ABSTRACT

WEN, ZHI. Single State Super Population Capture-Recapture Model Augmented with Information on Population of Origin. (Under the direction of Dr. Kenneth Pollock and Dr. James Nichols).

Ecologists applying capture-recapture models to animal populations sometimes have access to additional information about individuals' populations of origin. For example, tests that assign an individual's genotype to its most likely source population are increasingly used. Here we show how to augment a super population capture-recapture model with such information. We consider a single super population model without age structure, and split the entry probability into separate components due to births in situ and immigration. We show that it is possible to estimate these two probabilities separately. We first consider the case of perfect information about population of origin, where we can distinguish individuals born in situ from immigrants with certainty. Then we consider the more realistic case of imperfect information, where we use genetic or other information to assign probabilities to each individual's origin in situ or outside the population. We use a resampling approach to impute the perfect origination assignment data based on the imperfect assignment tests. The integration of data on population of origin with capture-recapture data allows us to determine the contributions of immigration and in situ reproduction to the growth of the population, an issue of importance to ecologists. Further, the augmentation of capture-recapture data with origination data should improve the precision of parameter estimates. We illustrate our new models with capture-recapture and genetic assignment test data from a population of banner-tailed kangaroo rats *Dipodomys spectabilis* in Arizona.

In chapter 4, we evaluate the value of marine reserves for fisheries using tag-return, tag-recapture and telemetry models. We estimate the patch-specific fishing mortality, natural mortality, and movement rates. We first focus on tag-return models for a two-site model with one area a marine reserve and one area a fishing area. We consider tag-return, tag-recapture and telemetry models in various combinations for two site models where one area is a marine reserve and one is subject to regular fishing. Then we illustrate our methods with a comprehensive simulation study.
Single State Super Population Capture-Recapture Model Augmented with Information on Population of Origin

by

Zhi Wen

A dissertation submitted to the Graduate Faculty of North Carolina State University in partial fulfillment of the requirements for the Degree of Doctor of Philosophy

Statistics and Operations Research

Raleigh, North Carolina

2009

APPROVED BY:

__________________________  ____________________________
Dr. James Nichols           Dr. Thomas Reiland

__________________________  ____________________________
Dr. Sujit Ghosh             Dr. Charles E. Smith
Co-Chair of Advisory Committee

__________________________  ____________________________
Dr. Kenneth Pollock         Chair of Advisory Committee
DEDICATION

TO MY HUSBAND, MY DAUGHTER, MY FATHER, AND MY SIBLINGS.
Zhi Wen was born on June 22, 1979 in Jiangxi, P.R.China. She received her Bachelors degree in Computational Mathematics from Nanchang University of China in July 2000. In April of 2004, she got her Masters degree in Operations Research from Southeast University of China. Then she enrolled at the Department of Applied Mathematics in Case Western Reserve University. After one year of study there she started her Ph.D. study at North Carolina State University in 2005, co-majoring in Statistics and Operations Research. Since June 2006, she has been working on a research assistantship on incorporating additional genetic information on recruitment sources into open capture-recapture models with Kenneth H. Pollock and James D. Nichols.
ACKNOWLEDGMENTS

First of all, I thank my advisor Kenneth Pollock, for his continuous support in the Ph.D. program. He was always there to listen and taught me how to ask questions and express my ideas. He showed me different ways to approach a research problem and the need to be persistent. He has been a friend and mentor. He taught me how to write academic papers, made me a better writer, had confidence in me when I doubted myself, and brought out the good ideas in me. He secured grants that financially supported me for three years which enabled me to focus on research without worrying my financial situation. Without his guidance and encouragement, I could not have finished this dissertation.

A special thank goes to James Nichols, who as my co-advisor guided my research. He contributed a lot to the writing of this dissertation as well as the challenging research that lies behind it. His insightful suggestions improved the quality of my dissertation and he showed patience in explaining complicated research problems to me simply. I would also like to express my gratitude to Peter Waser for providing us the Kangaroo-rat data set in Southern Arizona. He taught me about genetic assignment tests and how to use GENECLASS2 to analyze genetic data. I also thank Jim and Peter for securing the NSF funding that supported the research in Chapter 1-3. Also thanks to John Hoenig of Virginia Institute of Marine Science who provided funding for the research on fisheries tag-return and telemetry models in Chapter 4.

I would like to express my thanks to Sujit Ghosh, from whom I have learned using Bayesian approaches and WinBUGS to analyze statistical problems. His classes opened a window for me to the wonderful Bayesian world. I would like to thank the rest of my thesis committee. Charles Smith, my operations research co-chair who asked good questions and inspired me with new ideas, and Thomas Reiland, who taught me optimization algorithms and these algorithms improved the efficiency of my procedures. I am also greatly indebted to all the teachers having taught me in North Carolina State University. They were always there for questions and were patient in clearing my confusions.

