{\beta} &\sim \text{Unif(0,0.3)} \\ \tilde{p} &\sim \text{Unif(0,1)} \\ c_l &\sim \text{Discrete Uniform over the set } \mathbb{C}. \end{aligned}$$

For the priors I assume that researchers do (at least) a slightly better job at correctly identifying each sighting as being a duplicate sighting or a single sighting than flipping a coin would. That is why the flat, relatively non-informative Unif(0,0.3) priors were chosen

for $\tilde{\alpha}_s$, $\tilde{\alpha}_d$ and $\tilde{\beta}$. I don't make any prior assumptions about $\tilde{p}$ or c_l and that is why Unif(0,1) and Discrete Uniform over the set $\mathbb{C}$ were chosen for their respective parameters.

3.7 Markov Chain Monte Carlo

Using the framework above, the joint posterior distribution is

$$\begin{aligned}\Pr(\tilde{\alpha}_s, \tilde{\alpha}_d, \tilde{\beta}, \tilde{p}, c_l | Y, \mathbb{W}_1, \mathbb{W}_2) &\propto \Pr(\tilde{\alpha}_s)\Pr(\tilde{\alpha}_d)\Pr(\tilde{\beta})\Pr(\tilde{p})\Pr(c_l) \\ &\times \Pr(c_l|\tilde{p}, \mathbb{W}_1, \mathbb{W}_2) \\ &\times \Pr(Y|\tilde{\alpha}_s, \tilde{\alpha}_d, \tilde{\beta}, \tilde{p}, \mathbb{W}_1, \mathbb{W}_2, c_l)\end{aligned}$$

This expression is algebraically intractable so inference relies on the Markov Chain Monte Carlo approach (Gelman et al., 2014). However, sampling directly from the joint posterior distribution isn't possible so I will use a Metropolis within Gibbs MCMC strategy.

In a canonical setting, the Metropolis within Gibbs approach (e.g., Givens and Hoeting, 2005) for simulating from distribution $f(x)$ for the random p-dimensional $X = (X_1, \ldots, X_p)^T$ is:

1. Let $t = 0$ and let x^t denote initial starting values for all p parameters.
2. Choose an ordering of the components of x^t.
3. For $i = 1, \cdots, p$ indexing the ordering above, complete the following steps.
   (a) Sample candidate value X_i^* from a proposal distribution $g(\cdot|x_i^t)$.

(b) Compute the Metropolis-Hastings ratio:

$$R(x_i^t, X_i^*) = \frac{f(X_i^*|x_{-i}^t)g(x_i^t|X_i^*)}{f(x_i^t|x_{-i}^t)g(X_i^*|x_i^t)}$$

(c) Let

$$X_i^{t+1} = \begin{cases} X_i^* & \text{with probability min(R,1)} \\ x_i^t & \text{otherwise} \end{cases}$$

4. Increment t and return to step 2.

In our case $X = (\tilde{\alpha}_s, \tilde{\alpha}_d, \tilde{\beta}, \tilde{p}, c_l)^T$ and f(x) is the joint posterior distribution described earlier in this section. I will take the proposal distributions for $\tilde{\alpha}_s$, $\tilde{\alpha}_d$, and $\tilde{\beta}$ to be equal to the prior distributions, i.e., each is Unif(0,0.3). My proposal for $\tilde{p}$ is Unif(0,1). Note that I have chosen proposal distributions that do not depend on the parameters' current locations in the Metropolis step.

The proposal distribution for the set $\mathbb{C}$ is more complex. I have chosen to update a configuration, c_{l_i} in $c_l \in \mathbb{C}$ in a manner similar to a random walk in the sense that a proposed configuration differs only slightly from the current configuration. For any time interval i, denote the current configuration as e_i^t and the proposed configuration as e^*. The steps I use to propose one configuration given the current configuration are as follows:

1. Randomly select one sighting from the $w_{i,1}$ sightings at perch 1. Denote the selected sighting as s_{w_1}

2. Randomly select one sighting from the $w_{2,1}$ sightings at perch 2. Denote the selected sighting as s_{w_2}

3. If s_{w_1} and s_{w_2} are each other's complimentary sightings in e_i^t, then break the connection so that each sighting becomes a single sighting in e^*.

4. If s_{w_1} and s_{w_2} are both duplicate sightings but are not each other's complimentary sightings in e_i^t, then switch the connections. That is, s_{w_1}'s previous complimentary sighting becomes s_{w_2}'s previous complementary sighting complimentary sighting and s_{w_2} becomes s_{w_1}'s complimentary sighting to generate e^*.

5. If s_{w_1} is a single sighting and s_{w_2} is a duplicate sighting in e_i^t then

   (a) Break the connection between s_{w_2} and its complimentary sighting. s_{w_2}'s previous complimentary sighting becomes a single sighting in e^*.

   (b) Connect s_{w_1} with s_{w_2}. Then s_{w_1} becomes a duplicate sighting in e^*.

6. If s_{w_2} is a single sighting and s_{w_1} is a duplicate sighting in e_i^t then

   (a) Break the connection between s_{w_1} and its complimentary sighting. Then s_{w_1}'s complimentary sighting becomes a single sighting in e^*.

   (b) Connect s_{w_2} with s_{w_1}. Then s_{w_2} becomes a duplicate sighting in e^*.

7. If both s_{w_1} and s_{w_2} are single sighting in e_i^t, then connect them. That is the two single sightings, s_{w_1} and s_{w_2}, become each other's complimentary sighting in e^*.

A simple example is used to demonstrate the random walk like process for the proposal distribution of configurations, described above, for one time interval.

Table 3.14: Current configuration, e_1^t.

Current Configuration (e_1^t)	
Sighting	Complimentary
A	K
B	Q
C	N
D	none
E	none
K	A
L	none
M	none
N	C
O	none
P	none
Q	B

Let $w_{1,1} = 5$ and $w_{2,1} = 7$. The uniquely identified sightings at perch 1 and perch 2 are as follows:

$$\begin{aligned} \text{Perch 1} &= \text{A, B, C, D, E} \\ \text{Perch 2} &= \text{K, L, M, N, O, P Q} \end{aligned}$$

Now suppose that the current configuration at iteration t, denoted e_1^t is shown in Table 3.14.

1. Suppose that sighting A is selected from perch 1. Then s_{w_1}=A.

2. Suppose that sighting K is selected from perch 2. Then s_{w_2}=K.

Table 3.15: Current configuration, e_1^t, and proposed configuration, e^*, with s_{w_1}=A and s_{w_2}=K.

Current Configuration (e_1^t)		**Proposed Configuration** (e^*)	
Sighting	Complimentary	Sighting	Complimentary
A	K	A	none
B	Q	B	Q
C	N	C	N
D	none	D	none
E	none	E	none
K	A	K	none
L	none	L	none
M	none	M	none
N	C	N	C
O	none	O	none
P	none	P	none
Q	B	Q	B

3. Since s_{w_1}=A and s_{w_2}=K are each other's complimentary sightings we proceed with step 3, that is, we break the connection so that each sighting becomes a single sighting in the proposed configuration, e^*. The proposed configuration compared to the current configuration can be found in Table 3.15.

1. Suppose that sighting A is selected from perch 1. Then s_{w_1}=A.

2. Suppose that sighting N is selected from perch 2. Then s_{w_2}=N.

3. Since s_{w_1}=A and s_{w_2}=K are both duplicate sightings but not each other's complimentary sighting we proceed with step 4, that is, switch their connections. In other words, s_{w_1}=A's complimentary sighting is now s_{w_2}=N and s_{w_2}=N's complimentary sighting is now s_{w_1}=A in the proposed configuration, e^*. The proposed configuration compared to the current configuration can be found in Table 3.16.

Table 3.16: Current configuration, e_1^t, and proposed configuration, e^*, with s_{w_1}=A and s_{w_2}=N.

Current Configuration (e_1^t)		**Proposed Configuration** (e^*)	
Sighting	Complimentary	Sighting	Complimentary
A	K	A	N
B	Q	B	Q
C	N	C	K
D	none	D	none
E	none	E	none
K	A	K	C
L	none	L	none
M	none	M	none
N	C	N	A
O	none	O	none
P	none	P	none
Q	B	Q	B

1. Suppose that sighting D is selected from perch 1. Then s_{w_1}=D.

2. Suppose that sighting K is selected from perch 2. Then s_{w_2}=K.

3. Since s_{w_1}=D is a single sighting and s_{w_2}=K is a duplicate sighting we proceed with step 5, that is, they become each other's complimentary sightings. In other words, s_{w_1}=D and s_{w_2}=K become complimentary sightings and s_{w_2}=K's previous complimentary sighting, A, is now a single sighting in the proposed configuration, e^*. The proposed configuration compared to the current configuration can be found in Table 3.17.

1. Suppose that sighting D is selected from perch 1. Then s_{w_1}=D.

2. Suppose that sighting M is selected from perch 2. Then s_{w_2}=M.

Table 3.17: Current configuration, e_1^t, and proposed configuration, e^*, with s_{w_1}=D and s_{w_2}=K.

Current Configuration (e_1^t)		**Proposed Configuration (e^*)**	
Sighting	Complimentary	Sighting	Complimentary
A	K	A	none
B	Q	B	Q
C	N	C	N
D	none	D	K
E	none	E	none
K	A	K	D
L	none	L	none
M	none	M	none
N	C	N	C
O	none	O	none
P	none	P	none
Q	B	Q	B

3. Since both s_{w_1}=D and s_{w_2}=M are single sightings we proceed with step 7, that is, connect the two so that they become each other's complimentary sighting in the proposed configuration, e^*. The proposed configuration can be found in Table 3.17.

The proposal distribution for configurations satisfies the following equality:

$$\Pr(e_i^t|e^*) \quad = \quad \Pr(e^*|e_i^t)$$

To prove this equality it is important to note the possible configurations that can be created. If step 3 is conducted then the resulting proposed configuration has one fewer duplicate individual than the current configuration. Call the possible proposed configurations of step 3 the $r-1$ group, where r is the number of duplicates in e_i^t. If steps 5, 6 or 7 are conducted then the resulting proposed configuration has the same number of duplicate individuals as the current configuration. Call the possible proposed configurations of steps 5, 6 or 7 the r group. If step 4 is conducted then the resulting proposed configuration has one more dupli-

Table 3.18: Current configuration, e_1^t, and proposed configuration, e^*, with s_{w_1}=D and s_{w_2}=M.

Current Configuration (e_1^t)		**Proposed Configuration (e^*)**	
Sighting	Complimentary	Sighting	Complimentary
A	K	A	K
B	Q	B	Q
C	N	C	N
D	none	D	M
E	none	E	none
K	A	K	A
L	none	L	none
M	none	M	D
N	C	N	C
O	none	O	none
P	none	P	none
Q	B	Q	B

cate individual than the current configuration. Call the possible proposed configurations of step 4 the $r+1$ group.

There are $w_{1,i}$ ways to choose one sighting from perch 1 and $w_{2,i}$ ways to choose one sighting from perch 2. Overall, then, there are $w_{1,i} \times w_{2,i}$ different potential proposed configurations given a current configuration. We can split up the $w_{1,i} \times w_{2,i}$ different configurations into their different groups; the $r-1$ group, r group and $r+1$ group.

There are r different ways to select a configuration in the $r-1$ group, all of which are unique. That is, each configuration that can be produced in the group cannot be achieved in multiple ways. There are $(w_{1,i}-r)(w_{2,i}-r)$ different ways to get a configuration in the $r+1$ group. Once again all are unique. The r group needs to be further broken down into subgroups to determine the number of potential proposed configurations. The first subgroup is when step 6 or 7 is conducted, i.e., breaking one duplicate individual and creating another. There are $r(w_{1,i}-r)+r(w_{2,i}-r)$ ways to create a configuration in this subgroup. Again all of these are unique. The second subgroup is when step 5 is conducted. This step takes

two duplicate sightings and switches their respective complimentary sightings. There are $r(r-1)$ ways to achieve this. However, there are two ways to make a switch resulting in the same configuration. Therefore, this subgroup has $\frac{r(r-1)}{2}$ unique configurations. Notice, also, that this second subgroup has all the same duplicate sightings as the current configuration (although these duplicate sightings don't necessarily have the same complimentary sightings) while the other subgroup and groups have at least one duplicate sighting that differs from the current configuration. Now we can compute the transition probability. In summary:

$$\Pr(e^*|e_i^t) = \begin{cases} \frac{1}{(w_{1,i})(w_{2,i})} & e^* \in (r-1) \text{ group} \\ \frac{1}{(w_{1,i})(w_{2,i})} & e^* \in (r+1) \text{ group} \\ \frac{1}{(w_{1,i})(w_{2,i})} & e^* \in r \text{ group and subgroup 1} \\ \frac{2}{(w_{1,i})(w_{2,i})} & e^* \in r \text{ group and subgroup 2} \end{cases}$$

Assume e_i^t has r duplicate individuals and e^* has $r-1$ duplicate individuals. Then,

$$\Pr(e_i^t|e^*) = \frac{1}{(w_{1,i})(w_{2,i})}.$$

Notice that since e^* has $r-1$ duplicate individuals and e_i^t has r individuals that e_i^t is in the $r+1$ group compared to e^*. Now, if e^* has $r+1$ duplicate individuals,

$$\Pr(e_i^t|e^*) = \frac{1}{(w_{1,i})(w_{2,i})}.$$

Additionally, if e^* also has r duplicate individuals then

$$\Pr(e_i^t|e^*) = \begin{cases} \frac{1}{(w_{1,i})(w_{2,i})} & e_i^t \quad \text{in subgroup 1} \\ \frac{2}{(w_{1,i})(w_{2,i})} & e_i^t \quad \text{in subgroup 2} \end{cases}$$

Thus for all e_i^t and e^*

$$\Pr(e^*|e_i^t) \;\; = \;\; \Pr(e_i^t|e^*).$$

The proposal distribution, $\Pr(e^*|e_i^t)$, is used to explore the different possible configurations in $\mathbb{E}_i$. To apply my approach to all possible combinations of configurations in $\mathbb{C}$ for all time intervals, I randomly select a pre-specified number of time intervals to update (all with equal probability) then propose configurations for the chosen time intervals using $\Pr(e_i^t|e^*)$. Denote the proposal distribution for configurations for all intervals as $\Pr(c^*|c^t)$.

Now we can write out the Metropolis within Gibbs algorithm:

1. Let $t = 0$ and denote $\tilde{\alpha}_s^t, \tilde{\alpha}_d^t, \tilde{\beta}^t, \tilde{p}^t$ and c^t as each parameters value at iteration t.

2. Propose $\tilde{\alpha}_s^* \sim$ Unif(0,0.3)

3. Calculate

$$
\begin{aligned}
R &= \frac{\Pr(\tilde{\alpha}_s^*|\tilde{\alpha}_d^t\tilde{\beta}^t,\tilde{p}^t,c^t,Y,\mathbb{W}_1,\mathbb{W}_2)\Pr(\tilde{\alpha}_s^t)}{\Pr(\tilde{\alpha}_s^t|\tilde{\alpha}_d^t\tilde{\beta}^t,\tilde{p}^t,c^t,Y,\mathbb{W}_1,\mathbb{W}_2)\Pr(\tilde{\alpha}_s^*)} \\
&= \Big[\Pr(Y|\tilde{\alpha}_s^*,\tilde{\alpha}_d^t,\tilde{\beta}^t,\tilde{p}^t,c^t,\mathbb{W}_1,\mathbb{W}_2)\Pr(c^t|\tilde{p}^t,\mathbb{W}_1,\mathbb{W}_2) \\
&\quad \times \Pr(\tilde{\alpha}_s^*)\Pr(\tilde{\alpha}_d^t),\Pr(\tilde{\beta}^t)\Pr(\tilde{p}^t)\Pr(c^t)\Pr(\tilde{\alpha}_s^t)\Big] \\
&\div \Big[\Pr(Y|\tilde{\alpha}_s^t,\tilde{\alpha}_d^t,\tilde{\beta}^t,\tilde{p}^t,c^t,\mathbb{W}_1,\mathbb{W}_2)\Pr(c^t|\tilde{p}^t,\mathbb{W}_1,\mathbb{W}_2) \\
&\quad \times \Pr(\tilde{\alpha}_s^t)\Pr(\tilde{\alpha}_d^t),\Pr(\tilde{\beta}^t)\Pr(\tilde{p}^t)\Pr(c^t)\Pr(\tilde{\alpha}_s^*)\Big] \\
&= \frac{\Pr(Y|\tilde{\alpha}_s^*,\tilde{\alpha}_d^t,\tilde{\beta}^t,\tilde{p}^t,c^t,\mathbb{W}_1,\mathbb{W}_2)}{\Pr(Y|\tilde{\alpha}_s^t,\tilde{\alpha}_d^t,\tilde{\beta}^t,\tilde{p}^t,c^t,\mathbb{W}_1,\mathbb{W}_2)} \\
&= \frac{(1-\tilde{\alpha}_d^t-\tilde{\beta}^t)^{x_t^{TP}}(\tilde{\alpha}_d^t)^{x_t^{FP_d}}(\tilde{\beta}^t)^{(Jd_t)-x_t^{TP}-x_t^{FP_d}}(\tilde{\alpha}_s^*)^{x_t^{FP_s}}(1-\tilde{\alpha}_s^*)^{(Js_t)-x_t^{FP_s}}}{(1-\tilde{\alpha}_d^t-\tilde{\beta}^t)^{x_t^{TP}}(\tilde{\alpha}_d^t)^{x_t^{FP_d}}(\tilde{\beta}^t)^{(Jd_t)-x_t^{TP}-x_t^{FP_d}}(\tilde{\alpha}_s^t)^{x_t^{FP_s}}(1-\tilde{\alpha}_s^t)^{(Js_t)-x_t^{FP_s}}} \\
&= \frac{(\tilde{\alpha}_s^*)^{x_t^{FP_s}}(1-\tilde{\alpha}_s^*)^{(Js_t)-x_t^{FP_s}}}{(\tilde{\alpha}_s^t)^{x_t^{FP_s}}(1-\tilde{\alpha}_s^t)^{(Js_t)-x_t^{FP_s}}}.
\end{aligned}
$$

4. Assign a new $\tilde{\alpha}_s$ value according to

$$
\tilde{\alpha}_s^{t+1} = \begin{cases} \tilde{\alpha}_s^* & \text{with probability } \min(1,R) \\ \tilde{\alpha}_s^t & \text{otherwise.} \end{cases}
$$

5. Propose $\tilde{\alpha}_d^* \sim$ Unif(0,0.3)

6. Calculate

$$
\begin{aligned}
R &= \frac{\Pr(\tilde{\alpha}_d^*|\tilde{\alpha}_s^{t+1},\tilde{\beta}^t,\tilde{p}^t,c^t,Y,\mathbb{W}_1,\mathbb{W}_2)\Pr(\tilde{\alpha}_d^t)}{\Pr(\tilde{\alpha}_d^t|\tilde{\alpha}_s^{t+1},\tilde{\beta}^t,\tilde{p}^t,c^t,Y,\mathbb{W}_1,\mathbb{W}_2)\Pr(\tilde{\alpha}_d^*)} \\
&= \Big[\Pr(Y|\tilde{\alpha}_s^{t+1},\tilde{\alpha}_d^*,\tilde{\beta}^t,\tilde{p}^t,c^t,\mathbb{W}_1,\mathbb{W}_2)\Pr(c^t|\tilde{p}^t,\mathbb{W}_1,\mathbb{W}_2) \\
&\quad \times \Pr(\tilde{\alpha}_s^{t+1})\Pr(\tilde{\alpha}_d^*),\Pr(\tilde{\beta}^t)\Pr(\tilde{p}^t)\Pr(c^t)\Pr(\tilde{\alpha}_d^t)\Big] \\
&\div \Big[\Pr(Y|\tilde{\alpha}_s^{t+1},\tilde{\alpha}_d^t,\tilde{\beta}^t,\tilde{p}^t,c^t,\mathbb{W}_1,\mathbb{W}_2)\Pr(c^t|\tilde{p}^t,\mathbb{W}_1,\mathbb{W}_2) \\
&\quad \times \Pr(\tilde{\alpha}_s^{t+1})\Pr(\tilde{\alpha}_d^t),\Pr(\tilde{\beta}^t)\Pr(\tilde{p}^t)\Pr(c^t)\Pr(\tilde{\alpha}_d^*)\Big] \\
&= \frac{\Pr(Y_i|\tilde{\alpha}_s^{t+1},\tilde{\alpha}_d^*,\tilde{\beta}^t,\tilde{p}^t,c^t,\mathbb{W}_1,\mathbb{W}_2)}{\Pr(Y|\tilde{\alpha}_s^{t+1},\tilde{\alpha}_d^t,\tilde{\beta}^t,\tilde{p}^t,c^t,\mathbb{W}_1,\mathbb{W}_2)} \\
&= \frac{(1-\tilde{\alpha}_d^*-\tilde{\beta}^t)^{x_t^{TP}}(\tilde{\alpha}_d^*)^{x_t^{FP_d}}(\tilde{\beta}^t)^{(Jd_t)-x_t^{TP}-x_t^{FP_d}}(\tilde{\alpha}_s^{t+1})^{x_t^{FP_s}}(1-\tilde{\alpha}_s^{t+1})^{(Js_t)-x_t^{FP_s}}}{(1-\tilde{\alpha}_d^t-\tilde{\beta}^t)^{x_t^{TP}}(\tilde{\alpha}_d^t)^{x_t^{FP_d}}(\tilde{\beta}^t)^{(Jd_t)-x_t^{TP}-x_t^{FP_d}}(\tilde{\alpha}_s^{t+1})^{x_t^{FP_s}}(1-\tilde{\alpha}_s^{t+1})^{(Js_t)-x_t^{FP_s}}} \\
&= \frac{(1-\tilde{\alpha}_d^*-\tilde{\beta}^t)^{x_t^{TP}}(\tilde{\alpha}_d^*)^{x_t^{FP_d}}}{(1-\tilde{\alpha}_d^t-\tilde{\beta}^t)^{x_t^{TP}}(\tilde{\alpha}_d^t)^{x_t^{FP_d}}}.
\end{aligned}
$$