Last, but not least, I thank my family: my husband, Weihua Cao, my daughter, Amy Cao, my sister, Qing Wen, and my brother, Ming Wen. They were my inspiration to work hard and contribute to the development of society.
# TABLE OF CONTENTS

**LIST OF TABLES** ........................................................................................................ vii

**LIST OF FIGURES** ....................................................................................................... x

**1 A Single State Super Population Capture-Recapture Model Augmented with Information on Population of Origin** ........................................ 1

1.1 Introduction .................................................................................................................. 2

1.2 Notation and Assumptions .......................................................................................... 3

1.3 Model Structure .......................................................................................................... 5

1.3.1 Standard Super Population Capture-Recapture Model with Robust Design ........ 5

1.3.2 Super Population Capture-Recapture Model Augmented with Perfect Assignment Information ................................................................. 7

1.3.3 Super Population Capture-Recapture Model Integrated with Imperfect Assignment Information ............................................................. 8

1.3.4 Assumptions .......................................................................................................... 9

1.4 Simulation .................................................................................................................... 10

1.5 Example: Kangaroo Rats in Arizona ........................................................................... 16

1.6 Discussion and Future Research ................................................................................. 20

**Bibliography** .............................................................................................................. 27

**2 A Super Population Capture-Recapture Model for Single Population with Two Age Groups Augmented with Age Information or Genetic Data** ................................................................. 29

2.1 Introduction ................................................................................................................ 30

2.2 Model Structure ........................................................................................................ 32

2.2.1 Notation ............................................................................................................... 32

2.2.2 Standard Super Population Capture-Recapture Model for Single Population with Two Age Classes ................................................................. 33

2.2.3 Super Population Capture-Recapture Model for a Single Population with Two Age Groups Augmented with Perfect Genetic Assignment Test ............................................................................................................... 36

2.2.4 Super Population Capture-Recapture Model for a Single Population with Two Age Groups Augmented with Imperfect Genetic Assignment Test ........................................................................................................... 39

2.2.5 Assumptions ........................................................................................................ 40

2.3 Simulation ................................................................................................................... 41

2.4 Example: Kangaroo Rats in Southern Arizona ............................................................. 47

2.5 Discussion ................................................................................................................... 48
### 3 Multi-State Super Population Capture-Recapture Model Integrated with Origination Assignment Data

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Introduction</td>
<td>56</td>
</tr>
<tr>
<td>3.2 Model Structure</td>
<td>58</td>
</tr>
<tr>
<td>3.2.1 Notation</td>
<td>58</td>
</tr>
<tr>
<td>3.2.2 Standard Multi-State Super Population Capture-Recapture Model</td>
<td>61</td>
</tr>
<tr>
<td>3.2.3 Multi-State Super Population Model with Robust Design Augmented with Perfect Origination Assignment Information</td>
<td>63</td>
</tr>
<tr>
<td>3.2.4 Multi-State Super Population Capture-Recapture Model Augmented with Imperfect Origination Assignment Information</td>
<td>66</td>
</tr>
<tr>
<td>3.3 Assumption</td>
<td>67</td>
</tr>
<tr>
<td>3.4 Simulations</td>
<td>68</td>
</tr>
<tr>
<td>3.5 Example: Kangaroo Rats in Southern Arizona</td>
<td>72</td>
</tr>
<tr>
<td>3.6 Discussion and Future Research</td>
<td>80</td>
</tr>
</tbody>
</table>

### Bibliography

53

### 4 Evaluation of Marine Reserves for Fisheries using Tag-Return, Tag-Recapture and Telemetry Models: Estimation of Patch-Specific Fishing Mortality, Natural Mortality and Movement Rates

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 Introduction</td>
<td>88</td>
</tr>
<tr>
<td>4.2 Notation</td>
<td>90</td>
</tr>
<tr>
<td>4.3 Model Development</td>
<td>92</td>
</tr>
<tr>
<td>4.3.1 Model 1</td>
<td>92</td>
</tr>
<tr>
<td>4.3.2 Model 2</td>
<td>93</td>
</tr>
<tr>
<td>4.3.3 Model 3 (Telemetry Model)</td>
<td>96</td>
</tr>
<tr>
<td>4.3.4 Model 4 (Combines Regular Tags and Telemetry Tags)</td>
<td>101</td>
</tr>
<tr>
<td>4.3.5 Assumptions</td>
<td>102</td>
</tr>
<tr>
<td>4.4 Evaluation of Likelihood Methods Using Simulation</td>
<td>103</td>
</tr>
<tr>
<td>4.5 Discussion and Future Research</td>
<td>107</td>
</tr>
</tbody>
</table>

### Bibliography

85

111
LIST OF TABLES

Table 1.1 Comparing simulation results of the super population capture-recapture model with imperfect assignment information ("Imperfect" columns) and the standard super population capture-recapture model ("Standard" columns). The uncertainty level of the "imperfect" assignment procedure is 0.95, that is, \( \pi_i \in (0.95, 1) \), \( i = 1, \ldots, \sum_{i=1}^{K} (u_i^B + u_i^I) \). The super population size is \( N = 1000 \). 23

Table 1.2 Estimates of entry probabilities from the kangaroo rat data set under models with a “perfect” (\( M_1 \)) population assignment procedure, an “imperfect” population assignment procedure (\( M_2 \)), and without population assignment information (\( M_3 \)). 24

Table 1.3 Estimates of survival probabilities from the kangaroo rat data set under models incorporating “perfect” assignment information (“Perfect model” columns), imperfect assignment information (“Imperfect model” columns), and without assignment information (“Standard” columns). 25

Table 1.4 Kangaroo rat data set probability models, ranked from best (lowest \( QAIC_c \) score) to worst. \( p(c, c) \) represents the model where capture probability is constant over time and is the same for \textit{in situ} birth and immigrant. The first letter of parenthesis indicates whether the parameter is dependent on time (“t”) or is constant over time (“c”). The second letter of parenthesis indicates whether the parameter is different for \textit{in situ} birth and immigrant (“o”) or is the same for \textit{in situ} birth and immigrant (“c”). \( \beta t, o \) means the model whose entry probability is dependent on time and is the same for \textit{in situ} birth and immigrant. \( \phi(c, c) \) means the model whose survival probability is constant over time but different for \textit{in situ} birth. 26

Table 1.5 Estimate of Kangaroo rat data under the super population capture-recapture model added with imperfect genetic data (\( p(c, c)\beta(t, o)\varphi(c, o) \)), with capture probability be the same for all animals, entry probability is dependent on time and origination of animal, and survival probability is the same over time but different for \textit{in situ} birth and immigrant. 26

Table 2.1 Compare the simulation results of the super population capture-recapture model with perfect genetic assignment data (“P M” columns) and the standard
super population capture-recapture model ("S M" columns). The simulation results are under $\beta_{0B}^Y = 0.06$, $\beta_{0B}^O = 0.2$, $\beta_{0I}^I = 0.1$, $\beta_{1B}^Y = 0.05$, $\beta_{1I}^I = 0.11$ and other parameter values are listed in the table.

Table 2.2 Estimate the entry probability of Kangaroo rat data set under three different super population capture-recapture models, model with “perfect” genetic assignment test (“Perfect model” columns), model with imperfect genetic assignment test (“Imperfect model” columns), and model without genetic assignment test (“Standard model”).

Table 2.3 Estimate the survival probability of Kangaroo rat data set under three different super population capture-recapture models, model with “perfect” genetic assignment test (“Perfect model” columns), model with imperfect genetic assignment test (“Imperfect model” columns), and model without genetic assignment test (“Standard model”).

Table 3.1 Relative RMSE (%) of estimates of entry probabilities under multi-state super population capture-recapture model (standard model) and multi-state super population capture-recapture model augmented with perfect origination assignment information. The results are based on 500 simulations. “P” column corresponds to model with perfect origination information, “S” column corresponds standard model, and “Val” columns is the true parameter value.

Table 3.2 Relative RMSE (%) of estimates of entry probabilities under multi-state super population capture-recapture model (standard model) and multi-state super population capture-recapture model augmented with perfect origination assignment information. The results are based on 500 simulations.

Table 4.1 Symbolic representation for model 1 with 3 tagging periods and two fishing seasons.

Table 4.2 Simulation Result for parameters take value in column 2, comparison of the Relative Root Mean Square Error (%) of estimates for model 1 to model 4b. The first model 3a (4a) assumes $M_A \neq M_B$, while the second model 3a (4a) assumes $M_A = M_B$.

Table 4.3 Simulation Result for parameters take value in column 2, comparison of the Relative Root Mean Square Error (%) of estimates for model 1 to model 4b.
first model 3a (4a) assumes $M^A \neq M^B$, while the second model 3a (4a) assumes $M^A = M^B$.

Table 4.4 Simulation Result for parameters take value in column 2, comparison of the Relative Root Mean Square Error (%) of estimates for model 1 to model 4b. The first model 3a (4a) assumes $M^A \neq M^B$, while the second model 3a (4a) assumes $M^A = M^B$.

Table 4.5 Simulation Result for parameters take value in column 2, comparison of the Relative Root Mean Square Error (%) of estimates for model 1 to model 4b. The first model 3a (4a) assumes $M^A \neq M^B$, while the second model 3a (4a) assumes $M^A = M^B$.

Table 4.6 Simulation Result for parameters take value in column 2, comparison of the Relative Root Mean Square Error (%) of estimates for model 1 to model 4b. The first model 3a (4a) assumes $M^A \neq M^B$, while the second model 3a (4a) assumes $M^A = M^B$.

Table 4.7 Simulation Result for parameters take value in column 2, comparison of the Relative Root Mean Square Error (%) of estimates for model 1 to model 4b. The first model 3a (4a) assumes $M^A \neq M^B$, while the second model 3a (4a) assumes $M^A = M^B$.

Table 4.8 Simulation Result for parameters take value in column 2, comparison of the Relative Root Mean Square Error (%) of estimates for model 1 to model 4b. The first model 3a (4a) assumes $M^A \neq M^B$, while the second model 3a (4a) assumes $M^A = M^B$.

Table 4.9 Simulation Result for parameters take value in column 2, comparison of the Relative Root Mean Square Error (%) of estimates for model 1 to model 4b. The first model 3a (4a) assumes $M^A \neq M^B$, while the second model 3a (4a) assumes $M^A = M^B$. 