7. Assign a new $\tilde{\alpha}_d$ value according to

$$
\tilde{\alpha}_d^{t+1} = \begin{cases} \tilde{\alpha}_d^* & \text{with probability } \min(1,R) \\ \tilde{\alpha}_d^t & \text{otherwise.} \end{cases}
$$

8. Propose $\tilde{\beta}^* \sim$ Unif(0,0.3).

9. Calculate

$$
\begin{aligned}
R &= \frac{\Pr(\tilde{\beta}^{*}|\tilde{\alpha}_s^{t+1},\tilde{\alpha}_d^{t+1},\tilde{p}^t,c^t,Y,\mathbb{W}_1,\mathbb{W}_2)\Pr(\tilde{\beta}^t)}{\Pr(\tilde{\beta}^{t}|\tilde{\alpha}_s^{t+1},\tilde{\alpha}_d^{t+1},\tilde{p}^t,c^t,Y,\mathbb{W}_1,\mathbb{W}_2)\Pr(\tilde{\beta}^{*})} \\
&= \Big[\Pr(Y|\tilde{\alpha}_s^{t+1},\tilde{\alpha}_d^{t+1},\tilde{\beta}^{*},\tilde{p}_{eff}^t,c^t,\mathbb{W}_1,\mathbb{W}_2)\Pr(c_l^t|\tilde{p}^t,\mathbb{W}_1,\mathbb{W}_2) \\
&\quad \times \Pr(\tilde{\alpha}_s^{t+1})\Pr(\tilde{\alpha}_d^{t+1}),\Pr(\tilde{\beta}^{*})\Pr(\tilde{p}^t)\Pr(c^t)\Pr(\tilde{\beta}^t)\Big] \\
&\div \Big[\Pr(Y|\tilde{\alpha}_s^{t+1},\tilde{\alpha}_d^{t+1},\tilde{\beta}^{t},\tilde{p}^t,c^t,\mathbb{W}_1,\mathbb{W}_2)\Pr(c^t|\tilde{p}^t,\mathbb{W}_1,\mathbb{W}_2) \\
&\quad \times \Pr(\tilde{\alpha}_s^{t+1})\Pr(\tilde{\alpha}_d^{t+1}),\Pr(\tilde{\beta}^{t})\Pr(\tilde{p}^t)\Pr(c^t)\Pr(\tilde{\beta}^{*})\Big] \\
&= \frac{\Pr(Y|\tilde{\alpha}_s^{t+1},\tilde{\alpha}_d^{t+1},\tilde{\beta}^{*},\tilde{p}^t,c^t,\mathbb{W}_1,\mathbb{W}_2)}{\Pr(Y|\tilde{\alpha}_s^{t+1},\tilde{\alpha}_d^{t+1},\tilde{\beta}^{t},\tilde{p}^t,c^t,\mathbb{W}_1,\mathbb{W}_2)} \\
&= \frac{(1-\tilde{\alpha}_d^{t+1}-\tilde{\beta}^{*})^{x_t^{TP}}(\tilde{\alpha}_d^{t+1})^{x_t^{FP_d}}(\tilde{\beta}^{*})^{(Jd_t)-x_t^{TP}-x_t^{FP_d}}(\tilde{\alpha}_s^{t+1})^{x_t^{FP_s}}(1-\tilde{\alpha}_s^{t+1})^{(Js_t)-x_t^{FP_s}}}{(1-\tilde{\alpha}_d^{t+1}-\tilde{\beta}^{t})^{x_t^{TP}}(\tilde{\alpha}_d^{t+1})^{x_t^{FP_d}}(\tilde{\beta}^{t})^{(Jd_t)-x_t^{TP}-x_t^{FP_d}}(\tilde{\alpha}_s^{t+1})^{x_t^{FP_s}}(1-\tilde{\alpha}_s^{t+1})^{(Js_t)-x_t^{FP_s}}} \\
&= \frac{(1-\tilde{\alpha}_d^{t+1}-\tilde{\beta}^{*})^{x_t^{TP}}(\tilde{\beta}^{*})^{(Jd_t)-x_t^{TP}-x_t^{FP_d}}}{(1-\tilde{\alpha}_d^{t+1}-\tilde{\beta}^{t})^{x_t^{TP}}(\tilde{\beta}^{t})^{(Jd_t)-x_t^{TP}-x_t^{FP_d}}}.
\end{aligned}
$$

10. Assign a new $\tilde{\beta}^{t+1}$ value according to

$$
\tilde{\beta}^{t+1} = \begin{cases} \tilde{\beta}^{*} & \text{with probability } \min(1,R) \\ \tilde{\beta}^{t} & \text{otherwise.} \end{cases}
$$

11. Propose $\tilde{p}^{*} \sim$ Unif(0,1).

12. Calculate

$$
\begin{aligned}
R &= \frac{\Pr(\tilde{p}^*|\tilde{\alpha}_s^{t+1},\tilde{\alpha}_d^{t+1},\tilde{\beta}^{t+1},c^t,Y,\mathbb{W}_1,\mathbb{W}_2)\Pr(\tilde{p}^t)}{\Pr(\tilde{p}^t|\tilde{\alpha}_s^{t+1},\tilde{\alpha}_d^{t+1},\tilde{\beta}^{t+1},c^t,Y,\mathbb{W}_1,\mathbb{W}_2)\Pr(\tilde{p}^*)} \\
&= \Big[\Pr(Y|\tilde{\alpha}_s^{t+1},\tilde{\alpha}_d^{t+1},\tilde{\beta}^{t+1},\tilde{p}^*,c^t,\mathbb{W}_1,\mathbb{W}_2)\Pr(c^t|\tilde{p}^*,\mathbb{W}_1,\mathbb{W}_2) \\
&\quad \times \Pr(\tilde{\alpha}_s^{t+1})\Pr(\tilde{\alpha}_d^{t+1}),\Pr(\tilde{\beta}^{t+1})\Pr(\tilde{p}^*)\Pr(c^t)\Pr(\tilde{p}^t)\Big] \\
&\div \Big[\Pr(Y|\tilde{\alpha}_s^{t+1},\tilde{\alpha}_d^{t+1},\tilde{\beta}^{t},\tilde{p}^t,c^t,\mathbb{W}_1,\mathbb{W}_2)\Pr(c_l^t|\tilde{p}^t,\mathbb{W}_1,\mathbb{W}_2) \\
&\quad \times \Pr(\tilde{\alpha}_s^{t+1})\Pr(\tilde{\alpha}_d^{t+1}),\Pr(\tilde{\beta}^{t+1})\Pr(\tilde{p}^t)\Pr(c^t)\Pr(\tilde{p}^*)\Big] \\
&= \frac{\Pr(Y|\tilde{\alpha}_s^{t+1},\tilde{\alpha}_d^{t+1},\tilde{\beta}^{t+1},\tilde{p}^*,c^t,\mathbb{W}_1,\mathbb{W}_2)\Pr(c^t|\tilde{p}^*,\mathbb{W}_1,\mathbb{W}_2)}{\Pr(Y|\tilde{\alpha}_s^{t+1},\tilde{\alpha}_d^{t+1},\tilde{\beta}^{t+1},\tilde{p}^t,c^t,\mathbb{W}_1,\mathbb{W}_2)\Pr(c^t|\tilde{p}^t,\mathbb{W}_1,\mathbb{W}_2)} \\
&= \Big[(1-\tilde{\alpha}_d^{t+1}-\tilde{\beta}^{t+1})^{x_t^{TP}}(\tilde{\alpha}_d^{t+1})^{x_t^{FP_d}}(\tilde{\beta}^{t+1})^{Jd_t-x_t^{TP}-x_t^{FP_d}}(\tilde{\alpha}_s^{t+1})^{x_t^{FP_s}} \\
&\times \ (1-\tilde{\alpha}_s^{t+1})^{Js_t-x_t^{FP_s}}(\tilde{p}^*)^{Jd_t}(1-\tilde{p}^*)^{Js_t}\Big] \\
&\div \Big[(1-\tilde{\alpha}_d^{t+1}-\tilde{\beta}^{t+1})^{x_t^{TP}}(\tilde{\alpha}_d^{t+1})^{x_t^{FP_d}}(\tilde{\beta}^{t+1})^{Jd_t-x_t^{TP}-x_t^{FP_d}}(\tilde{\alpha}_s^{t+1})^{x_t^{FP_s}} \\
&\times \ (1-\tilde{\alpha}_s^{t+1})^{Js_t-x_t^{FP_s}}(\tilde{p}^t)^{Jd_t}(1-\tilde{p}^t)^{Js_t}\Big] \\
&= \frac{(\tilde{p}^*)^{Jd_t}(1-\tilde{p}^*)^{Js_t}}{(\tilde{p}^t)^{Jd_t}(1-\tilde{p}^t)^{Js_t}}.
\end{aligned}
$$

13. Assign a new $\tilde{p}^{t+1}$ value according to

$$
\tilde{p}^{t+1} = \begin{cases} \tilde{p}^* & \text{with probability } \min(1,R) \\ \tilde{p}^t & \text{otherwise.} \end{cases}
$$

14. Propose c^* via the random walk procedure described above.

15. Calculate

$$
\begin{aligned}
R &= \frac{\Pr(c^*|\tilde{\alpha}_s^{t+1},\tilde{\alpha}_d^{t+1},\tilde{\beta}^{t+1},\tilde{p}^{t+1},Y,\mathbb{W}_1,\mathbb{W}_2)\Pr(c^t|c^*)}{\Pr(c^t|\tilde{\alpha}_s^{t+1},\tilde{\alpha}_d^{t+1},\tilde{\beta}^{t+1},\tilde{p}^{t+1},Y,\mathbb{W}_1,\mathbb{W}_2)\Pr(c^*|c^t)} \\
&= \Big[\Pr(Y|\tilde{\alpha}_s^{t+1},\tilde{\alpha}_d^{t+1},\tilde{\beta}^{t+1},\tilde{p}^{t+1},c^*,\mathbb{W}_1,\mathbb{W}_2)\Pr(c^*|\tilde{p}^{t+1},\mathbb{W}_1,\mathbb{W}_2) \\
&\quad \times \Pr(\tilde{\alpha}_s^{t+1})\Pr(\tilde{\alpha}_d^{t+1}),\Pr(\tilde{\beta}^{t+1})\Pr(\tilde{p}^{t+1})\Pr(c^*)\Pr(c^t|c^*)\Big] \\
&\div \Big[\Pr(Y|\tilde{\alpha}_s^{t+1},\tilde{\alpha}_d^{t+1},\tilde{\beta}^{t+1},\tilde{p}^{t+1},c^t,\mathbb{W}_1,\mathbb{W}_2)\Pr(c^t|\tilde{p}^{t+1},\mathbb{W}_1,\mathbb{W}_2) \\
&\quad \times \Pr(\tilde{\alpha}_s^{t+1})\Pr(\tilde{\alpha}_d^{t+1}),\Pr(\tilde{\beta}^{t+1})\Pr(\tilde{p}^{t+1})\Pr(c^t)\Pr(c^*|c^t)\Big] \\
&= \frac{\Pr(Y|\tilde{\alpha}_s^{t+1},\tilde{\alpha}_d^{t+1},\tilde{\beta}^{t+1},\tilde{p}^{t+1},c^*,\mathbb{W}_1,\mathbb{W}_2)\Pr(c^*|\tilde{p}^{t+1},\mathbb{W}_1,\mathbb{W}_2)}{\Pr(Y|\tilde{\alpha}_s^{t+1},\tilde{\alpha}_d^{t+1},\tilde{\beta}^{t+1},\tilde{p}^{t+1},c^t,\mathbb{W}_1,\mathbb{W}_2)\Pr(c^t|\tilde{p}^{t+1},\mathbb{W}_1,\mathbb{W}_2)} \\
&= \Big[(1-\tilde{\alpha}_d^{t+1}-\tilde{\beta}^{t+1})^{x_*^{TP}}(\tilde{\alpha}_d^{t+1})^{x_*^{FP_d}}(\tilde{\beta}^{t+1})^{Jd_*-x_t^{TP}-x_*^{FP_d}}(\tilde{\alpha}_s^{t+1})^{x_*^{FP_s}} \\
&\quad \times(1-\tilde{\alpha}_s^{t+1})^{Js_*-x_*^{FP_s}}(\tilde{p}^{t+1})^{Jd_*}(1-\tilde{p}^{t+1})^{Js_*}\Big] \\
&\div \Big[(1-\tilde{\alpha}_d^{t+1}-\tilde{\beta}^{t+1})^{x_t^{TP}}(\tilde{\alpha}_d^{t+1})^{x_t^{FP_d}}(\tilde{\beta}^{t+1})^{Jd_t-x_t^{TP}-x_t^{FP_d}}(\tilde{\alpha}_s^{t+1})^{x_t^{FP_s}} \\
&\quad \times(1-\tilde{\alpha}_s^{t+1})^{Js_t-x_t^{FP_s}}(\tilde{p}^{t+1})^{Jd_t}(1-\tilde{p}^{t+1})^{Js_t}\Big].
\end{aligned}
$$

Note that $\Pr(c^0|c^*)$ and $\Pr(c^*|c^0)$ cancel.

16. Assign new c^1 value according to

$$
c_l^1 = \begin{cases} c^* & \text{with probability } \min(1,R) \\ c^0 & \text{otherwise.} \end{cases}
$$

17. Increment t and return to step 2.

3.8 Discussion

This chapter provided the necessary notation for my model, introduced the important concept of effective parameters, and gave an in-depth look at the data, model and estimation framework for my second approach to estimating detection probability for capture-recapture surveys with uncertain recaptures. The next chapter provides the framework and performance evaluation results for a simulation study of this approach.

CHAPTER 4

MONTE CARLO EVALUATION OF PERFORMANCE

4.1 Sampling Procedure

4.1.1 Overview

I conducted a simulation study to examine estimation performance for my approach. For simplicity and computational ease, my primary evaluation study examined only one time interval. Though a multiple time interval study would be desirable, determining whether my model performs well for one time interval is a necessary step, for if my model does not perform well for one interval a multiple interval study becomes moot.

There are various factors to consider in a Monte Carlo experiment. At the highest level, performance and behaviour are examined for different passage rates for a specified length of time for one time interval. This is a way of controlling survey population sizes in a way that can be compared to empirical evidence from the survey, and my choices for these parameters are based on that evidence. Let us call this highest level of simulation factors "scenario". The next level, nested in the first, examines different $w_{1,1}$ and $w_{2,1}$ pairs, i.e., different number of whales seen at each perch. This allows me to evaluate the performance for estimating sighting rates. Next, for given $w_{1,1}$ and $w_{2,1}$ values, several latent capture histories, without matching errors, are sampled. Finally, given the previous information, matching errors are applied randomly to generate the final error prone data sets with known effective parameter values. These are the data sets to which my method is applied and estimation of performance is evaluated based on these analyses.

4.1.2 First Level of Sampling

Having given an overview of the simulation framework, I now describe the sample process in detail. If there are N_1 true number of animals migrating past in time interval 1 then there are $\sum_{j=0}^{N_1} \frac{N_1!}{(N_1-j)!(j)!}$ different possible $w_{1,1}$ values and the same number of different possible $w_{2,1}$ values. For all combinations of those two variables, there are K_1 ways to achieve each given $w_{1,1}$ and $w_{2,1}$ pair where K_1 is defined in Section 3.1 and further described in Section 3.5. The enormity of the sampling space precluded using a large random sample at this phase, so instead of randomly selecting many $w_{1,1}$ and $w_{2,1}$ pairs for each scenario, I selected five $w_{1,1}$ and $w_{2,1}$ pairs in such a way as to give a representative feel to the overall distribution of the $w_{1,1}$ and $w_{2,1}$ pairs. Further discussion on the choices for N_i and the 5 different pairs of $w_{1,1}$ and $w_{2,1}$ for each scenario can be found in section 4.1.5.

I treat these five samples as case studies, keeping in mind that the information each contains about the parameters is imperfect. The way that I chose the five cases is as follows. Although N_1 and p vary depending on the scenario, the overall sampling scheme is the same. Let $X \sim Bin(N_1, p)$ and $Y \sim Bin(N_1, p)$ where the realizations of X and Y represent the number of sightings at each perch and X and Y are independent. Now define $W = (W_1, W_2)$ where $W_1 = min(X, Y)$ and $W_2 = max(X, Y)$. The realizations of W_1 and W_2 represent the number of sightings at perch 1 and perch 2 respectively (the choice of which perch has the greater number of sightings is irrelevant). The mean and standard deviation for W_1 and W_2 are easily calculated. Denote μ_1 as the mean of W_1 and σ_1 as the standard deviation of W_1. Define μ_2 and σ_2 for W_2 similarly. The five pairs of $W = (w_{1,1}, w_{2,1})$ that were chosen are:

1. (μ_1, μ_2)
2. $(\mu_1, \mu_2 - \sigma_2)$
3. $(\mu_1 - \sigma_1, \mu_2)$

4. $(\mu_1, \mu_2 + \sigma_2)$

5. $(\mu_1 + \sigma_1, \mu_2)$

It is important to note that $\mu_1, \mu_2, \sigma_1, \sigma_2$ are not whole numbers but $w_{1,1}$ and $w_{2,1}$ always are. Therefore, I rounded to the nearest integer.

4.1.3 Second Level of Sampling

Given a $(w_{1,1}, w_{2,1})$ pair recall that there are K_1 different possibilities regarding which sightings are part of a duplicate individual and which of those sightings are part of a single individual. However, sampling directly from the different possibilities is difficult since I would need to enumerate all K_1 possibilities. Instead, I sample from a duplicate sighting type, T_R^d, because there are only $w_{1,1}$ different duplicate sighting types. Recall from Section 3.2.2 that a duplicate sighting type is defined as a certain combination of duplicate and single sightings given $w_{1,1}$ and $w_{2,1}$ in the space after removal of illogical connections. Once I have sampled a duplicate type I can randomly select an outcome of which sightings are duplicate sightings and which are single sightings within that duplicate type because all outcomes within a duplicate type have equal probability. This randomly selected outcome represents a latent capture history without matching errors applied during the study in time interval 1. I repeat this sampling scheme 20 times. In summary:

$$\text{Duplicate Type} = \begin{cases} 1 & \text{with probability} \quad \Pr(T_{R_1}^d | p, w_{1,1}, w_{2,1}) \\ 2 & \text{with probability} \quad \Pr(T_{R_2}^d | p, w_{1,1}, w_{2,1}) \\ \vdots & \\ w_{1,1} & \text{with probability} \quad nT_{R_{w_{1,1}}}^d \Pr(T_{R_{w_{1,1}}}^d | p, w_{1,1}, w_{2,1}) \end{cases}$$

and the latent capture history without matching errors applied, given the selected duplicate Type $= j$, is selected with probability $\frac{1}{nT_{R_j}^d}$ for $j = 1, \ldots, N_{T_R}^d$.