LIST OF FIGURES

Figure 1.1  Relative bias of 1000 simulations for the super population capture-recapture model augmented with different assignment procedures. $p_{ij} = 0.5, \varphi_i = 0.4, \beta^B_0 = 0.2, \beta^I_i = 0.12, \beta^I_i = 0.06, N = 1000$ ................................. 13

Figure 1.2  Relative sample standard error of 1000 simulations for the super population capture-recapture model augmented with different assignment procedures. $p_{ij} = 0.5, \varphi_i = 0.4, \beta^B_0 = 0.2, \beta^B_i = 0.12, \beta^I_0 = 0.1, \beta^I_i = 0.06, N = 1000$ ................................. 14

Figure 1.3  Relative RMSE of 1000 simulations for the super population capture-recapture model integrated with different assignment procedures. $p_{ij} = 0.5, \varphi_i = 0.4, \beta^B_0 = 0.2, \beta^B_i = 0.12, \beta^I_0 = 0.1, \beta^I_i = 0.06, N = 1000$ ................................. 15

Figure 1.4  Relative RMSE for the simulation results of the super population capture-recapture model integrated with four different levels of uncertainty of assignment information ($\pi = 1, \pi \in (0.95, 1), \pi \in (0.9, 1)$ and $\pi \in (0.8, 1)$). Parameters take a complete combinations of $p_{ij} = 0.5, 0.2, \varphi_i = 0.4, 0.15, N = 1000, 10000, \beta^B_0 = 0.4, \beta^B_i = 0.1, \beta^B_i = 0.12, \beta^I_i = 0.06, i = 1, \cdots, K - 1$ ................................. 16

Figure 1.5  Compare the entry probability and survival probability of Kangaroo rat data set under two different models, super population capture-recapture model with perfect genetic data (Perfect genetic model) and standard super population model capture-recapture model without genetic data (Standard model).................. 18

Figure 1.6  Comparison of capture probabilities and the 95% confidence interval of its estimates from the kangaroo rat data set under three different models, super population capture-recapture model with perfect assignment information (Perfect genetic model), super population capture-recapture model with imperfect assignment information (Imperfect genetic model) and standard super population model capture-recapture model without assignment information (standard model)........ 19

Figure 2.1  Relative bias of 1000 simulations for the super population capture-recapture model integrated with different genetic assignment tests. $p^Y_{ij} = 0.3, p^O_{ij} = 0.27, \varphi^Y_i = 0.8, \varphi^O_i = 0.8, \beta^{YB}_0 = 0.2, \beta^{YB}_i = 0.11, \beta^{OB}_0 = 0.2, \beta^{OB}_i = 0.1, \beta^{OI}_0 = 0.05, N = 1000$ ................................. 43

Figure 2.2  Relative sample standard error of 1000 simulations for the super population capture-recapture model integrated with different genetic assignment tests. $p^Y_{ij} = 0.3, p^O_{ij} = 0.27, \varphi^Y_i = 0.8, \varphi^O_i = 0.8, \beta^{YB}_0 = 0.2, \beta^{YB}_i = 0.11, \beta^{OB}_0 = 0.2, \beta^{OB}_i = 0.1, \beta^{OI}_0 = 0.05, N = 1000$ ................................. 44
Figure 2.3  Relative root of mean square error (RMSE) of the estimates of the transition parameter under five different models. From left to right, the x-axis indicates the parameter values of $\pi = 1, \pi \in (0.95, 1), \pi \in (0.9, 1)$ and $\pi \in (0.8, 1)$. Parameters take a complete combinations of $p_{ij}^Y = 0.5, p_{ij}^O = 0.2, \varphi_i^Y = 0.4, \varphi_i^{Yhh} = 0.05, \varphi_i^{Ohh} = 0.45, \varphi_i^{Yhq} = 0.07$, the upper right picture corresponds to parameter values of $p_{ij}^Y = 0.5, p_{ij}^O = 0.2, \varphi_i^{Yhq} = 0.03, \varphi_i^{Ohh} = 0.3, \varphi_i^{Yhh} = 0.05$, the lower left picture corresponds to parameter values of $p_{ij}^Y = 0.2, p_{ij}^O = 0.5, \varphi_i^{Yhq} = 0.05, \varphi_i^{Ohh} = 0.45, \varphi_i^{Yhh} = 0.07$, and the lower right corresponds to parameter values of $p_{ij}^Y = 0.2, p_{ij}^O = 0.5, \varphi_i^{Yhq} = 0.03, \varphi_i^{Ohh} = 0.3, \varphi_i^{Yhh} = 0.05$. Each line of the pictures represents the relative RMSE of a parameter under five different models. From left to right, the x-axis indicates the uncertainty levels are (0.8, 1), (0.9, 1), (0.95, 1), 1. The “standard model” corresponds to the multi-state super population capture-recapture model without origination information. .......................... 46

Figure 2.4  Relative RMSE for the simulation results of the super population capture-recapture model integrated with four different uncertainty levels of genetic assignment tests ( $\pi = 1, \pi \in (0.95, 1), \pi \in (0.9, 1)$ and $\pi \in (0.8, 1)$). Parameters take a complete combinations of $p_{ij}^Y = 0.5, p_{ij}^O = 0.2, \varphi_i^Y = 0.3, \varphi_i^O = 0.35, \beta_i^Y = 0.2, \beta_i^O = 0.2, \beta_i^O = 0.2, \beta_i^O = 0.1, \beta_i^O = 0.05$. .......................... 45

Figure 2.5 Compare the entry probability and survival probability of Kangaroo rat data set under two different models, super population capture-recapture model with imperfect genetic data (Imperfect model)and standard super population model capture-recapture model without genetic data (Standard model) .......................... 48
models are multi-state super population capture-recapture with imperfect origination assignment information and the uncertain levels are (0.8, 1), (0.9, 1), (0.95, 1), 1. The “standard model” corresponds to the multi-state super population capture-recapture model without origination information.

Figure 3.3 Relative root of mean square error (RMSE) of the estimates of total 76 parameters based on 1000 simulations. The upper left picture corresponds to parameter values of $p_{ij}^Y = 0.5, p_{ij}^O = 0.2, \varphi_i^{Yhh} = 0.4, \varphi_i^{Yhq} = 0.05, \varphi_i^{Ohh} = 0.45, \varphi_i^{Yhq} = 0.07$, the upper right picture corresponds to parameter values of $p_{ij}^Y = 0.2, \varphi_i^{Yhh} = 0.25, \varphi_i^{Yhq} = 0.03, \varphi_i^{Ohh} = 0.3, \varphi_i^{Yhq} = 0.05$, the lower left picture corresponds to parameter values of $p_{ij}^Y = 0.2, p_{ij}^O = 0.5, \varphi_i^{Yhh} = 0.4, \varphi_i^{Yhq} = 0.05, \varphi_i^{Ohh} = 0.45, \varphi_i^{Yhq} = 0.07$, and the lower right corresponds to parameter values of $p_{ij}^Y = 0.2, p_{ij}^O = 0.5, \varphi_i^{Yhh} = 0.25, \varphi_i^{Yhq} = 0.03, \varphi_i^{Ohh} = 0.3, \varphi_i^{Yhq} = 0.05$. Each line of the pictures represents the relative RMSE of a parameter under four different models. From left to right, the x-axis indicates the models are multi-state super population capture-recapture with imperfect origination assignment information and the uncertain levels are (0.8, 1), (0.9, 1), (0.95, 1), 1.

Figure 3.4 Relative bias of the estimates of total 76 parameters based on 1000 simulations. The upper left picture corresponds to parameter values of $p_{ij}^Y = 0.5, p_{ij}^O = 0.2, \varphi_i^{Yhh} = 0.4, \varphi_i^{Yhq} = 0.05, \varphi_i^{Ohh} = 0.45, \varphi_i^{Yhq} = 0.07$, the upper right picture corresponds to parameter values of $p_{ij}^Y = 0.5, p_{ij}^O = 0.2, \varphi_i^{Yhh} = 0.25, \varphi_i^{Yhq} = 0.03, \varphi_i^{Ohh} = 0.3, \varphi_i^{Yhq} = 0.05$, the lower left picture corresponds to parameter values of $p_{ij}^Y = 0.2, p_{ij}^O = 0.5, \varphi_i^{Yhh} = 0.4, \varphi_i^{Yhq} = 0.05, \varphi_i^{Ohh} = 0.45, \varphi_i^{Yhq} = 0.07$, and the lower right corresponds to parameter values of $p_{ij}^Y = 0.2, p_{ij}^O = 0.5, \varphi_i^{Yhh} = 0.25, \varphi_i^{Yhq} = 0.03, \varphi_i^{Ohh} = 0.3, \varphi_i^{Yhq} = 0.05$. Each line of the pictures represents the relative bias of a parameter under four different models. From left to right, the x-axis indicates the models are multi-state super population capture-recapture with imperfect origination assignment information and the uncertain levels are (0.8, 1), (0.9, 1), (0.95, 1), 1.

Figure 3.5 Relative standard error (S.E.) of the estimates of total 76 parameters based on 1000 simulations. The upper left picture corresponds to parameter values of $p_{ij}^Y = 0.5, p_{ij}^O = 0.2, \varphi_i^{Yhh} = 0.4, \varphi_i^{Yhq} = 0.05, \varphi_i^{Ohh} = 0.45, \varphi_i^{Yhq} = 0.07$, the upper right picture corresponds to parameter values of $p_{ij}^Y = 0.5, p_{ij}^O = 0.2, \varphi_i^{Yhh} = 0.25, \varphi_i^{Yhq} = 0.03, \varphi_i^{Ohh} = 0.3, \varphi_i^{Yhq} = 0.05$, the lower left picture corresponds to parameter values of $p_{ij}^Y = 0.2, p_{ij}^O = 0.5, \varphi_i^{Yhh} = 0.4, \varphi_i^{Yhq} = 0.05, \varphi_i^{Ohh} = 0.45, \varphi_i^{Yhq} = 0.07$, and the lower right corresponds to parameter values of $p_{ij}^Y = 0.2, p_{ij}^O = 0.5, \varphi_i^{Yhh} = 0.25, \varphi_i^{Yhq} = 0.03, \varphi_i^{Ohh} = 0.3, \varphi_i^{Yhq} = 0.05$. Each line of the pictures represents the relative S.E. of a parameter under four different models. From left to right, the x-axis indicates the models are multi-state super
population capture-recapture with imperfect origination assignment information and the uncertain levels are $(0.8, 1), (0.9, 1), (0.95, 1)$.

Figure 3.6 Compare capture probabilities and the 95% confidence interval of its estimates of the Kangaroo rat data set under three different models, multi-state super population capture-recapture models with perfect origination assignment information (perfect genetic), imperfect genetic assignment test (imperfect genetic), and without genetic assignment test (standard model).