4.1.4 Third Level of Sampling

The third level of sampling is similar to the second level of sampling in the sense that types are randomly selected (a duplicate type for the second level of sampling or an error set type for the third level of sampling). Given one latent capture history without matching errors applied (generated from the second level of sampling, described above) I can enumerate all the possible error set types in the reduced space described in Section 3.2.1. Enumeration of these error set types is challenging due to the enormity of possible error set types and extensive coding was needed to accomplish the enumeration. Thus, I can select an error set type based on each error set type's probability, $\Pr(T^e_{R_t}|\alpha_s, \alpha_d, \beta, w_{1,1}, w_{2,1})$. Once an error set type is selected, I randomly select 5 outcomes (i.e., which sightings are false positive duplicates, false positive single, or false negatives) where each outcome in an error set type has equal probability. The five selections represent the matching process conducted by 5 different researchers. I repeat the third level of sampling 25 times. Notice that since a latent capture history is known along with α_s, α_d, β and p, I can calculate the effective parameter values for all sampled data sets. I am then able to compare the estimates from my model to these effective parameter values.

4.1.5 Summary

The final outcome of the sampling scheme is an error prone data set with known effective parameter values. To summarize, it is obtained as follows. There are four scenarios that will be explored. For each of these scenarios, 5 pairs of $w_{1,1}$ and $w_{2,1}$ are selected in the first sampling step. Given these pairs, 20 latent capture histories are randomly selected in step 2. Given the outcomes of the previous two steps, 25 error prone data sets are selected in step 3. Finally, step 3 is repeated for 8 different combination of α_s, α_d and β values.

Table 4.1: $w_{1,1}$ and $w_{2,1}$ pairs chosen for each scenario.

	Scenario 1	**Scenario 2**	**Scenario 3**	**Scenario 4**
(μ_1, μ_2)	(22, 28)	(46, 54)	(23, 27)	(47, 53)
$(\mu_1, \mu_2 - \sigma_2)$	(22, 24)	(46, 48)	(23, 24)	(47, 49)
$(\mu_1 - \sigma_1, \mu_2)$	(19, 28)	(41, 54)	(20, 27)	(43, 53)
$(\mu_1, \mu_2 + \sigma 2)$	(22, 31)	(46, 59)	(23, 30)	(47, 57)
$(\mu_1 + \sigma 1, \mu_2)$	(26, 28)	(52, 54)	(26, 27)	(51, 53)

I call these combination of errors "cases". Therefore, the simulation study involves 2,500 data sets for each case in each scenario. Thus, overall my model will be tested against 80,000 data sets.

Scenarios 1 and 3 were chosen to resemble the bowhead data in the sense that the simulated data might correspond to a whale passage rate roughly equal to the mode of observed passage rates during the bowhead study. Specifically, N and p are chosen to achieve the desired passage rate. Scenario 1 has $N = 125$ and $p = 0.2$. Scenario 3 has N=50 and $p = 0.5$. Scenarios 2 and 4 were chosen to have passage rates of 20 whales per hour, which corresponds to the 95th percentile of the observed passage rates for the bowhead study. Specifically, scenario 2 has $N = 250$ and $p = 0.2$. Scenario 4 has $N = 100$ and $p = 0.5$. In all four scenarios I used a 2.5 hour time interval, similar to the time intervals used during the bowhead study, although this choice is unimportant except for calibrating total sample sizes.

For each scenario, five pairs of $w_{1,1}$ and $w_{2,1}$ were chosen using the process described in Section 4.1.2. A summary of those values can be found in Table 4.1. As you can see in Figure 4.1 the five pairs of $w_{1,1}$ and $w_{2,1}$ chosen for scenario 1 do a decent job of representing the overall distribution. Similar results are seen for the other three scenarios.

For each $(w_{1,1}, w_{2,1})$ pair in each scenario, eight different cases where examined. A summary of those values can be found in Table 4.2. Recall that the effective parameter values are cal-

Table 4.2: Typical true values of $\tilde{\alpha}_s$, $\tilde{\alpha}_d$, $\tilde{\beta}$ for the 8 cases.

	Case 1	**Case 2**	**Case 3**	**Case 4**	**Case 5**	**Case 6**	**Case 7**	**Case 8**
$\tilde{\alpha}_s$	.1	.1	.05	.1	.05	.05	.05	.05
$\tilde{\alpha}_d$	.05	.1	.15	.05	.05	.1	.05	.05
$\tilde{\beta}$	.05	.1	.15	.1	.05	.1	.1	.2

culated based on the unrestricted parameter values (see Chapter 3). I chose the unrestricted parameter values and calculated the effective parameter values for each data set. Note that each simulated data set has its own varying effective parameter values, that is, there are 80,000 different effective parameter values and due to the variability between every data set, not all effective parameters within each case will be the same. However, all have values near the values shown in Table 4.2. So instead of listing them all out and since all are similar, only the typical values are shown.

4.2 Model Performance Results

4.2.1 Overview

For every simulated data set I set the initial $\tilde{\alpha}_s$, $\tilde{\alpha}_d$, $\tilde{\beta}$ and $\tilde{p}$ values as .1, .1, .25, and .5 respectively. I set the initial configuration as the data produced from the first matching process. The Markov Chain Monte Carlo procedure was run for 40,000 iterations and a burn-in of 20,000 was used.

Colorado State University's Cray super computer was used to run the simulations. To use the Cray my simulations were broken up into 10 batch jobs per case for each scenario. Thus, a total of 320 jobs were run where each job completed analysis for 250 data sets using 192 cores. Each job for scenario 1 and 3 took anywhere between 12 and 24 hours, while each job for scenario 2 and 4 took anywhere from 24 and 48 hours. The reason for the length of each job has to do with the enumeration and generation of error prone data sets and not the

Markov Chain Monte Carlo estimation process applied to each data set. Scenario 2 and 4 took almost twice the time as scenario 1 and 3 because the number of potential error prone data sets was significantly larger. If I assume that only one job was run at a time then the maximum time needed to complete all of scenario 1 and 3 was between 80 and 160 days and the maximum time needed to complete all of scenario 2 and 4 was between 160 and 320 days. However, depending on the computational needs of other Cray users, up to 10 jobs could have been run simultaneously, thus reducing the overall computational time for this simulation study to a more reasonable but still lengthy amount of time.

4.2.2 Informative Data

In classical capture recapture studies a higher detection probability provides more information about the population. However, with the presence of errors, especially with respect to duplicate sightings, a higher detection probability provides more opportunities to make these errors, and in such situations there is a huge number of ways to do so. Replications in the multiple matching processes are where the statistical evidence of these errors come from and since these matching processes are made independently, the chances of the matchers making the exact same errors are low. Thus, in a sense, there is actually less information about each sightings true capture history because of this and there is certainty less information available for use in estimation. One potential way to counteract this dilution of evidence is to increase the number of matching decisions processes, or replications. However, the amount of replications needed to be beneficial is not known and I do not explore this further in this dissertation.

Let's look at an example. Suppose that the true number of duplicate sightings is 6 (implying the number of duplicate individuals is 3). If the first match process yields 4 true positive matches (recall that the errors are defined in terms of sightings and not individuals), there are $\binom{3}{2} = 3$ ways to do that. Note that if one duplicate sighting is determined to be a

true positive then its complimentary sighting is automatically also determined to be a true positive. If the second match process also calls for 4 true positive matches, there are again $\binom{3}{2} = 3$ ways to do that. The chances that the two match processes make the same true positive matches is $\frac{1}{3} \times \frac{1}{3} = \frac{1}{9}$, due to the independence of the matching process. Now suppose that the true number of duplicate sightings is 30 (thus the number of duplicate individuals is 15). If the first match process makes 4 true positive matches, there are $\binom{15}{2} = 105$ ways to make those true positive matches. If the second match process also makes 4 true positive matches there are again $\binom{15}{2} = 105$ ways to make those matches. The chances of match process 1 and match process 2 making the same true positive matches is $\frac{1}{105} \times \frac{1}{105} = \frac{1}{11025}$. Therefore, the higher the detection probability (i.e. the more duplicate sightings) with the presence of errors, the lower the chances of the matchers making identical match decisions and thus less information is provided for estimation of the parameters, all else being equal.

The above situation can be seen in the results discussed below, where scenario 1 and 2 (with lower detection probabilities) overall perform better than scenario 3 and 4 (with higher detection probabilities).

4.2.3 Summary Statistics and Visual Graphics

I calculated the coverage rates (for 95% credible intervals) for the estimates from each simulated data set. That is, I found the percentage of 95% credible intervals that contain the true effective parameter value for each case in each scenario. The results for each scenario can be found in Tables 4.3, 4.4, 4.5, and 4.6. Figures 4.2, 4.3, and 4.4 provide visual comparisons. Figure 4.2 separately graphs coverage rates for estimating each parameter, spread out across cases in the x-axis, with different symbols for each scenario. Figure 4.3 graphs coverage rates across effect parameter on the x-axis with each symbol and color representing each

Table 4.3: 95% Coverage Rates of 95% credible intervals for $\tilde{\alpha}_s$, $\tilde{\alpha}_d$, $\tilde{\beta}$ and $\tilde{p}$ for the 8 cases in scenario 1.

Scenario 1								
	Case 1	**Case 2**	**Case 3**	**Case 4**	**Case 5**	**Case 6**	**Case 7**	**Case 8**
$\tilde{\alpha}_s$	.936	.898	.872	.888	.918	.933	.900	.806
$\tilde{\alpha}_d$	.902	.887	.915	.933	.900	.885	.931	.942
$\tilde{\beta}$	.924	.943	.956	.938	.912	.948	.900	.901
$\tilde{p}$	.940	.933	.936	.934	.941	.941	.940	.910

Table 4.4: 95% Coverage Rates of 95% credible intervals for $\tilde{\alpha}_s$, $\tilde{\alpha}_d$, $\tilde{\beta}$ and $\tilde{p}$ for the 8 cases in scenario 2.

Scenario 2								
	Case 1	**Case 2**	**Case 3**	**Case 4**	**Case 5**	**Case 6**	**Case 7**	**Case 8**
$\tilde{\alpha}_s$	.928	.865	.861	.848	.937	.936	.903	.755
$\tilde{\alpha}_d$	.890	.886	.910	.898	.895	.887	.906	.918
$\tilde{\beta}$	.936	.916	.903	.870	.924	.935	.890	.829
$\tilde{p}$	.954	.929	.923	.920	.952	.951	.951	.886

case and scenario. Figure 4.4 separately graphs coverage rates for each effective parameter across scenario on the x-axis. The different colors represent the different cases.

As I have stated previously, I simulated 80,000 data sets. Checking the performance of all the 80,000 data sets is not practical. Instead, I randomly chose one data set for case 1 and case 8 for each scenario, and provided each effective parameters trace plots. Case 1 and case 8 where chosen to represent one that resulted in decent estimates and one that resulted in poor estimates, respectively. These can be found in Figures 4.5, 4.6, 4.7, 4.8, 4.9, 4.10, 4.11, and 4.12. What is shown in these figures is broadly similar to what is seen with the remaining unshown cases.

Table 4.5: 95% Coverage Rates of 95% credible intervals for $\tilde{\alpha}_s$, $\tilde{\alpha}_d$, $\tilde{\beta}$ and $\tilde{p}$ for the 8 cases in scenario 3.

	Scenario 3							
	Case 1	**Case 2**	**Case 3**	**Case 4**	**Case 5**	**Case 6**	**Case 7**	**Case 8**
$\tilde{\alpha}_s$	.912	.860	.756	.828	.903	.921	.868	.662
$\tilde{\alpha}_d$	.889	.892	.878	.904	.902	.838	.900	.927
$\tilde{\beta}$	.920	.910	.855	.879	.919	.913	.877	.782
$\tilde{p}$	.944	.909	.823	.911	.945	.944	.936	.754

Table 4.6: 95% Coverage Rates of 95% credible intervals for $\tilde{\alpha}_s$, $\tilde{\alpha}_d$, $\tilde{\beta}$ and $\tilde{p}$ for the 8 cases in scenario 4.

	Scenario 4							
	Case 1	**Case 2**	**Case 3**	**Case 4**	**Case 5**	**Case 6**	**Case 7**	**Case 8**
$\tilde{\alpha}_s$	.916	.810	.750	.773	.935	.934	.880	.568
$\tilde{\alpha}_d$	.891	.910	.862	.924	.886	.913	.913	.885
$\tilde{\beta}$	.913	.890	.828	.856	.895	.921	.876	.679
$\tilde{p}$	.940	.901	.848	.888	.942	.942	.936	.720

4.2.4 Scenario 1 Results

Based on Table 4.3 and Figures 4.2, 4.3 and 4.4 estimation of each effective parameter in scenario 1 is relatively successful for most parameters in most cases. Although coverage rates are generally a bit lower than the target 95%, the outcomes are not too bad, especially considering that my statistical model is the first of its kind to deal with the type of data described here.

Model Performance

The trace plots found in Figure 4.5 and Figure 4.6, without burn-in, for case 1 scenario 1 and case 8 scenario 1 look good and don't suggest any problems with mixing or convergence.

4.2.5 Scenario 2 Results

Based on Table 4.4 and Figures 4.2, 4.3 and 4.4, estimating the effective parameters in scenario 2 shows variable performance with somewhat lower coverage rates and the range of the coverage rates is more than for scenario 1.

Effective Parameter Results

Coverage rate performance for $\tilde{\alpha}_s$ is best for cases 1, 5, 6 and 7 and worst in cases 2, 3, 4 and 8. Coverage rate performance for $\tilde{\alpha}_d$ is best for cases 3, 4, 5, 7 and 8 and worst in cases 1, 2 and 6. Coverage rate performance for $\tilde{\beta}$ is best for cases 1, 2, 3, 5 and 6 and worst in cases 3, 4 and 8. Coverage rate performance for $\tilde{p}$ is best for all but case 8.

Case Results

While examining each effective parameters performance is important, examining each case and how the effective parameters collectively perform is a better indicator of overall performance.

Cases 1, 5, and 7 provided good results, cases 2, 6 and 3 provided fairly decent results and cases 4 and 8 provided poor results. What distinguished cases 1, 5 and 7 from the other cases is that their overall error rates where lower than the overall error rates for the other cases. And as the overall error rates increase the performance tends to decrease. The one exception is case 3. Case 3 has the highest overall error rate of all the cases but performed slightly better than case 4 and much better than case 8. This is an interesting results which will need further exploration in future studies.

Model Performance

Just like the with scenario 1, the trace plots for scenario 2, found in Figures 4.7 and 4.8, look good and don't suggest any problems with mixing or convergence.

4.2.6 Scenario 3 Results

Less than adequate performance is seen in scenario 3 which has coverage rates ranging from fairly good to poor in some of the hardest cases as seen in Table 4.5 and Figures 4.2, 4.3 and 4.4.

Effective Parameter Results

Coverage rate performance for $\tilde{\alpha}_s$ is best for cases 1, 5, and 6 and worst in cases 2, 3, 4, 7 and 8. Coverage rate performance for $\tilde{\alpha}_d$ is best for cases 4, 5, 7 and 8 and worst in cases 1, 2, 3, and 6. Coverage rate performance for $\tilde{\beta}$ is best for cases 1, 2, 5 and 6 and worst in cases 3, 4, 7 and 8. Coverage rate performance for $\tilde{p}$ is best for all but cases 3 and 8.

Case Results

Cases 1, and 5 outperform cases 2, 6 and 7 which then outperformed cases 3, 4 and 8. Similar to scenario 2, scenario 3's performance decreases as the overall error rates increase. Scenario 3 also shows that as the $\tilde{\beta}$ value increase (simultaneously with the increase in overall error rates) the performance decreases. Notice that scenario 3 doesn't perform as well as scenario 1 and 2. This is attributed to the discussion in Section 4.2.2, stating that data sets with larger detection probability (i.e. the data set has more duplicate sightings) don't provide as much statistical evidence as data sets with lower detection probability, unless more matching opportunities are available.

Model Performance

Trace plots for scenario 3, case 1, found in Figure 4.9 look good, just like with the trace plots for scenario 1 and 2. However, the trace plots for case 8 found in Figure 4.10 show more jumpy behavior. Note that case 1 has overall lower error rates for duplicate sightings than does case 8. The poor performance for case 8 is the result of the match processes not having many overlapping matches, that is, there are not a lot of replicate match decisions between the matching processes, and the MCMC process is having trouble converging to a configuration.

4.2.7 Scenario 4 Results

Performance in scenario 4 are pretty similar to scenario 3 in all cases for all parameters with coverage rates ranging from fairly good to poor in some of the hardest cases.

Effective Parameter Results

Coverage rate performance for $\tilde{\alpha}_s$ is best for cases 1, 5, and 6 and worst in cases 2, 3, 4, 7 and 8. Coverage rate performance for $\tilde{\alpha}_d$ is best for cases 2, 4, 6, 7 and 8 and worst in cases 1, 3, 5, and 8. Coverage rate performance for $\tilde{\beta}$ is best for cases 1, 2, 5 and 6 and worst in cases 3, 4, 7 and 8. Coverage rate performance for $\tilde{p}$ is best for cases 1, 2, 5, 6 and 7 and worst for cases 3, 4 and 8.

Case Results

Cases 1, 5 and 6 had the best performance out of all the cases in scenario 4. Cases 2 and 7 performed decently while cases 3, 4 and 8 performed very poorly. These results are very similar to the results for scenario 3 with respect to the model performance for each case.

However, overall scenario 3 out performed scenario 4 in all cases except for case 6. More specifically, estimation of $\tilde{\alpha}_d$ improves quit a bit. Its coverage rate goes from .838 in scenario 3 to .913 in scenario 4. This seems counter-intuitive and further exploration of this phenomenon is needed.

Model Performance

Trace plots for scenario 4, case 1 and case 8, found in Figure 4.11 and Figure 4.12 show more jumpy behavior. This behavior is explained in Section 4.2.5.3 for scenario 3. Again, this is the reason that overall the coverage rates for scenario 4 are not as good as the coverage rates for scenario 1 and 2.

4.3 Summary

My model's performance was tested against 80,000 simulated data sets with varying effective parameter values. Performance for scenarios 1 and 2 was fairly good with almost all coverage rates being 90% and above. Performance for scenarios 3 and 4 was adequate but the number of data sets with high amount of errors is overwhelming and replication of matches between raters is low. This lack of replication between the matching processes proves difficult for mixing among configurations and discovery of highly likely configurations. Overall, though, I believe my model performed reasonably well and proves that detection probability estimation is possible even in the presence of errors in very complex capture-recapture datasets.

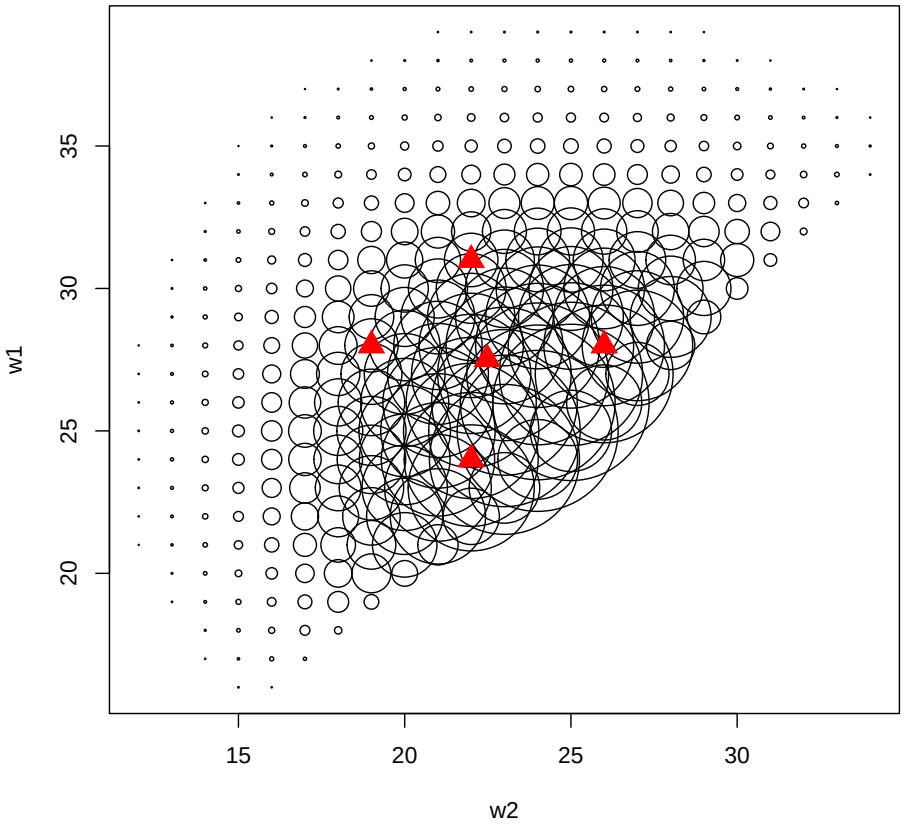

Figure 4.1: Plot of probabilities of $w_{1,1}$ and $w_{2,1}$ pairs (circle areas) for scenario 1. The five pairs of values examined in my trials correspond to the red triangles.