Figure 3.7 Compare the transition probability of Kangaroo rat data set under three different models, multi-state super population capture-recapture models with perfect origination assignment information (perfect genetic), imperfect genetic assignment test (imperfect genetic), and without genetic assignment test (standard model). For the upper picture, the upper three lines are estimates of $\varphi_{i}^{Y22}$, the middle three lines are estimates of $\varphi_{i}^{Y11}$, and the bottom three lines for estimates of $\varphi_{i}^{Y12}$ and $\varphi_{i}^{Y21}$. For the bottom picture, the upper three lines are estimates of $\varphi_{i}^{O11}$, the middle three lines are estimates of $\varphi_{i}^{O22}$, and the bottom three lines for estimates of $\varphi_{i}^{O12}$ and $\varphi_{i}^{O21}$.

Figure 3.8 Compare the entry probability of Kangaroo rat data set under three different models, multi-state super population capture-recapture models with perfect origination assignment information (perfect genetic), imperfect genetic assignment test (imperfect genetic), and without genetic assignment test (standard model). For the upper two pictures, the upper three lines are estimates of $\beta_{i}^{Y1B}$, the middle three lines are estimates of $\varphi_{i}^{O1I}$. For the bottom picture, every three points in the same vertical line are the standard error of the estimates of entry probability under three different models. We define $\beta_{0h}^{Oh \text{ genetic model}} = \sum_{q=1,2} \beta_{0hBq}^{Oh \text{ genetic model}} + \beta_{0hI}^{Oh \text{ genetic model}}, h = 1, 2$.

Figure 4.1 The depiction of the fishing season and tagging period of area A and area B.
Chapter 1

A Single State Super Population Capture-Recapture Model Augmented with Information on Population of Origin

**summary** Ecologists applying capture-recapture models to animal populations sometimes have access to additional information about individuals’ populations of origin. For example, tests that assign an individual’s genotype to its most likely source population are increasingly used. Here we show how to augment a super population capture-recapture model with such information. We consider a single super population model without age structure, and split the entry probability into separate components due to births *in situ* and immigration. We show that it is possible to estimate these two probabilities separately. We first consider the case of perfect information about population of origin, where we can distinguish individuals born *in situ* from immigrants with certainty. Then we consider the more realistic case of imperfect information, where we use genetic or other information to assign probabilities to each individual’s origin *in situ* or outside the population. We use a re-sampling approach to impute the perfect origination assignment data based on the imperfect assignment tests. The integration of data on population of origin with capture-recapture data allows us to determine the contributions of immigration and *in situ* reproduction to the
growth of the population, an issue of importance to ecologists. Further, the augmentation of capture-recapture data with origination data should improve the precision of parameter estimates. We illustrate our new models with capture-recapture and genetic assignment test data from a population of banner-tailed kangaroo rats \textit{Dipodomys spectabilis} in Arizona.

1.1 Introduction

The Jolly-Seber (Jolly 1965, Seber, 1982) method is a very popular capture-recapture method for open animal populations. In its original version, animals are assumed to belong to a single population subject to birth, death, emigration and immigration. The super-population model is one modern representation of the Jolly-Seber model. It was first proposed by Crosbie and Manly (1985) and then developed further by Schwarz and Arnason (1996). Here we consider a single state population without age structure where animals may be recruited either by births or by immigration from outside the population. The sampling study is composed of \( K \) primary periods and each primary period has \( l_i (i = 1, 2, \ldots, K) \) shorter secondary periods. The super-population model treats all the individuals exposed to sampling at any time of the study as a part of a super population \((N)\). That is, the super population is composed of the recruits during all sampling periods \((i = 1, 2, \ldots, K)\) with \( N = \sum_{i=1}^K B_i \), where \( B_i \) is the number of new recruits that enter the population between sampling period \( i \) and \( i+1 \), \( i = 1, 2, \ldots, K-1 \). Note that \( B_0 = N_1 \) are all the animals in the population when the initial sampling takes place. The animals of the super population are assumed to enter the super population at one sampling period \((i)\) with entry probability \((\beta_i, \beta_i \in (0, 1), \sum_{i=0}^K \beta_i = 1)\).

Throughout this paper, all our capture-recapture models use the Robust Design (first proposed by Pollock, 1982), an important sampling scheme which combines the features of both closed and open population studies. In the Robust Design, the investigator samples the population at two different time scales, with \( l_i \) (where \( l_i \geq 2, i = 1, 2, \ldots, K \)) shorter secondary sampling periods within the longer primary sampling periods \( i \) \((i = 1, 2, \ldots, K)\). The population is assumed to be closed within each primary period (Kendall and Nichols, 1995, Kendall \textit{et al.}, 1997) and open between primary periods. That is, there can be losses (deaths, emigration) or gains (births, immigration) between primary periods. With the Robust Design, we can estimate some parameters (such as the first and last period
capture probabilities and population sizes) which can not be estimated under the general

Estimating the relative contribution of \textit{in situ} birth and immigration to the growth
of population is of great interest to ecologists. However, in the standard super population
model (Schwarz and Arnason, 1996), it is not possible to separate immigrants from animals
born on site. Here we propose a new approach to combine information on animals’ pop-
ulations of origin with the standard super population capture-recapture model. Methods
to infer the site of origin of a captured animal have been developing rapidly; they include
genetic assignment tests, genetic parentage analysis, and feather isotope analysis (we should
cite some papers here). These methods allow the investigator to “assign” an individual to
a probable population of origin, hence we refer to them as “assignment procedures”.

We shall begin with the assumption that the assignment procedures provide perfect
information (i.e. without uncertainty) and later generalize to a model for “imperfect”
assignment procedures (i.e., the uncertainty in the assignment is accounted for). A “perfect”
assignment procedure means that each animal captured can be assigned without error either
as an animal born in the population or an immigrant from outside the population. That
is, a “birth state” (immigrant or \textit{in situ} birth) is assigned to any captured animal where
this state is otherwise unknown. In reality, most assignment procedures will be imperfect,
and the assignment of an individual to a particular population of origin will be subject to
uncertainty. In our new approach, we will split the entry probability into two components:
an on site recruitment entry probability and an immigrant entry probability. We will
show that it is possible to estimate these two entry probabilities separately. The fusion of
origination information and capture-recapture data will allow us to determine the separate
contributions of immigration and \textit{in situ} reproduction to the growth of the population.

1.2 Notation and Assumptions

Consider a single state population without age structure. There are $K$ primary
periods and each primary period has $l_i$ shorter secondary periods.

$p_{ij}$ : the secondary period capture probability, \textit{i.e.} the probability that an animal is
captured during secondary period $j$ of primary period $i$, $q_{ij} = 1 - p_{ij}$.

$p_i^\star$ : the primary period capture probability, \textit{i.e.} the probability that an animal is
caught in at least one of the secondary periods during primary period \( i \), \( p_i^* = 1 - \prod_{j=1}^{h_i}(1 - p_{ij}) \).

\( \varphi_i \) : the probability that an animal is alive in primary period \( i + 1 \) given that it was alive during primary period \( i \).

\( B_i \) : the number of new recruits (\textit{in situ} births or immigrants) to the population between primary period \( i - 1 \) and primary period \( i \). \( B_0 = N_1 \) is the number of animals in the population at the initial sampling period.

\( u_i \) : the number of unmarked animals caught on at least one secondary occasion within primary period \( i \). The newly captured unmarked animals \( u_i \) can be

- animals that enter the super population before period \( i \), survive to period \( i \), but are never captured before period \( i \)
- animals enter the super population at period \( i \).

\( \beta_i \) : the probability that a member of the super population enters it between primary period \( i \) and \( i + 1 \), and survives until sampling period \( i + 1 \). \( \beta_0 \) is the probability that a member of the super-population came into the population before the study started and survived to primary period 1.

\( m_{ij} \) : the number of animals caught in primary period \( j \) that were last caught in primary period \( i \) \((1 \leq i < j \leq K)\).

\( R_i \) : the number of marked animals that are released back into the population following primary period \( i \).

\( \chi_{\omega}^i \) : the number of animals having capture history \( \omega \) over the \( l_i \) secondary periods of primary period \( i \) given that the animals are part of \( R_i \).

Assuming no tag mortality and all captured animals are released back to the population, we have \( R_i = u_i + \sum_{j=1}^{i-1} m_{ji} \). If we consider an individual’s population of origin, we can decompose \( u_i = u_i^B + u_i^I \), \( \beta_i = \beta_i^B + \beta_i^I \), \( R_i = R_i^B + R_i^I \), and \( m_{ij} = m_{ij}^B + m_{ij}^I \), where

\( u_i^B \) (\( u_i^I \)) : the number of newly caught animals at primary period \( i \) that are \textit{in situ} births (immigrants).
**1.3 Model Structure**

**1.3.1 Standard Super Population Capture-Recapture Model with Robust Design**

In the standard super population capture-recapture model, the likelihood for the model can be written as

\[ L = L_1 * L_2 * L_3 \]  \hspace{1cm} (1.1)

where \( L_1 \) deals with the capture of unmarked animals during the primary periods (the super population part), \( L_2 \) is the probability distribution of the recapture data of primary periods conditional on the releases \( R_i \), and \( L_3 \) is the likelihood corresponding to the secondary period captures when using the Robust Design. For ease of presentation, we here assume that there are two secondary periods within each primary period, but other numbers of secondary periods are easily modeled.

For the newly captured unmarked animals, the random variables \( \{ B_1, B_2, \cdots, B_K \} \) are assumed to follow a multinomial distribution with parameters \( \{ \beta_0, \beta_1, \cdots, \beta_{K-1} \} \). Define \( \psi_i \) as the probability that a member of the super population is alive and unmarked at primary period \( i \) (Schwarz and Arnason 1996). It is convenient to model the numbers of newly captured animals during primary periods with a recursive formula.

\[
\psi_1 = \beta_0 \\
\psi_{i+1} = \psi_i (1 - p_i^*) * \varphi_i + \beta_i \\
E(u_i) = N * \psi_i * p_i^* 
\]  \hspace{1cm} (1.2)
We have

\[ L_1(\{\{\{N, \{\beta_i\}, \{p_i^\star\}, \{\varphi_i\}|u_i}) = \{\frac{N!}{u!(N-u)!}(\sum_{i=1}^{K}\psi_ip_i^\star)^u * (1 - \sum_{i=1}^{K}\psi_ip_i^\star)^{N-u}\}
\]

*{\frac{u!}{u_1!u_2! \cdots u_K!}(\prod_{i=1}^{K}(\sum_{i=1}^{K}\psi_ip_i^\star))}\]

(1.3)

where \(u = \sum_{i=1}^{K} u_i\).