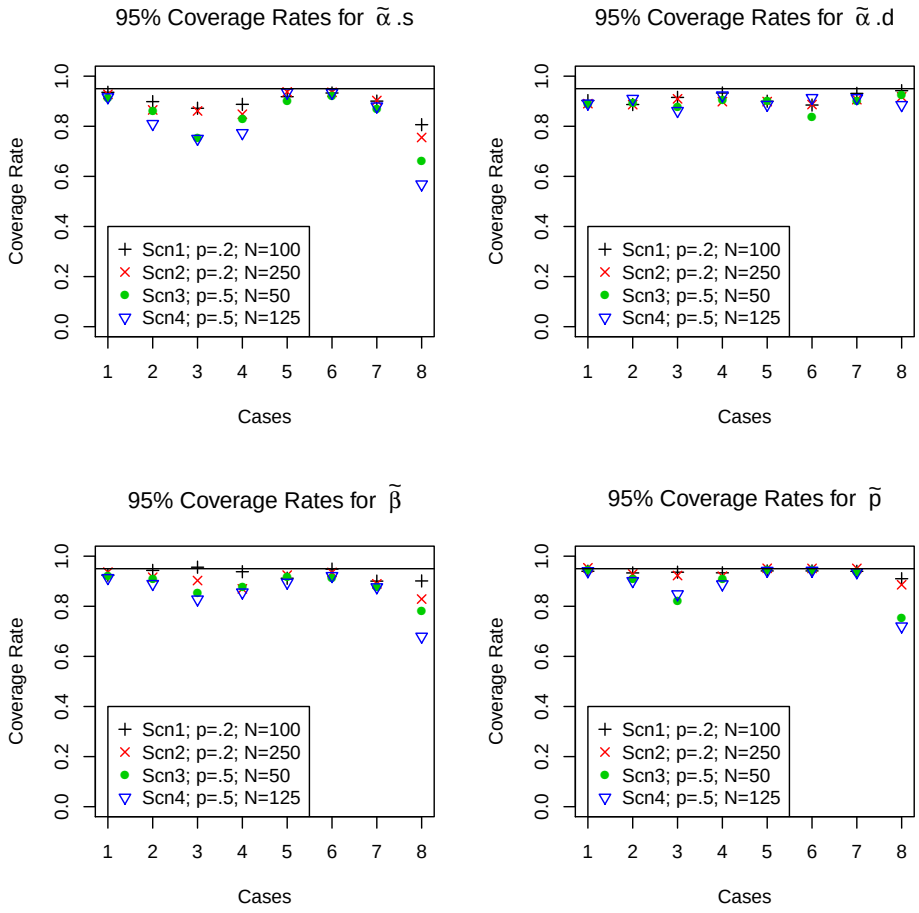

Figure 4.2: Coverage rates for 95% credible intervals for each case for $\tilde{\alpha}_s$, $\tilde{\alpha}_d$, $\tilde{\beta}$, and $\tilde{p}$.

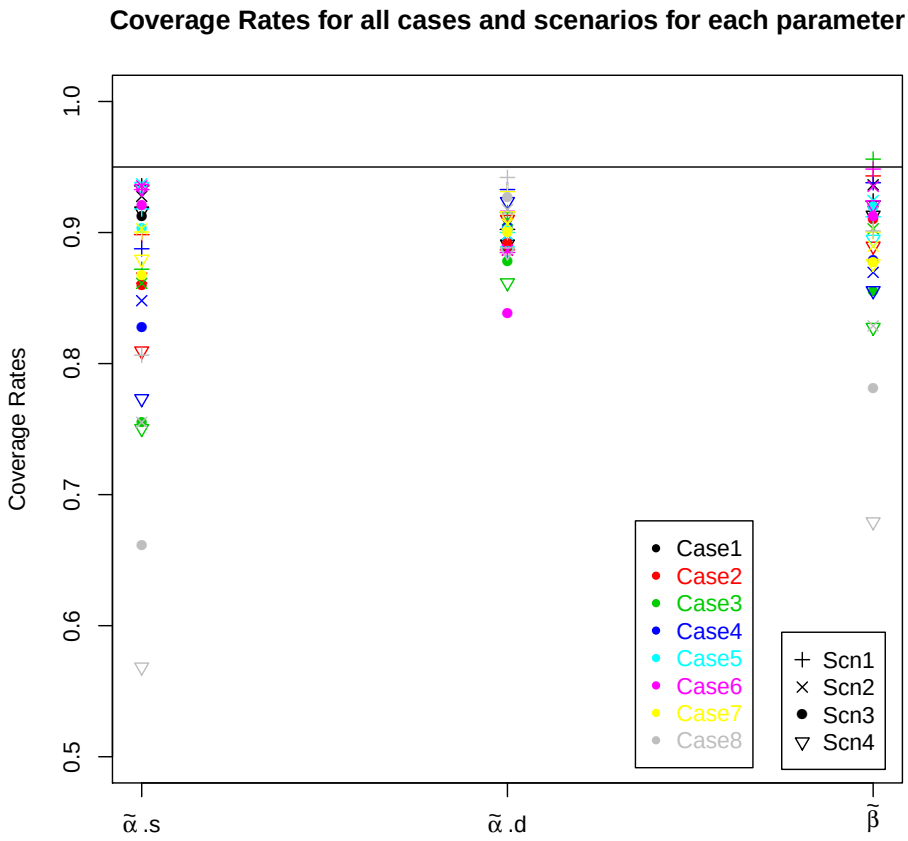

Figure 4.3: Coverage rates for 95% credible intervals for each parameter, $\tilde{\alpha}_s$, $\tilde{\alpha}_d$, and $\tilde{\beta}$.

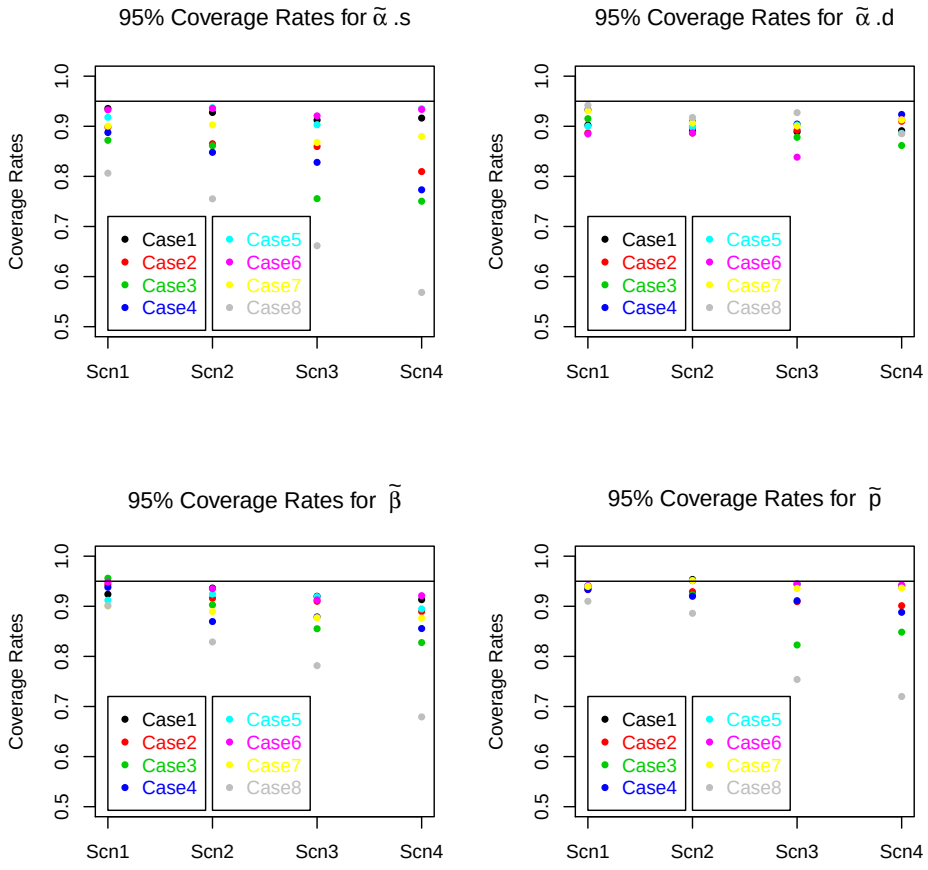

Figure 4.4: Coverage rates for 95% credible intervals for each scenario for $\tilde{\alpha}_s$, $\tilde{\alpha}_d$, $\tilde{\beta}$, and $\tilde{p}$.

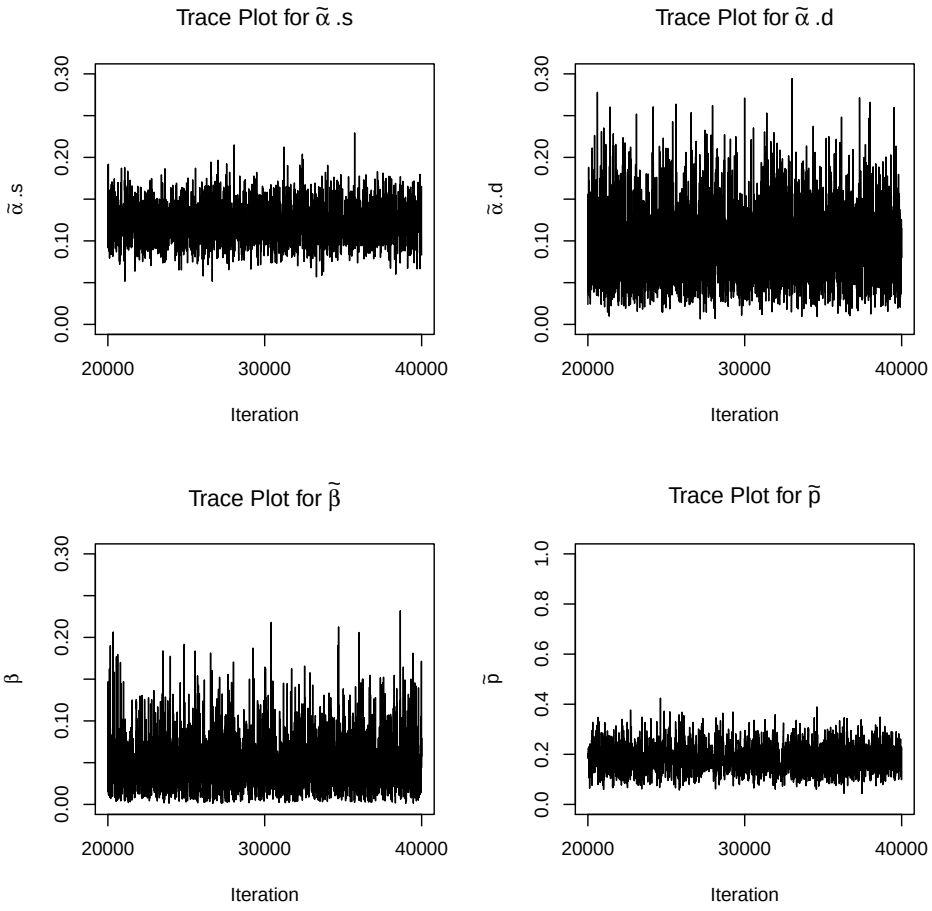

Figure 4.5: Trace plots for a simulated data set in scenario 1 case 1.

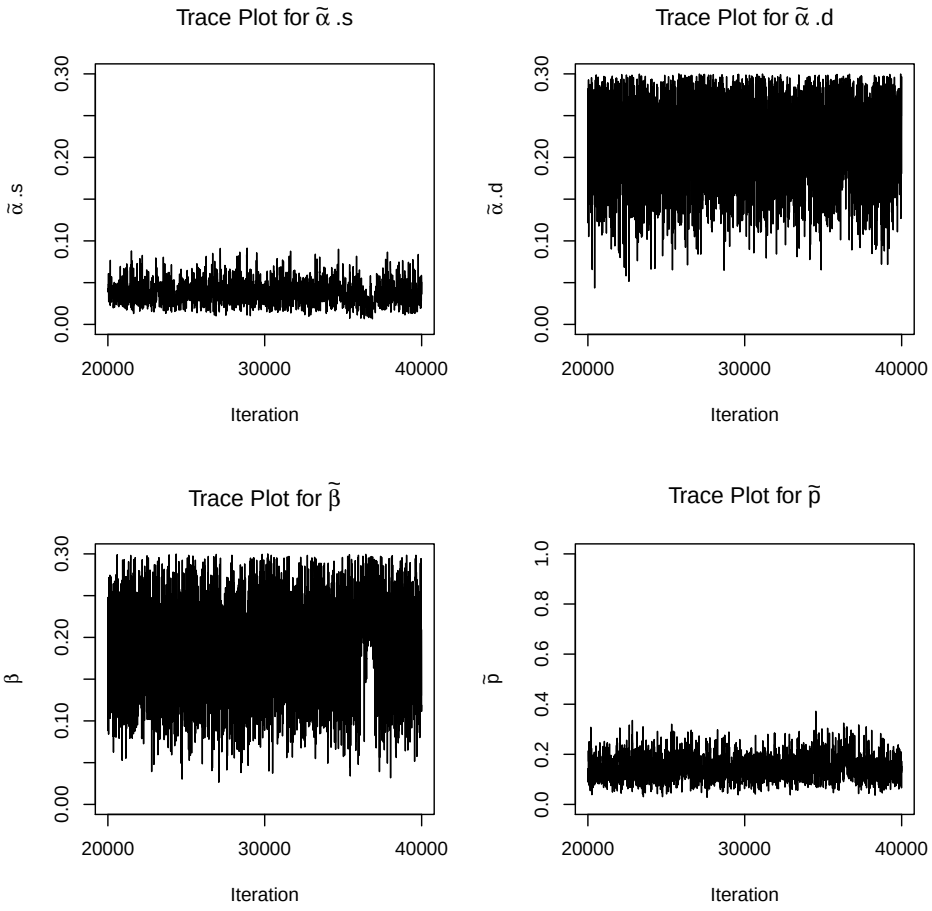

Figure 4.6: Trace plots for a simulated data set in scenario 1 case 8.

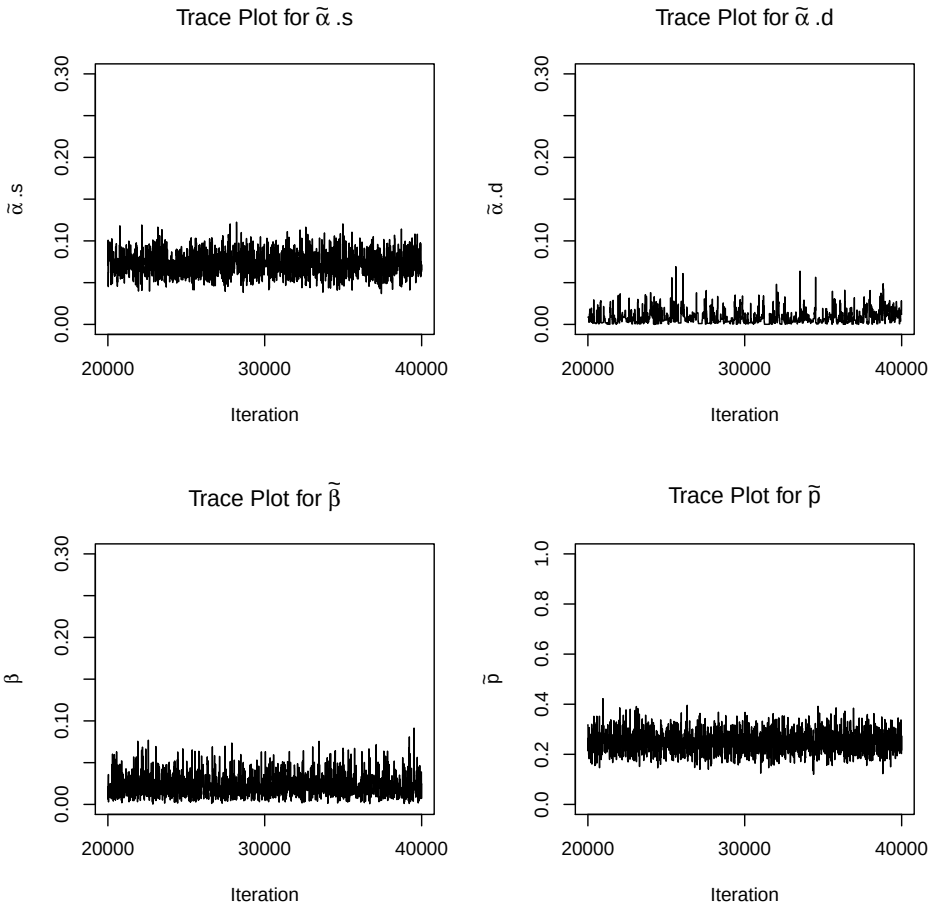

Figure 4.7: Trace plots for a simulated data set in scenario 2 case 1.

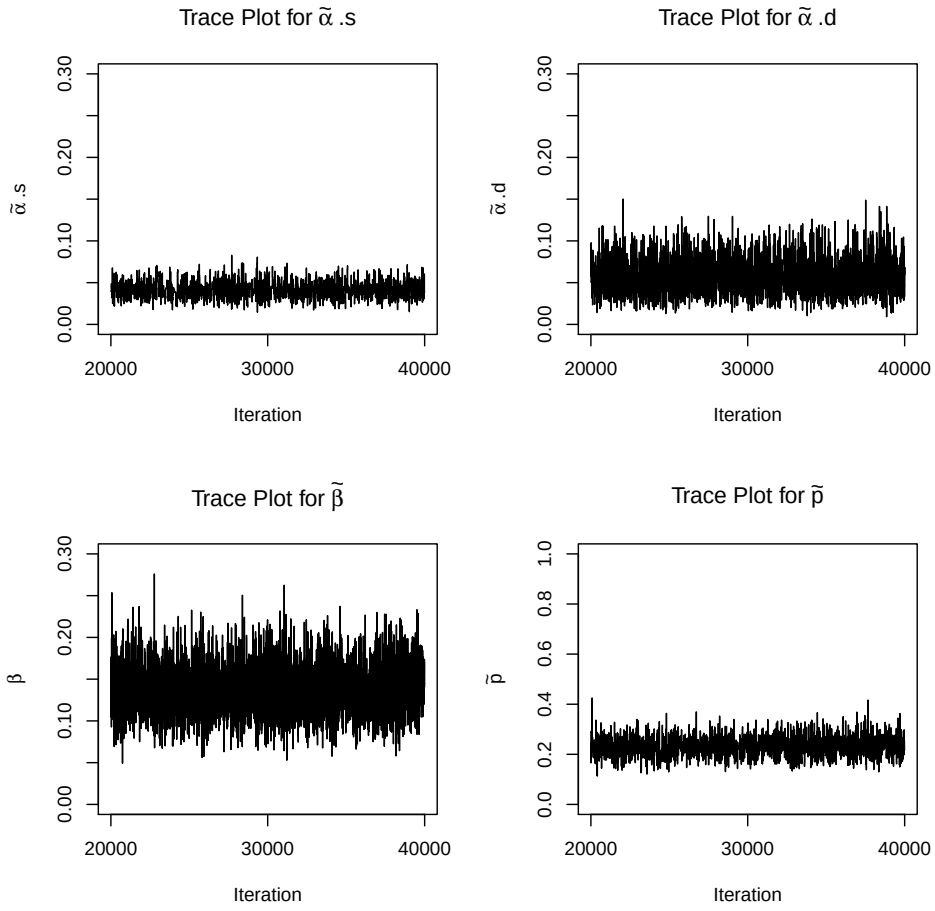

Figure 4.8: Trace plots for a simulated data set in scenario 2 case 8.

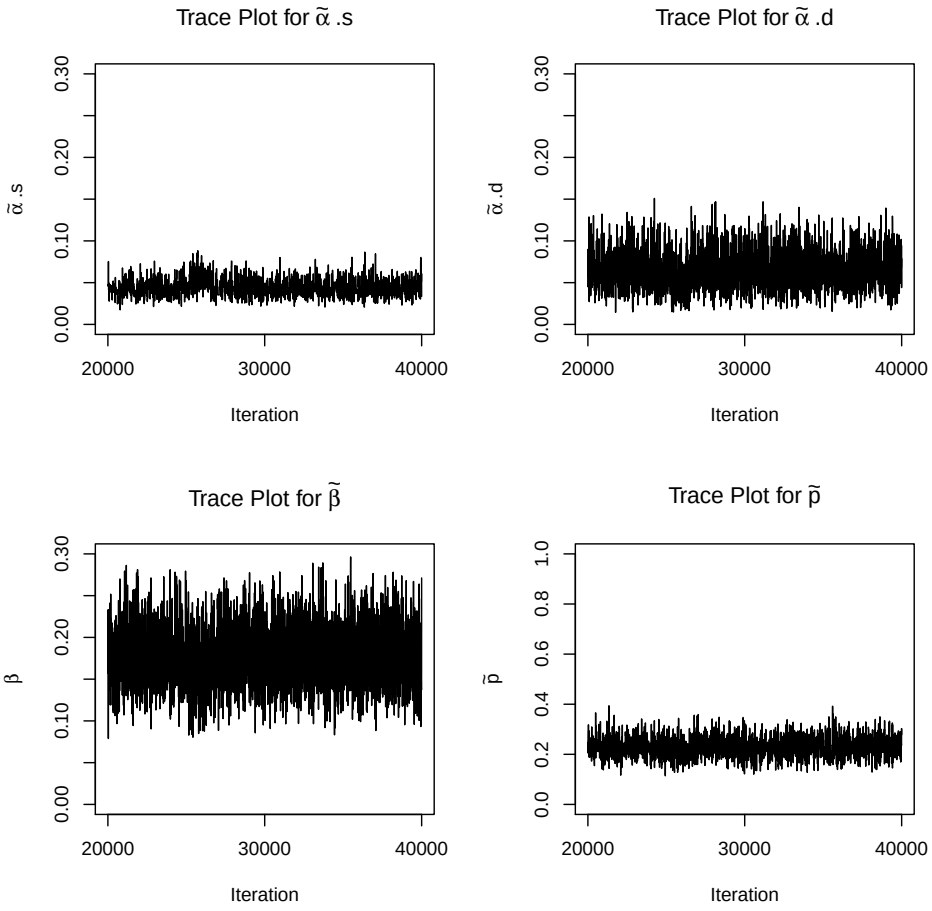

Figure 4.9: Trace plots for a simulated data set in scenario 3 case 1.

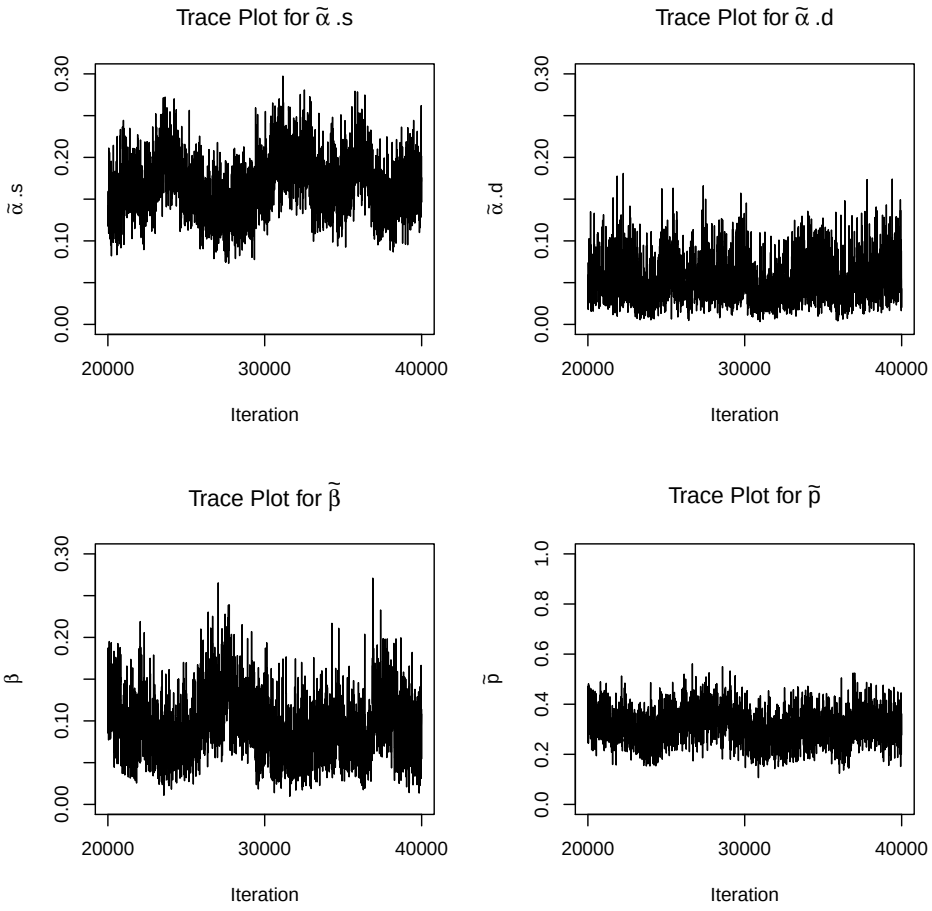

Figure 4.10: Trace plots for a simulated data set in scenario 3 case 8.

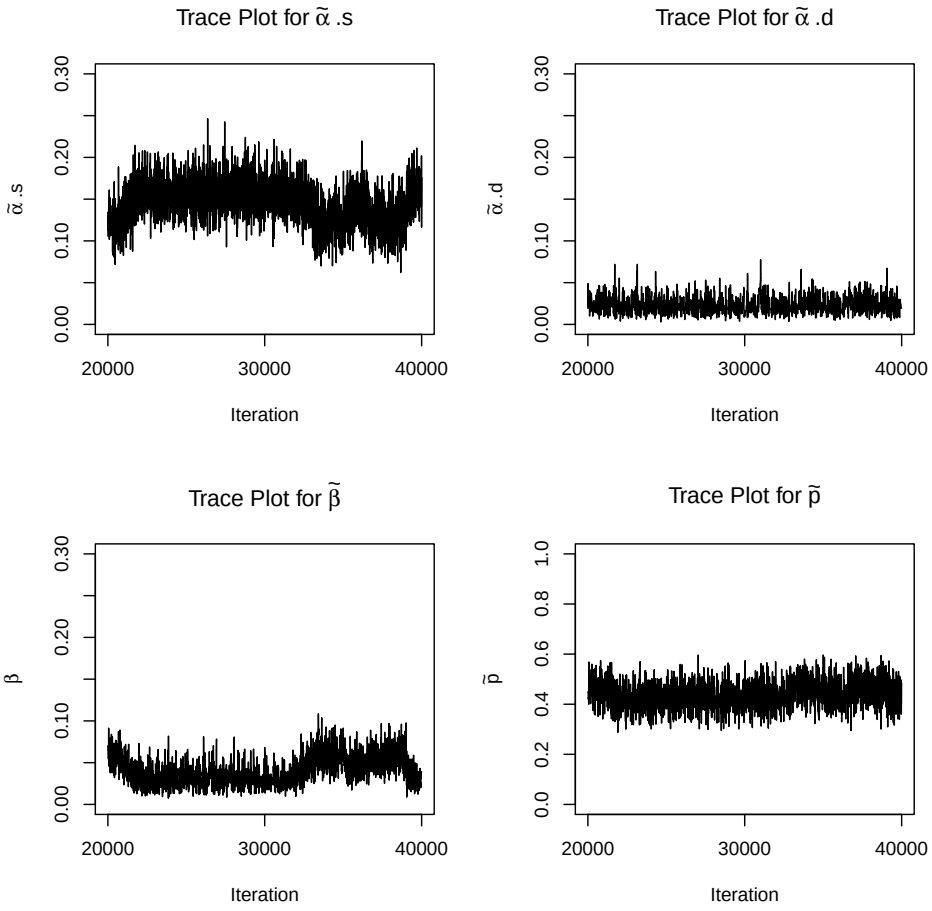

Figure 4.11: Trace plots for a simulated data set in scenario 4 case 1.

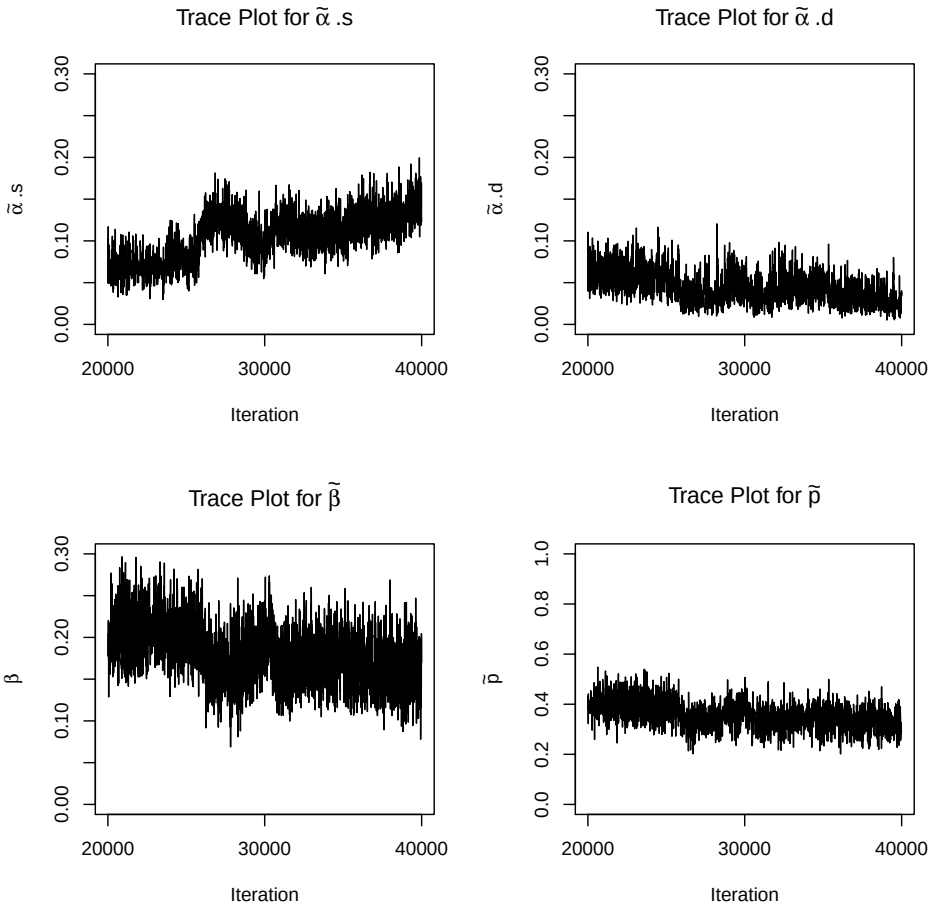

Figure 4.12: Trace plots for a simulated data set in scenario 4 case 8.

CHAPTER 5

BOWHEAD WHALE EXAMPLE

5.1 Data

I applied my model described in Chapter 3 to a subset of the 2010 bowhead whale survey data. In 2010, the raw sightings data were matched on three separate occasions. The three matching efforts were completed in May, August and September/October. For my case study, I chose a 2.5 hour time interval that had similar $w_{1,1}$ and $w_{2,1}$ values as scenarios 1 and 3 in my Monte Carlo simulations. This time interval was on April 8, 2010, between the hours of 14:15 and 16:45. During this period there were 22 sightings at the South perch and 25 sightings the North perch. Without loss of generality I refer to the South perch as perch 1 and the North perch as perch 2. Thus, $w_{1,1} = 22$ and $w_{2,1} = 25$.

Table 5.1 shows the matching data for perch 1, and Table 5.2 shows the matching data for perch 2. These tables provide the exact same information and only one is needed for analysis. Perch 1 sightings are identified using unique tags whose values range from 1026 to 1047 and perch 2 sightings are identified using unique tags whose values range from 10001 to 10025. The columns Rep 1, Rep 2 and Rep 3 represent the matching decisions for May, August, and September/October efforts respectively. In these columns, an index number identifies the sighting at the opposite perch that researchers believed to be the same whale as the sighting at the current perch (first column). A 0 indicates that the researchers believed that whale was only seen once. For example, in Table 5.1 sighting 1029 from perch 1 was thought to be the same whale as sighting 10012 from perch 2 in matching decision processes 1 and 3 (Rep 1 and 3 respectively) and thought to be sighted only at perch 1 in the second match decision process (Rep 2).

Table 5.1: Bowhead Data; 4/8/10; 14:15-16:45; Perch 1.

Perch 1 Sightings	**Rep 1**	**Rep 2**	**Rep 3**
1026	10001	10001	10001
1027	10003	10003	10003
1028	10007	10007	10007
1029	10012	0	10012
1030	10005	10005	10005
1031	10023	10006	10006
1032	0	0	10023
1033	10002	10002	10002
1034	10004	0	0
1035	0	10004	10004
1036	0	10004	10004
1037	0	0	10013
1038	10025	10025	10025
1039	0	0	10009
1040	0	0	0
1041	0	0	0
1042	0	0	0
1043	0	0	0
1044	0	0	0
1045	0	0	0
1046	0	0	0
1047	0	0	0

Table 5.2: Bowhead Data; 4/8/10; 14:15-16:45; Perch 2.

Perch 2 Sightings	**Rep 1**	**Rep 2**	**Rep 3**
10001	1026	1026	1026
10002	1033	1033	1033
10003	1027	1027	1027
10004	1034	1035	1035
10005	1030	1030	1030
10006	0	1031	1031
10007	1028	1028	1028
10008	0	1036	1036
10009	0	0	1039
10010	0	0	0
10011	0	0	0
10012	1029	0	1029
10013	0	0	1037
10014	0	0	0
10015	0	0	0
10016	0	0	0
10017	0	0	0
10018	0	0	0
10019	0	0	0
10020	0	0	0
10021	0	0	0
10022	0	0	0
10023	1031	0	1032
10024	0	0	0
10025	1038	1038	1038

5.2 Analysis

I applied my model described in chapter 3 to analyse the data. I used the same prior distributions, proposal distributions and initial starting values for each parameter as I did for the Monte Carlo simulations. Recall that in my simulations I set the initial configuration equal to the matching decisions made in the first matching process. I did the same for the bowhead data. I ran the MCMC sampler for 40,000 iterations and then used a burn-in of 20,000.

5.3 Results

The trace plots for $\tilde{\alpha}_s$, $\tilde{\alpha}_d$, $\tilde{\beta}$, and $\tilde{p}$ are shown in Figure 5.1. The red horizontal lines are the median posterior values for the MCMC estimates of each effective parameter after a burn-in period of 20,000 iterations. The trace plots for each effective parameter seem to show satisfactory mixing and appear to explore each effective parameter's space well. The acceptance rates for $\tilde{\alpha}_s$, $\tilde{\alpha}_d$, $\tilde{\beta}$, $\tilde{p}$ and configuration are .351, .247, .395, .223 and .0018, respectively. The rates for $\tilde{\alpha}_s$, $\tilde{\alpha}_d$, $\tilde{\beta}$, and $\tilde{p}$ fall within a desirable range (Givens and Hoeting, 2005) and are another indicator that the MCMC chain is mixing well. The acceptance rate for configuration is lower than the desired range. The current proposal distribution proposes configurations very similar to the current configuration, however there are a vast amount of potential configurations near the current one where most won't be supported by the data. This could be one reason for the low acceptance rate. Perhaps a different proposal distribution could be explored making less realistic configurations less likely to be proposed. That would require a new set of decision rules to determine which are less realistic configurations, however, that type of information is not clearly known. While the current proposal distribution results in a low acceptance rate, it does have the ability to explore realistic configurations. The

autocorrelation plots in Figure 5.3 show lingering lag correlation. This could be reduced by subsampling the chain, although I didn't pursue that approach.

Histograms of the posterior distributions of $\tilde{\alpha}_s$, $\tilde{\alpha}_d$, $\tilde{\beta}$, and $\tilde{p}$ can be found in Figure 5.2, and their respective summary statistics can be found in Table 5.3. The red vertical lines in the histograms indicate the posterior medians for each effective parameter. The histograms for $\tilde{\alpha}_s$ and $\tilde{p}$ are fairly symmetric, while the histograms for $\tilde{\alpha}_d$ and $\tilde{\beta}$ show slight skewness to the left. This behavior is also apparent in the summary statistics with the mean for $\tilde{\alpha}_d$ and $\tilde{\beta}$ having a slightly larger value than their median.

The mean and median for $\tilde{\alpha}_s$, $\tilde{\alpha}_d$, and $\tilde{\beta}$ are .104 and .107, .044 and .049, and .109 and .114 respectively. These results indicate that the researchers are more likely to falsely match single sightings (whales) and fail to correctly match duplicate sightings to their complimentary sighting, while they are less likely to mis-match a duplicate sighting to another sighting other than its complimentary one. The mean and median for $\tilde{p}$, .429 and .427, are near the mean detection probability (.5) that was calculated in our first approach in Chapter 2. It is reassuring that - while each approach accounts for the recapture errors in different ways - both provide similar estimates of detection probability to the level of resolution likely attainable from such a small sample size used here.

The 95% credible intervals from each parameter's posterior distributions can be found in Table 5.3. The range for $\tilde{\alpha}_s$, $\tilde{\alpha}_d$, $\tilde{\beta}$, and $\tilde{p}$ are .137, .102, .161, and .273 respectively. These credible intervals are sufficiently narrow that they provide useful information about their respective parameters. The 95% posterior credible interval for $\tilde{p}$ (.296, .569) is narrower than the 95% confidence interval calculated from our first method in Chapter 2.

Table 5.3: Posterior distribution summary statistics.

Parameters	2.5%	50%	95%	Mean
$\tilde{\alpha}_s$	.048	.104	.185	.107
$\tilde{\alpha}_d$	.0102	.044	.112	.049
$\tilde{\beta}$	.046	.109	.207	.114
$\tilde{p}$	.296	.427	.569	.429

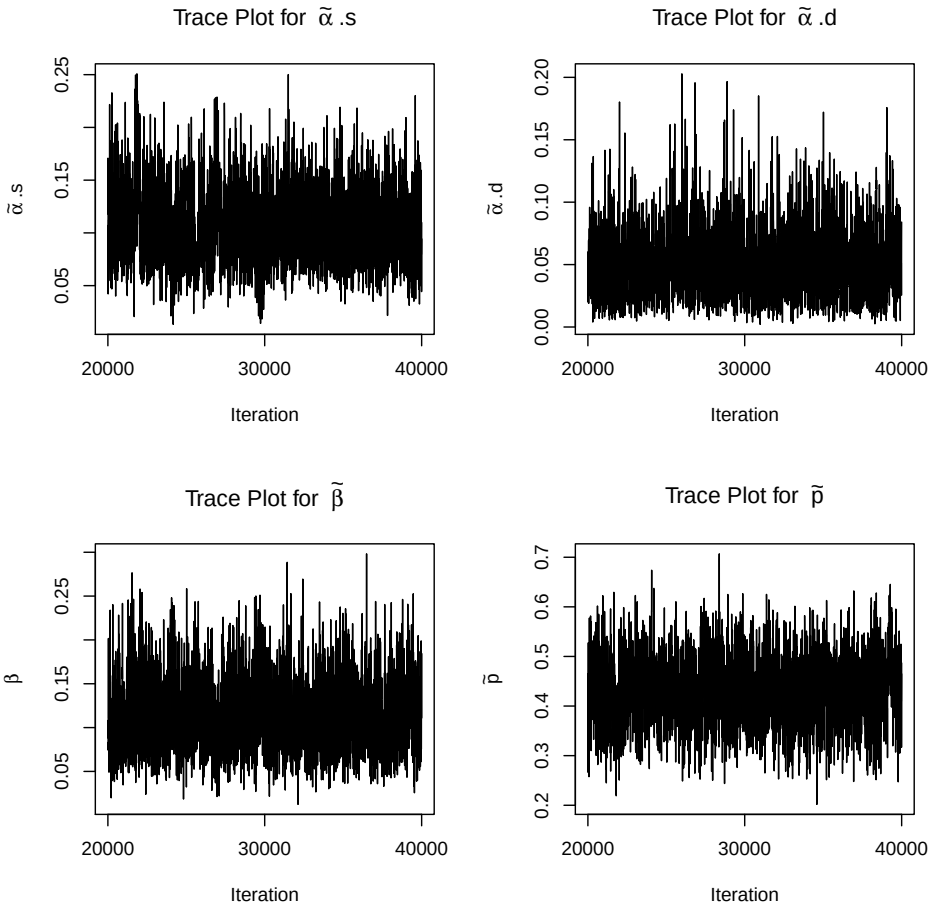

Figure 5.1: Trace plots for $\tilde{\alpha}_s$, $\tilde{\alpha}_d$, $\tilde{\beta}$, and $\tilde{p}$.

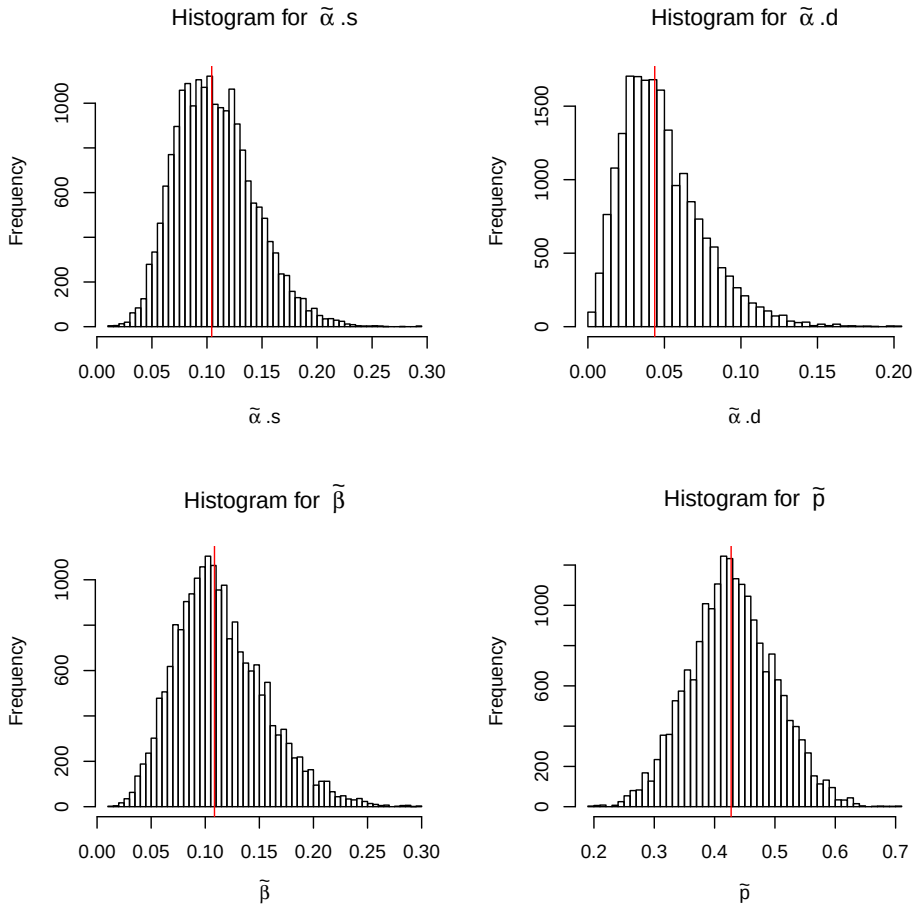

Figure 5.2: Histograms for posterior samples for $\tilde{\alpha}_s$, $\tilde{\alpha}_d$, $\tilde{\beta}$, and $\tilde{p}$.

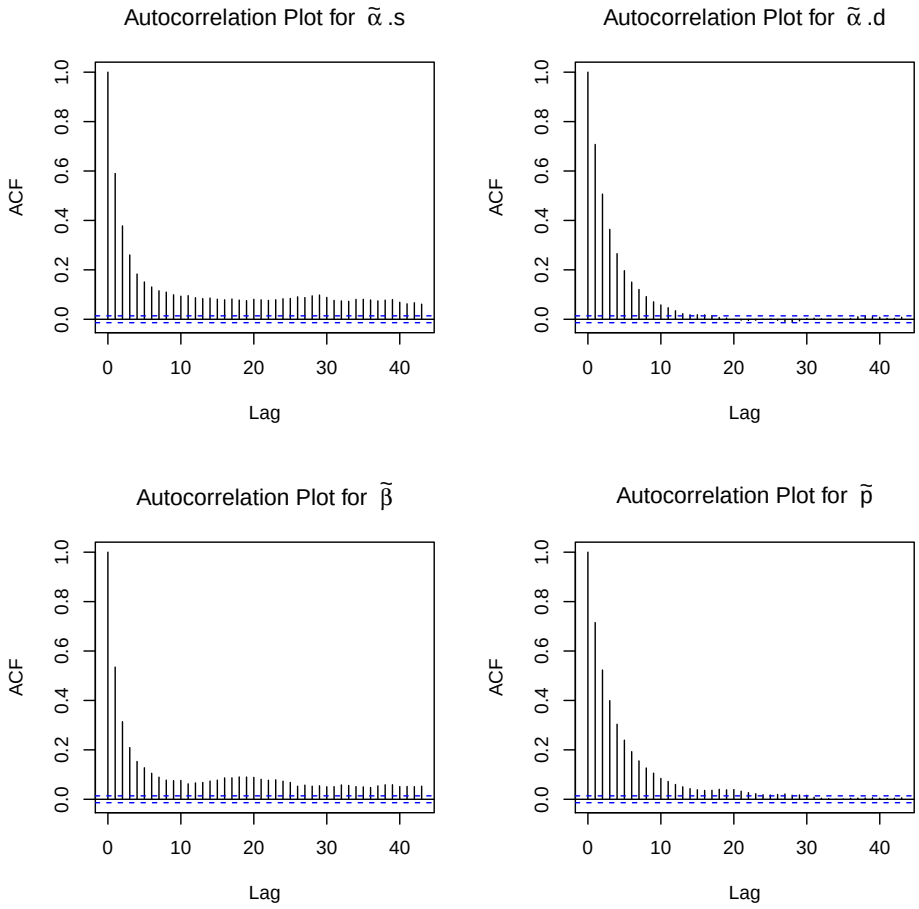

Figure 5.3: Autocorrelation plots for MCMC simulations for $\tilde{\alpha}_s$, $\tilde{\alpha}_d$, $\tilde{\beta}$, and $\tilde{p}$.

5.4 Sensitivity Analysis for Configurations

In general, mixing among configurations is the most challenging part of the MCMC analysis (see, e.g., Sections 4.2.6 and 4.2.7). To investigate the issue here, I analyzed the bowhead whale data setting the initial configuration at two extreme cases. The first extreme starting point was the configuration without any duplicate individuals. The other extreme value was one (out of many) configuration with the maximum number (22) of possible duplicate individuals. The reason I chose these as extreme cases is because these are the configurations that most likely would be the furthest from the latent configuration.

5.4.1 Results for The First Extreme Configuration

For the first sensitivity test I ran my model using the configuration with no duplicate individuals as my initial configuration; the starting values for the effective parameters were unchanged. Mirroring the first analysis, I used 40,000 iterations and a burn-in of 20,000. The results are shown in Figure 5.4 (trace plots for each effective parameter) and Figure 5.5 (the posterior histograms for each effective parameter). The trace plots show that the model is still exploring the configuration space and hasn't approached any equilibrium state yet. This is indicated by the sudden large jumps in the trace line. Specifically, it appears that the MCMC simulations become "stuck" in one local region of configuration space until a certain set of values for the other parameters allows the chain to "break out" and start exploring other regions. It appears that more MCMC iteration may be needed to achieve full mixing.

The acceptance rates for $\tilde{\alpha}_s$, $\tilde{\alpha}_d$, $\tilde{\beta}$, $\tilde{p}$ and configuration are .360, .230, .320, .210 and .0026 respectively. These acceptance rates are similar to the original analysis. However, the QQ plots between the results of the original analysis for all effective parameters and the currents results for all effective parameters shown in Figure 5.6 indicate that the estimated posterior

distributions for the original analysis and the current analysis are different. The QQ plots for $\tilde{\beta}$ and $\tilde{p}$ indicate similar shapes to their respective posterior distributions though differing in location. The QQ plot for $\tilde{\alpha}_d$ shows that the posterior distributions between the two results are practically the same. This is a very good result. Finally, the QQ plot for $\tilde{\alpha}_s$ shows different posterior distributions for the two results.

5.4.2 Results for the Second Extreme Configuration

Next, I ran my model using the configuration with the maximum number of duplicate individuals as my initial configuration. I kept the other starting values for the parameters unchanged. Again, I first ran my MCMC analysis with 40,000 iterations and a burn-in of 20,000. These results can be found in Figure 5.7 and Figure 5.8. Figure 5.7 shows the trace plots for each effective parameter and Figure 5.8 shows posterior histograms for each effective parameter. Similar to the results for the first extreme case, the trace plots show that the chain is still sticking in the configuration space.

The acceptance rates for $\tilde{\alpha}_s$, $\tilde{\alpha}_d$, $\tilde{\beta}$, $\tilde{p}$ and configuration are .344, .220, .334, .218 and .018 respectively. These acceptance rates are similar to the original analysis for all but the configurations. The QQ plots between the results from the original analysis and the currents results for all effective parameters shown in Figure 5.6 indicate that the estimated posterior distributions for the original analysis versus the current analysis are different except for $\tilde{\alpha}_d$. The QQ plots for $\tilde{\beta}$ and $\tilde{p}$ indicate similar shapes to their respective posterior distributions though differing in location except in the upper tails of each distribution. Once there, the distributions are the same. The QQ plot for $\tilde{\alpha}_d$ show that the posterior distributions between the two results are practically the same. Finally, the QQ plot for $\tilde{\alpha}_s$ show different posterior distributions for the two results.

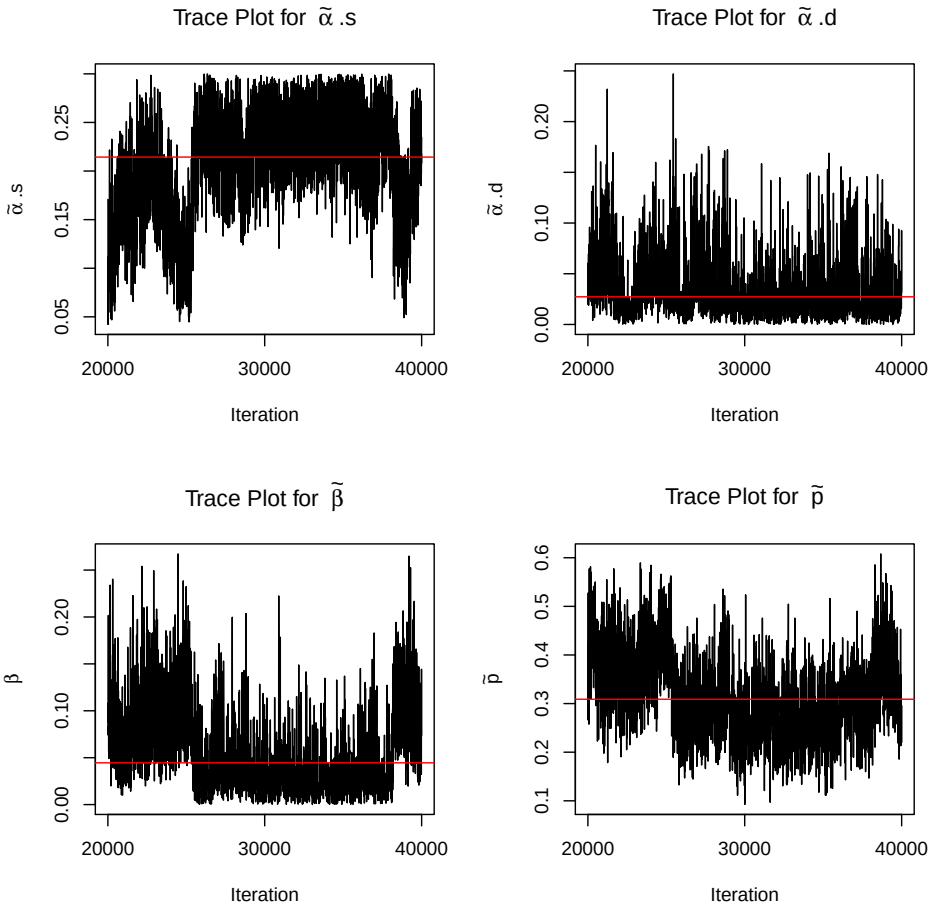

Figure 5.4: Trace plots for $\tilde{\alpha}_s$, $\tilde{\alpha}_d$, $\tilde{\beta}$, and $\tilde{p}$ for extreme case 1; 40,000 iterations, burn-in 20,000.

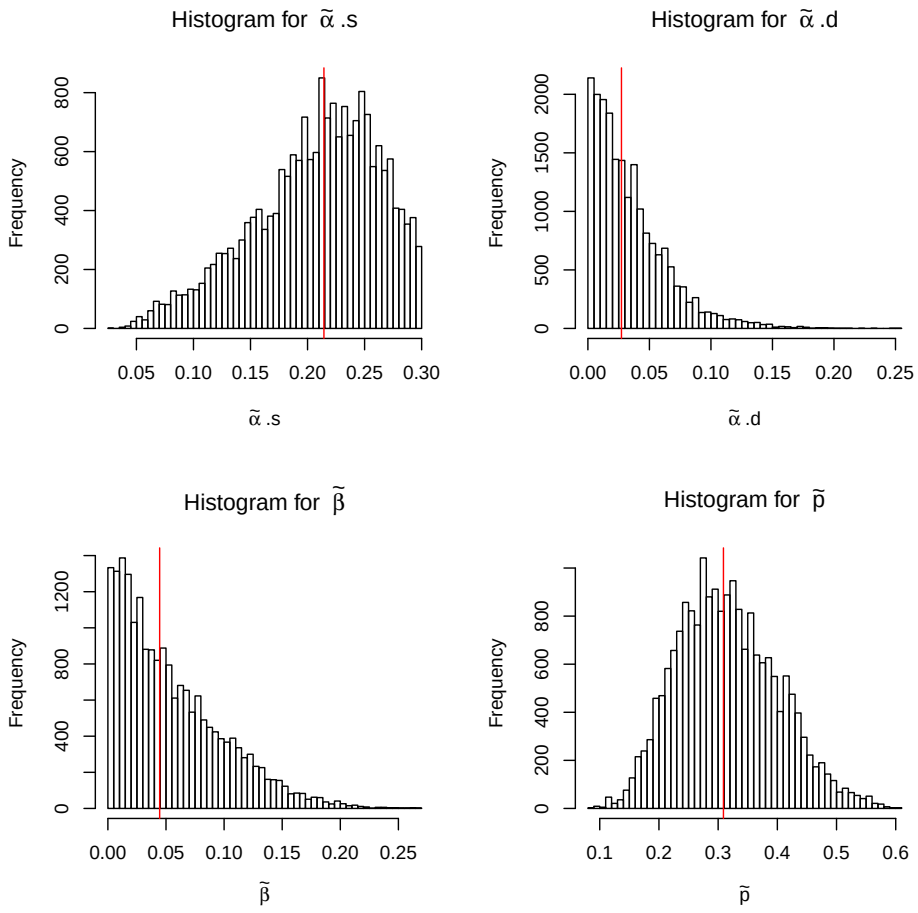

Figure 5.5: Histograms for $\tilde{\alpha}_s$, $\tilde{\alpha}_d$, $\tilde{\beta}$, and $\tilde{p}$ for extreme case 1; 40,000 iterations, burn-in 20,000.

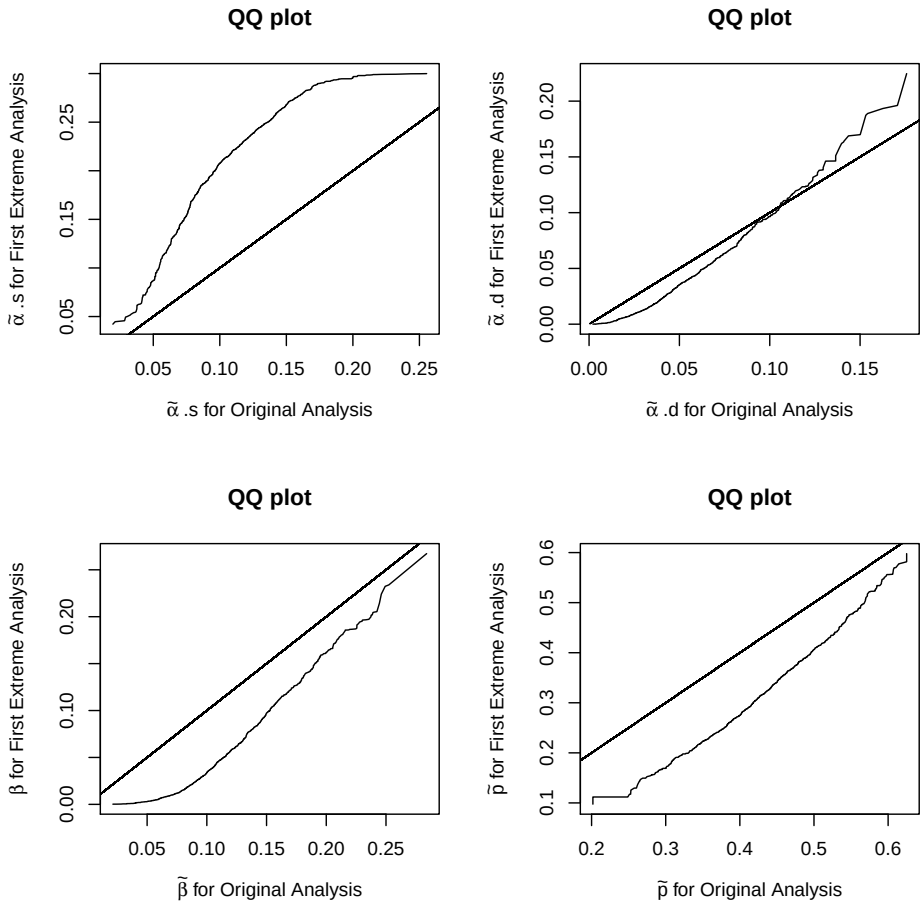

Figure 5.6: QQ plot comparing the results from the original analysis and the first extreme configuration analysis for each parameter.

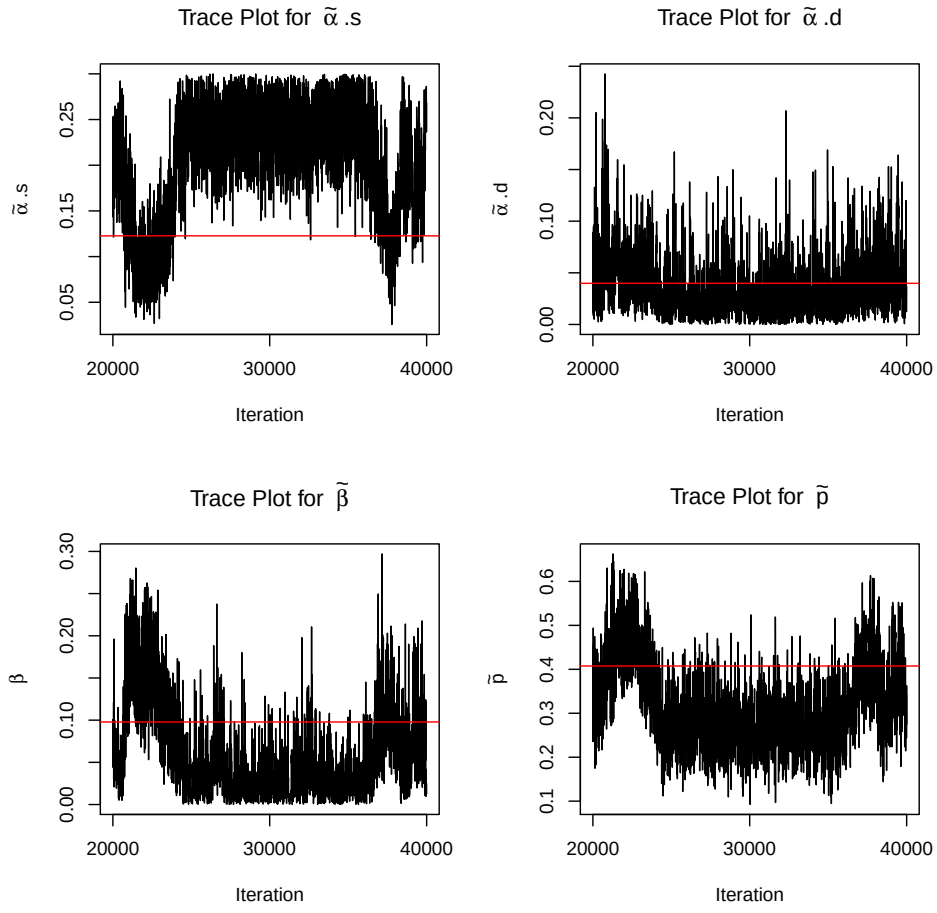

Figure 5.7: Trace plots for $\tilde{\alpha}_s$, $\tilde{\alpha}_d$, $\tilde{\beta}$, and $\tilde{p}$ for extreme case 2; 40,000 iterations, burn-in 20,000.

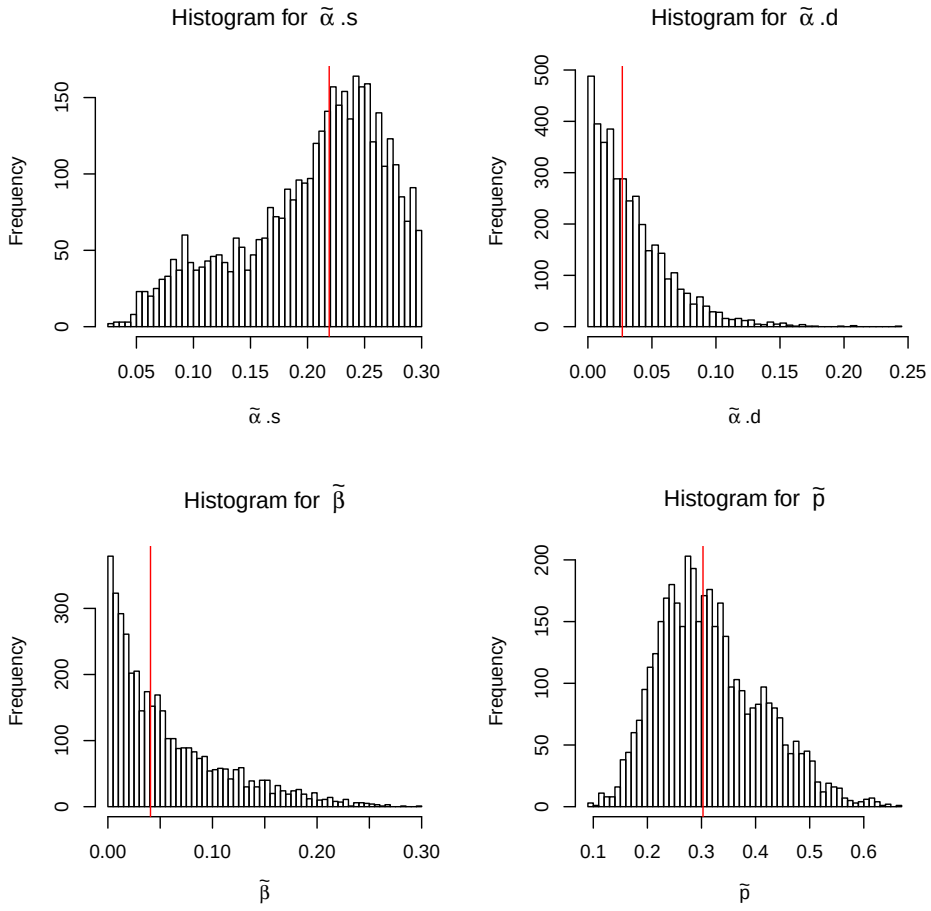

Figure 5.8: Histograms for $\tilde{\alpha}_s$, $\tilde{\alpha}_d$, $\tilde{\beta}$, and $\tilde{p}$ extreme case 2; 40,000 iterations, burn-in 20,000.

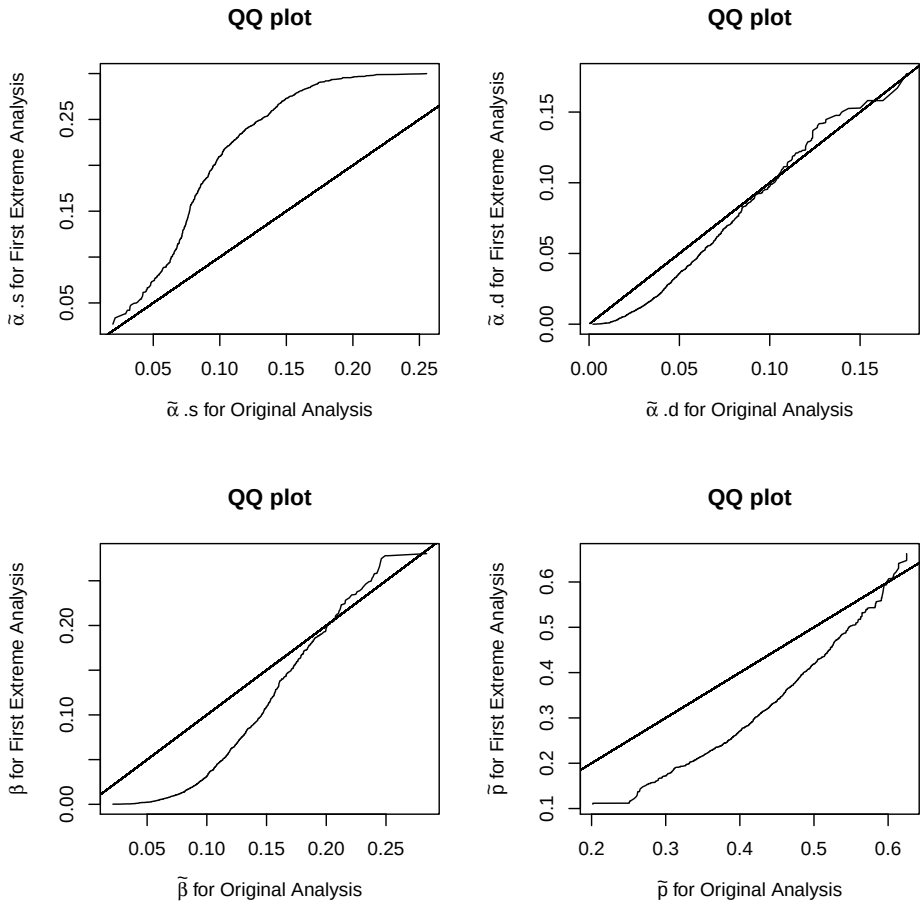

Figure 5.9: QQ plot comparing the results from the original analysis and the second extreme configuration analysis for each parameter.

Table 5.4: Posterior distribution summary statistics when using Rep 1, Rep 2 and Rep 3 as initial configuration.

	Rep 1				**Rep 2**				**Rep 3**			
Effective Parameters	2.5%	50%	95%	Mean	2.5%	50%	95%	Mean	2.5%	50%	95%	Mean
$\tilde{\alpha}_s$	.048	.104	.185	.107	.052	.112	.226	.118	.046	.119	.262	.129
$\tilde{\alpha}_d$	.010	.044	.112	.049	.009	.043	.112	.048	.010	.046	.125	.052
$\tilde{\beta}$	.046	.109	.207	.114	.037	.103	.199	.107	.027	.101	.224	.106
$\tilde{p}$	.296	.427	.569	.429	.265	.419	.563	.418	.228	.412	.575	.409

5.4.3 Sensitivity Analysis for Realistic Configurations

Due to the poor results seen when the initial configuration is considered extreme, I ran additional tests using the other two matching process decisions as initial configurations (keeping everything else the same as the original analysis) to show that with realistic starting configurations my model performs well.

To compare the results for the three different initial configurations (which are Rep 1, Rep 2 and Rep 3 respectively) I created QQ plots for all combinations (Figure 5.10, Figure 5.11 and Figure 5.12) and provided the posterior distributions summary statistics for the three runs (Table 5.4). The QQ plots show that for $\tilde{\alpha}_d$, $\tilde{\beta}$ and $\tilde{p}$ the posterior distributions are practically the same. This result is also seen in the summary statistics for those three parameters for all three runs. The QQ plots for each of the three combinations of results for $\tilde{\alpha}_s$ show that the posterior distributions are similar except for the upper tail portion of the distribution. These results are also apparent in the summary statistics for $\tilde{\alpha}_s$ for each run. The 2.5% and 50% quantiles are very similar but the 95% quantiles increase as initial configuration increases from Rep 1 to Rep 3.

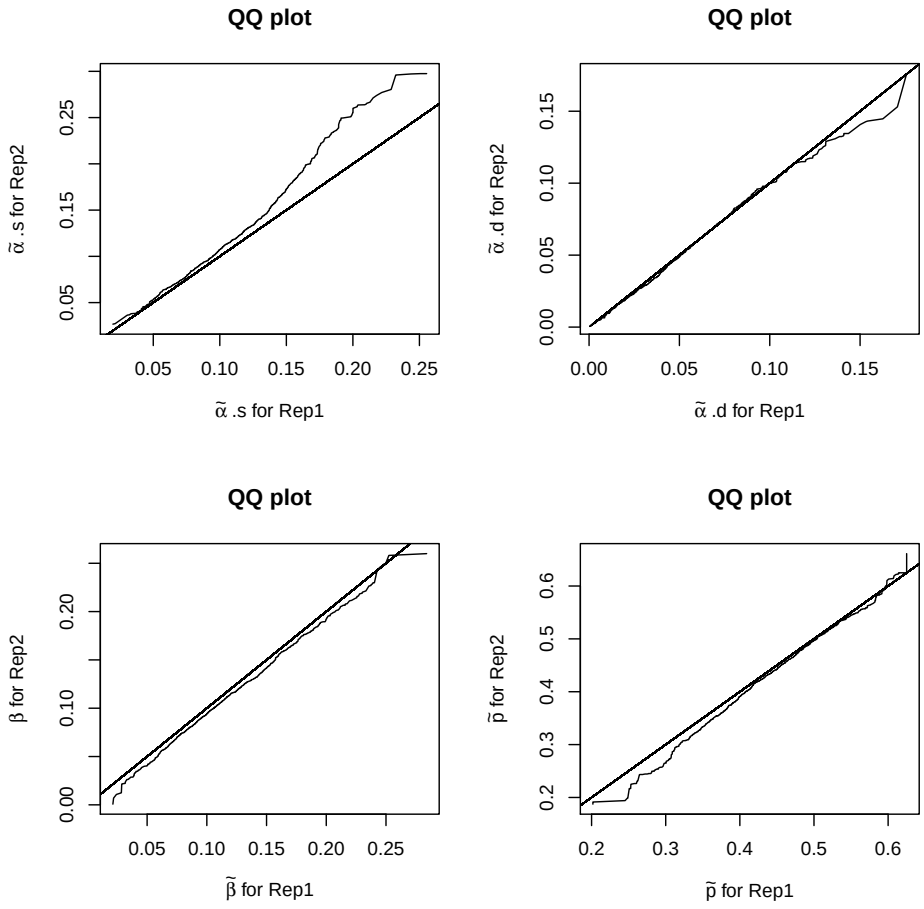

Figure 5.10: QQ plot comparing the results from the original analysis and Rep 2 as initial configuration for each parameter.

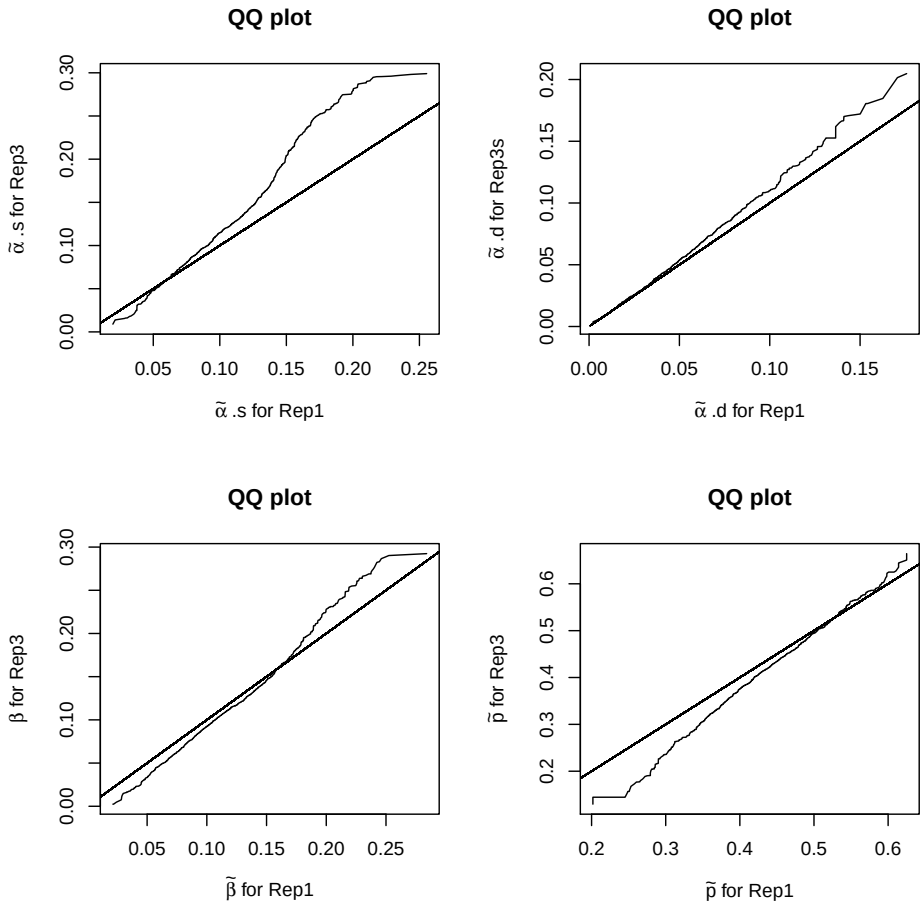

Figure 5.11: QQ plot comparing the results from the original analysis and Rep 3 as initial configuration for each parameter.

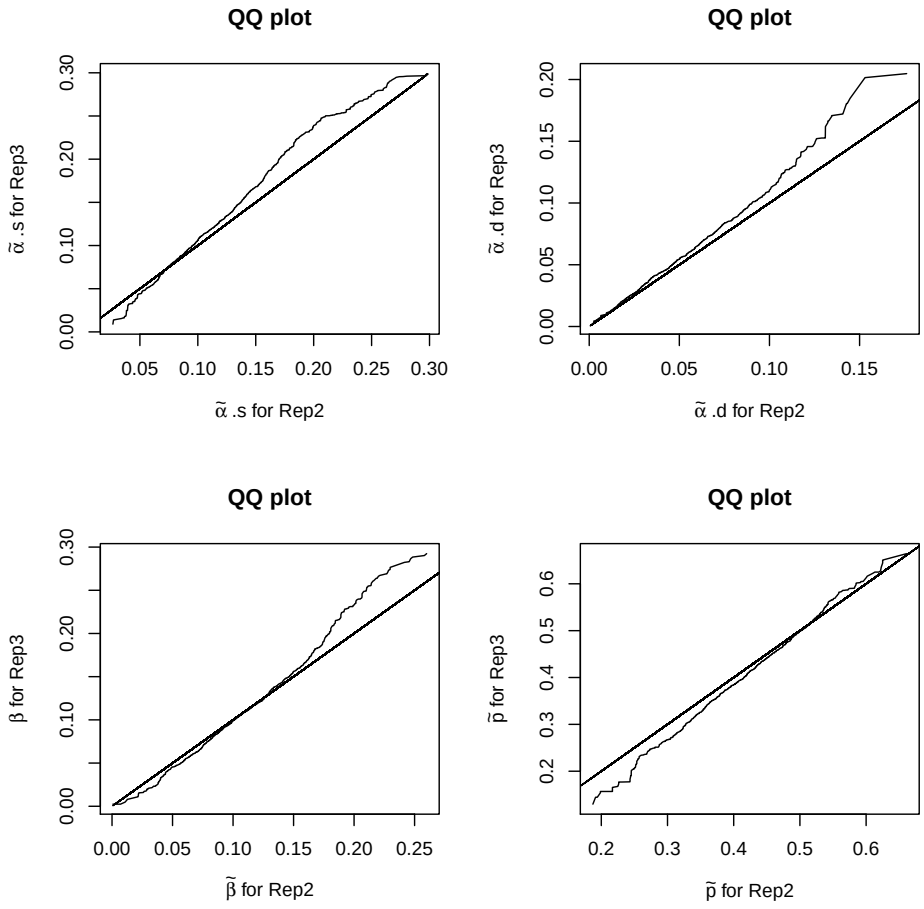

Figure 5.12: QQ plot comparing the results from Rep 2 as initial configuration and Rep 3 as initial configuration for each parameter.

Table 5.5: Initial starting values for the three combinations.

Effective Parameters	Range of Possible Values	Lower Extreme	Mid Range	Upper Extreme
$\tilde{\alpha}_s$	(0, .3)	.05	.15	.25
$\tilde{\alpha}_d$	(0, .3)	.05	.15	.25
$\tilde{\beta}$	(0, .3)	.05	.15	.25
$\tilde{p}$	(0, 1)	.1	.5	.8

5.4.4 Discussion of Sensitivity Analysis Results for Configurations

From this sensitivity analysis I found that my model can be sensitive to the chosen starting configurations, if the configuration is unrealistically extreme. However, when MCMC was started from more realistic points, the MCMC behavior was excellent.

The approach that I would take when analysing data using my model would be to use realistic initial configurations. Any of the decisions made in any match decision process seem to work well and provide similar results. I would be hesitant when assigning the extreme configurations not remotely plausible. Overall, these results are promising, especially with respect to the estimation of $\tilde{\alpha}_d$, $\tilde{\beta}$ and $\tilde{p}$.

5.5 Sensitivity Analysis for the Effective Parameters

Next, I will run a sensitivity analysis on the initial starting values for $\tilde{\alpha}_s$, $\tilde{\alpha}_d$, $\tilde{\beta}$ and $\tilde{p}$ while keeping the initial starting configuration the same as the original analysis. I focus on three combinations of initial starting values. The first will have all effective parameters near the lower extreme of their respective possible values, the second will have all near the middle of their respective possible values, and the third will have all near the higher extreme of their respective possible values. Table 5.5 provides the range of possible values, lower extreme, mid range, and upper extreme for each effective parameter. Again, like the original analysis I use 40,000 iterations with a burn-in of 20,000.

Table 5.6: Posterior distribution summary statistics for the lower extreme, upper extreme and mid range initial starting values for each effective parameter.

	Lower Extreme				**Upper Extreme**				**Mid Range**			
Effective Parameters	2.5%	50%	95%	Mean	2.5%	50%	95%	Mean	2.5%	50%	95%	Mean
$\tilde{\alpha}_s$	.046	.184	.290	.175	.049	.108	.200	.112	.048	.104	.185	.108
$\tilde{\alpha}_d$	.002	.034	.108	.040	.010	.043	.113	.048	.010	.044	.112	.049
$\tilde{\beta}$	.002	.062	.202	.072	.044	.107	.206	.113	.046	.109	.207	.114
$\tilde{p}$	.179	.345	.550	.350	.289	.423	.567	.426	.296	.427	.570	.429

5.5.1 Results Sensitivity Analysis for the Effective Parameters

I compare the three different analyses to the original results. Recall that the initial starting values for $\tilde{\alpha}_s$, $\tilde{\alpha}_d$, $\tilde{\beta}$ and $\tilde{p}$ in the original analysis were .1, .1, .25 and .5 respectively. Table 5.6 provides the posterior distribution summary statistics for each analysis (including the original). Figures 5.13, 5.14, and 5.15 are the QQ plots comparing the original analysis results to the other three analyses.

The results for the lower extreme combination vary significantly from the original results except for $\tilde{\alpha}_d$. Estimation for $\tilde{\alpha}_s$ in the lower extreme combination is higher than for the original analysis. The estimation for $\tilde{\beta}$ and $\tilde{p}$ is lower for the lower extreme combination than the original analysis. The results for the upper extreme combination and the mid range combination provided almost identical results compared to the original analysis. These results indicate that a longer chain or improved mixing is needed.

5.5.2 Discussion of Sensitivity Analysis Results for Effective Parameters

From this sensitivity analysis I found that my model can be sensitive to the chosen starting effective parameter values, especially if all initial values are in the lower extreme of their respective possible values. However, when MCMC was started from more realistic points, the MCMC behavior was excellent.

The approach that I would take when analysing data using my model would be to use initial values not at any extreme. While the upper extreme combination performed well for this small data set that might not always be the case. Overall, these results are promising, especially with respect to the estimation of $\tilde{\alpha}_d$, $\tilde{\beta}$ and $\tilde{p}$.

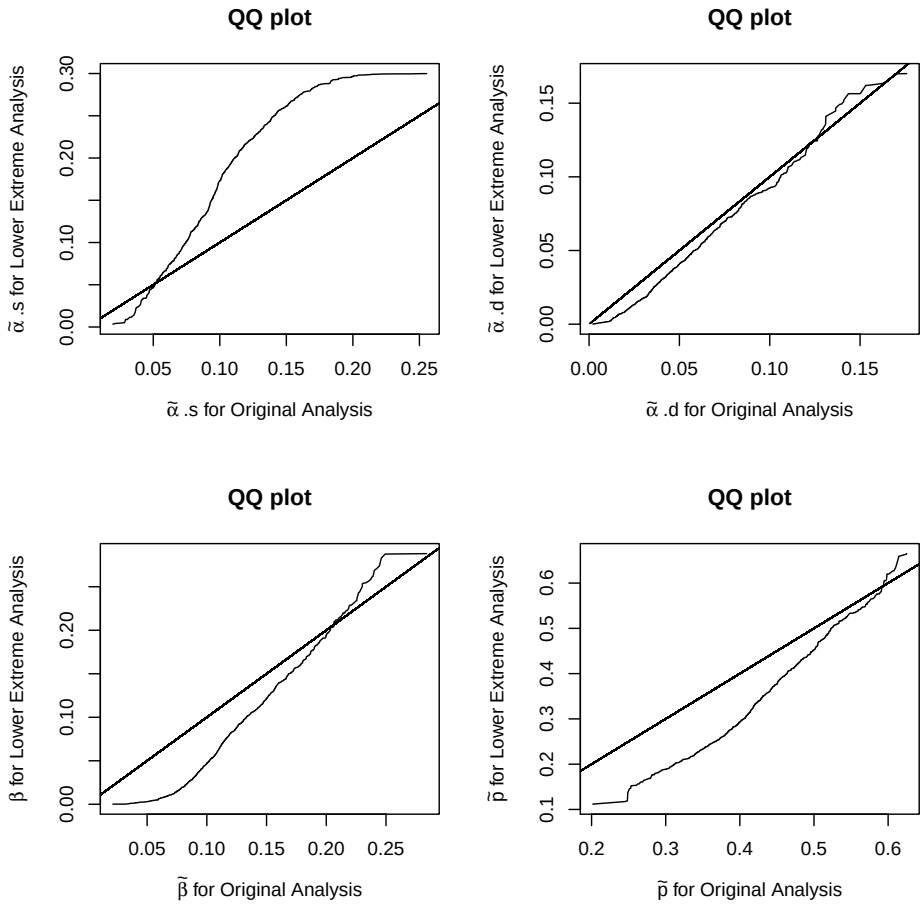

Figure 5.13: QQ plot comparing the results from the original analysis and the lower extreme analysis for each parameter.

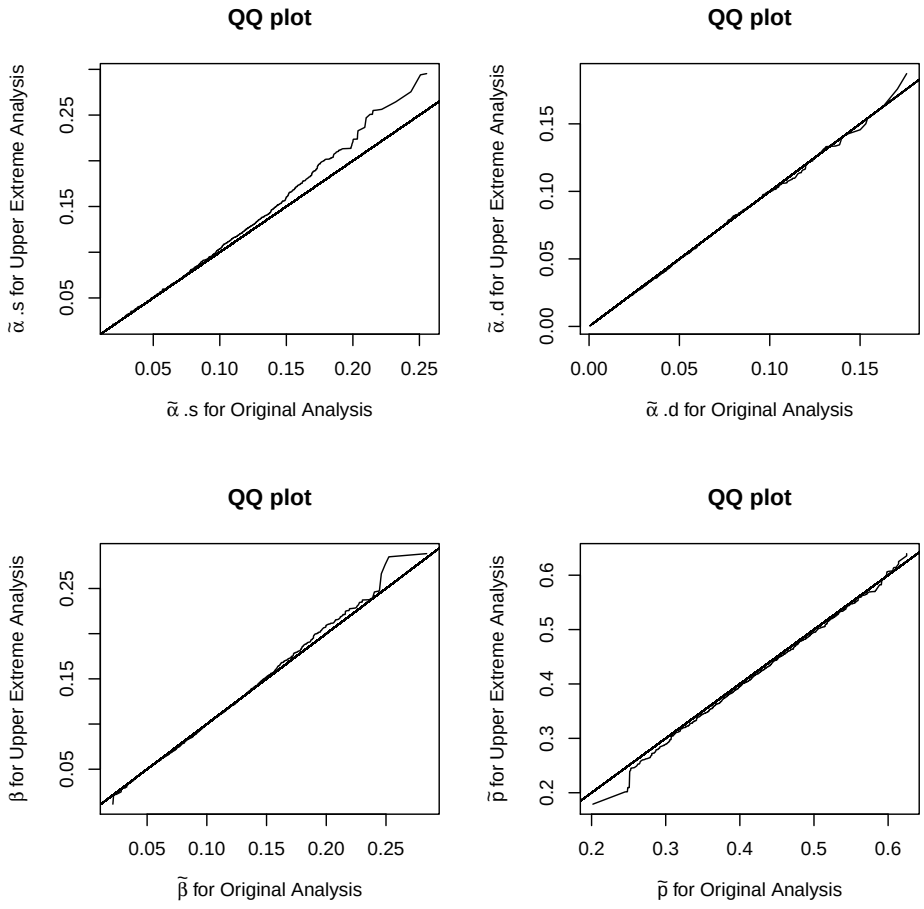

Figure 5.14: QQ plot comparing the results from the original analysis and the upper extreme analysis for each parameter.

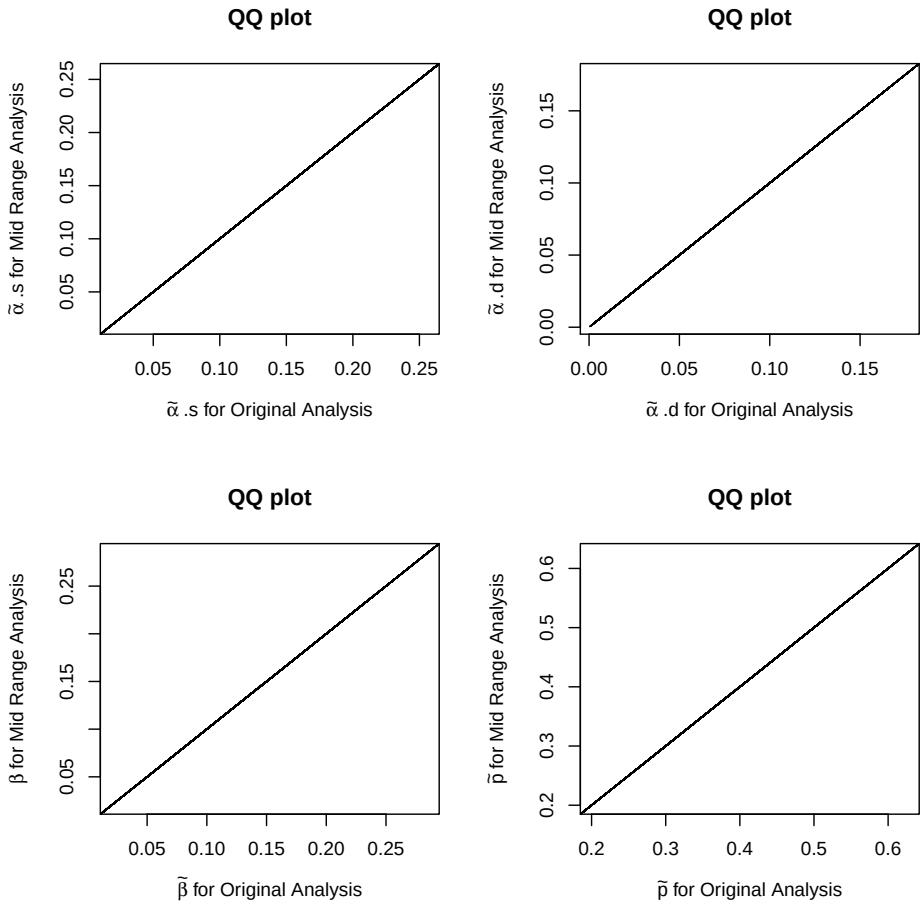

Figure 5.15: QQ plot comparing the results from the original analysis and the mid range analysis for each parameter.

CHAPTER 6

CONCLUSION

In this dissertation I provided two novel approaches to estimating detection probability in the presence of capture, recapture and animal identification errors. In this concluding chapter I will review the methods presented in the previous chapters and discuss directions for future work in this area of study.

6.1 Review

Chapter 2 described the approach for modeling capture-recapture histories, with weights to account for three sources of identification or match uncertainty; namely capture - conditional sightings, availability and match uncertainty via confidence ratings. I also explored many methods on how to handle group size inconsistencies within chains. My approaches were described with a detailed application to the whale survey data. The sample mean detection probability for my recommended analysis of the 2011 bowhead data is .501. What is most interesting about this chapter is that it presents a new way to address uncertainties that are often present but rarely acknowledged.

Chapter 3 discussed my second approach which includes the matching errors rate parameters into the likelihood. Due to the constraints imposed by the matching processes, I introduced the notation of effective parameters. My model estimates these effective parameters, which pertain to the probabilities of false positive and false negative matches. I took a Bayesian approach to parameter estimation treating the configurations of capture histories as a parameter to be estimated. An extensive simulation study to evaluate the performance of my second approach was conducted in Chapter 4. The simulations tested four different scenarios, where each of the scenarios varied based on detection probability and the true number of whales passing by in one time interval. Estimation for the scenarios with lower detection

probabilities outperformed scenarios with higher detection probabilities. These results are due to the fact that in the presence of errors a higher detection probability provides less statistical information for parameter estimation, all else being equal. The more sightings there are the more potential errors there are and thus less statistical evidence is available. Even in the face of these difficulties my model performs satisfactorily, especially considering the complexity of the problem.

Chapter 5 applied my second approach to a portion of the bowhead data. A sensitivity analysis was conducted to investigate how my model reacts to several different initial parameter values. I found that when extreme values are used my model has a difficult time mixing and converging. Potential fixes to this problem would be to increase the number of iterations and or the burn-in period, or increase the number of matching decision processes. The results with realistic starting values provided similar results to the method conducted in Chapter 2. The mean for the posterior distribution of effective detection probability is .429 with a 95% credible of (.296, .569). It is reassuring that - while each approach accounts for the recapture errors in different ways - both provide similar estimates of detection probability to the level of resolution likely attainable from such a small sample size used here.

6.2 Future Work

There is a lot of work that still can be explored to improve and advance my second approach for analysis of capture-recapture studies in the presence of capture, recapture and animal identification errors. First, an exploration of a new proposal distribution for configurations that explores more realistic configurations more often. This may improve mixing and the results when initial parameter values are extreme. Second, the model can be expanded to assume that detection probability (and even the error probabilities) depend on covariates, analogous to the (Huggins, 1989) model. We know from my weighted likelihood method that,

for example, lead condition and distance from perch affects a whales detection probability. Also, it would be very interesting to determine, for example, if the probability of a false negative match depended on whale swim speed. Third, determining how best to apply the results estimated from my second approach to provide an estimate of abundance is needed. Having surmounted the major challenge of developing and testing suitable models, I look forward in the coming years to exploring ideas like these.

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. on Automatic Control*, 19:716–723.

Amstrup, S., McDonald, T., and Manly, B. (2005). *Handbook of Capture-Recapture Analysis*. Princeton University Press.

Braham, H., Krogman, B., Leatherwood, S., Marquette, W., Rugh, D., Tillman, M., Johnson, J., and Carroll, G. (1979). Preliminary report of the 1978 spring bowhead whale research program results. *Rep. Int. Whal. Commn.*, 29:291–306.

Burnham, K. and Anderson, D. (1998). *Model Selection and Inference: a Practical Information-Theoretic Approach*. Springer-Verlag, New York, NY.

Burnham, K. and Anderson, D. (2002). *Model selection and multimodel inference: a practical information-theoretic approach*. Springer Science & Business Media.

Carothers, A. (1973). The effects of unequal catchability on Jolly-Seber estimates. *Biometrics*, 29:79–100.

Carothers, A. (1979). Quantifying unequal catchability and its effect on survival estimates in an actual population. *J. Animal Ecology*, 48:863–869.

Cormack, R. (1968). The statistics of capture-recapture methods. *Oceanogr. Mar. Biol. Ann. Rev*, 6:455–506.

Creel, S., Spong, G., Sands, J., Rotella, J., Zeigle, J., Joe, L., Murphy, K., and Smith, D. (2003). Population size estimation in yellowstone wolves with error-prone noninvasive microsatellite genotypes. *Molecular ecology*, 12(7).

Gelman, A., Carlin, J., Stern, H., and Rubin, D. (2014). *Bayesian data analysis*, volume 2. Taylor & Francis.

George, J., Givens, G., Herreman, J., DeLong, R., Tudor, B., Suydam, R., and Kendell, L. (2011). Report of the 2010 bowhead whale survey at Barrow with emphasis on methods for matching bowhead whale sightings from paired independent observations. Paper SC/63/BRG3 presented to the IWC Scientific Committee, June 2011.

George, J. C., Zeh, J., Suydam, R., and Clark, C. (2004). Abundance and population trend (1978-2001) of the western Arctic bowhead whales surveyed near Barrow, Alaska. *Marine Mammal Science*, 20:755–773.

Givens, G., Edmondson, S., George, J., Tudor, B., DeLong, R., and Suydam, R. (2014). Weighted likelihood recapture estimation of detection probabilities from an ice-based survey of bowhead whales. *Environmetrics*, 26(1):1–16.

Givens, G. and Hoeting, J. (2005). *Computational statistics*, volume 483. Wiley-Interscience.

Hammond, P. (1986). Estimating the size of naturally marked whale populations using capture-recapture techniques. *Reports of the International Whaling Commission*, 8(Special Issue):253–282.

Huggins, R. M. (1989). On the statistical analysis of capture experiments. *Biometrika*, 76(1):pp. 133–140.

Hurvich, C. and Tsai, C. (1989). Regression and time series model selection in small samples. *Biometrika*, 76:297–307.

Hwang, W.-D. and Chao, A. (1995). Quantifying the effects of unequal catchabilities on Jolly-Seber estimates via sample coverage. *Biometrics*, 51:128–141.

Koski, W., Zeh, J., Mocklin, J., Davis, A., Rugh, D., George, J., and Suydam, R. (2010). Abundance of Bering-Chukchi-Beaufort bowhead whales (*Balaena mysticetus*) in 2004 estimated from photo-identification data. *Journal of Cetacean Research and Management*, 11:89–99.

Laake, J. (2011). RMark package for R. http://www.phidot.org/software/mark/rmark.

Langtimm, C. A., Beck, C. A., Edwards, H. H., Fick-Child, K. J., Ackerman, B. B., Barton, S. L., and Hartley, W. C. (2004). Survival estimates for florida manatees from the photo-identification of individuals. *Marine Mammal Science*, 20(3):438–463.

Link, W., Yoshizaki, J., Bailey, L., and Pollock, K. (2010). Uncovering a latent multinomial: Analysis of markrecapture data with misidentification. *Biometrics*, 66(1):178–185.

Lukacs, P. and Burnham, K. (2005). Research notes: estimating population size from dna-based closed capture-recapture data incorporating genotyping error. *Journal of Wildlife Management*, 69(1):396–403.

Otis, D. L., Burnham, K. P., White, G. C., and Anderson, D. R. (1978). Statistical inference from capture data on closed animal populations. *Wildlife Monographs*. No. 62.

Paetkau, D. (2003). An empirical exploration of data quality in dna-based population inventories. *Molecular ecology*, 12(6):1375–1387.

Pledger, S. and Efford, M. (1998). Correction of bias due to heterogeneous capture probability in capture-recapture studies of open populations. *Biometrics*, 54:888–898.

Pledger, S. and Phillpot, P. (2008). Using mixtures to model heterogeneity in ecological capture-recapture studies. *Biometrical Journal*, 50:1022–1034.

Pollock, K., Hoenig, J., and Jones, C. (1991). Surveys for biological analysis. In *Creel and Angler Surveys in Fisheries Management: Proceedings of the International Symposium and Workshop on Creel and Angler Surveys in Fisheries Management, Held at Houston, Texas, USA, 26-31 March 1990*, volume 12, pages 423–434. Amer Fisheries Society.

Pollock, K., Nichols, J., Brownie, C., and Hines, J. (1990). Statistical inference for capture-recapture experiments. *Wildlife Monographs*. No. 107.

Särndal, C., Swensson, B., and Wretman, J. (2003). *Model Assisted Survey Sampling*. Springer Series in Statistics. Springer.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6:461–464.

Seber, G. (1982). *The Estimation of Animal Abundance and Related Parameters, 2nd Edition*. Griffin, London, UK.

Stevick, P., Palsbøll, P., Smith, T., Bravington, M., and Hammond, P. (2001). Errors in identification using natural markings: rates, sources, and effects on capture recapture estimates of abundance. *Canadian Journal of Fisheries and Aquatic Sciences*, 58(9):1861–1870.

Taberlet, P., Camarra, J., Griffin, S., Uhres, E., Hanotte, O., Waits, L., Dubois-Paganon, C., Burke, T., and Bouvet, J. (1997). Noninvasive genetic tracking of the endangered pyrenean brown bear population. *Molecular Ecology*, 6(9):869–876.

Taberlet, P., Waits, L., and Luikart, G. (1999). Noninvasive genetic sampling: look before you leap. *Trends in Ecology & Evolution*, 14(8):323–327.

Waits, L. (2004). Using noninvasive genetic sampling to detect and estimate abundance of rare wildlife species. *Sampling rare or elusive species: concepts, designs, and techniques for estimating population parameters. Island Press, Washington, DC, USA*, pages 211–228.

Waits, L. and Paetkau, D. (2005). Noninvasive genetic sampling tools for wildlife biologists: a review of applications and recommendations for accurate data collection. *Journal of Wildlife Management*, 69(4):1419–1433.

White, G. (1982). *Capture-recapture and removal methods for sampling closed populations*. Los Alamos National Laboratory.

White, G. and Burnham, K. (1999). Program MARK: survival estimation from populations of marked animals. *Bird Study*, 46, Suppl.:120–138.

Williams, B., Nichols, J., and Conroy, M. (2002). *Analysis and management of animal populations*. Academic Press.

Wright, J., Barker, R., Schofield, M., Frantz, A., Byrom, A., and Gleeson, D. (2009). Incorporating genotype uncertainty into mark–recapture-type models for estimating abundance using dna samples. *Biometrics*, 65(3):833–840.

Yoshizaki, J. (2007). Use of natural tags in closed population capture-recapture studies: modeling misidentification.

Zeh, J., Clark, C., George, J., Withrow, D., Carroll, G., , and Koski, W. (1993). Current population size and dynamics. In Burns, J., Montague, J., and Cowles, C., editors, *The Bowhead Whale*, pages 409–489. Special Publication No. 2 of the Society for Marine Mammaology, Lawrence, KS.

Zeh, J., George, J., Raftery, A., and Carroll, G. (1991). Rate of increase, 1978-1988, of bowhead whales, *Balaena mysticetus*, estimated from ice-based census data. *Marine Mammal Science*, 7:105–122.

Zeh, J., Ko, D., Krogman, B., and Sonntag, R. (1986). A multinomial model for estimating the size of a whale population from incomplete census data. *Biometrics*, 42:1–14.

Zeh, J. and Punt, A. (2005). Updated 1978-2001 abundance estimates and their correlations for the Bering-Chukchi-Beaufort Seas stock of bowhead whales. *Journal of Cetacean Research and Management*, 7:169–175.