Similarly, for recaptured animals of primary period \((L_2)\), define new parameters \(\phi_i\),

\[
\phi_{i,i+1} = \varphi_i, \ i = 1, 2, \cdots, K
\]

\[
\phi_{i,j} = \phi_{i,j-1} * (1 - p_j^\star) * \varphi_j, \ j = \text{min}(i+2, K), \cdots, K
\]

(1.4)

Let \(m_i = \sum_{j=i+1}^{K} m_{ij}\) and \(\phi_i = 1 - \sum_{j=i+1}^{K} \phi_{i,j} * p_j^\star\). Conditional on the releases during primary period \(i\) \((R_i)\), all newly recaptured animals \(\{m_{i,i+1}, m_{i,i+2}, \cdots, m_{i,K}, m_i\}\) from the following sampling periods are assumed to follow a multinomial distribution. We also assume that the animals released in one primary period are independent from the animals released in another different primary period. The likelihood of \(L_2\) will then be the product of the multinomial distributions conditional on \(R_1, R_2, \cdots, R_{K-1}\). Then the likelihood for the recapture part of the primary period is

\[
L_2(\{\{\varphi_i\}, \{p_i^\star\}|\{R_i\}, \{m_{ij}\}) = \prod_{i=1}^{K-1} \frac{R_i!}{m_i!(R_i - m_i)!} * \prod_{j=i+1}^{K-1} (\phi_{i,j}p_j^\star)^{m_{ij}} * (\phi_i)^{R_i - m_i}
\]

(1.5)

Within every primary period, the population is closed. The animals can be captured during any of the secondary periods. It is reasonable to assume that the capture-recapture animals of all the secondary periods within each primary period follow a multinomial distribution conditional on the number of animals released to that primary period \(R_i\). Then

\[
L_3(\{p_{ij}\}|\{\chi_i^\star, \{R_i\}) = \prod_{i=1}^{K} \frac{R_i}{\chi_i^{10} \chi_i^{01} \chi_i^{11}!} * \left(\frac{q_{i1}q_{i2}}{p_i^\star} \chi_i^{10} \left(\frac{q_{i1}p_{i2}}{p_i^\star} \chi_i^{01} \left(\frac{q_{i1}q_{i2}}{p_i^\star} \chi_i^{11}\right)\right)\right)
\]

(1.6)
1.3.2 Super Population Capture-Recapture Model Augmented with Perfect Assignment Information

Now we consider incorporating information on population of origin from an assignment procedure into the super population model. If the assignment procedure is “perfect”, we can split newly captured unmarked animals \( u_i = u^B_i + u^I_i \) according to their assigned populations of origin. Correspondingly, we can decompose the entry probability according to population of origin as \( \beta_i = \beta^B_i + \beta^I_i \). Without generality, we can assume that individuals born on site have different survival probability from immigrants. The likelihood of the super population capture-recapture model augmented with perfect assignment information now becomes

\[
L = L^*_1 * L^*_2 * L_3
\]  

where \( L^*_1 \) is the likelihood for the newly captured individuals incorporating information on population of origin, \( L^*_2 \) is the modified likelihood for the recapture component of primary periods conditional on the releases \( R^B_i \) and \( R^I_i \), and \( L_3 \) is the likelihood for the secondary period component. \( L_3 \) is the same as in the standard super population capture-recapture model if we assume that individuals born \textit{in situ} and immigrants have the same capture probabilities.

Conditional on the super population \( N \), the newly captured unmarked individuals born \textit{in situ} and immigrants \((u^B_1, u^B_2, \ldots, u^B_K, u^I_1, u^I_2, \ldots, u^I_K, N - \sum (u^B_i + u^I_i))\) are assumed to follow a multinomial distribution. For animals born \textit{in situ}:

\[
\begin{align*}
\psi^B_1 &= \beta^B_0 \\
\psi^B_{i+1} &= \psi^B_i (1 - p^*_{i}) * \phi^B_i + \beta^B_i, \ i = 1, \ldots, K - 1 \\
E(u^B_i) &= N * \psi^B_i * p^*_{i}, \ i = 1, \ldots, K
\end{align*}
\]  

(1.8)

For immigrants:

\[
\begin{align*}
\psi^I_1 &= \beta^I_0 \\
\psi^I_{i+1} &= \psi^I_i (1 - p^*_{i}) * \phi^I_i + \beta^I_i, \ i = 1, \ldots, K - 1 \\
E(u^I_i) &= N * \psi^I_i * p^*_{i}, \ i = 1, \ldots, K
\end{align*}
\]  

(1.9)
Let $u. = \sum_{i=1}^{K}(u^B_i + u^I_i)$ and $p. = \sum_{i=1}^{K}(\psi^B_i * p^*_i + \psi^I_i * p^*_i)$, then

$$L^*_1(N, \beta^B_i, \beta^I_i, p^*_i, \varphi^B_i, \varphi^I_i | u^B_i, u^I_i)$$

$$= \frac{N!}{u.!(N-u.)!}(p.)^{u.}(1-p.)^{N-u.}\frac{u.}{u^B_1 \cdots u^B_K \cdots u^I_K}$$

$$\times \prod_{i=1}^{K} \left( \frac{\psi^B_i * p^*_i}{p.} \right)^{u^B_i} \left( \frac{\psi^I_i * p^*_i}{p.} \right)^{u^I_i}$$

(1.10)

For the recapture part of primary periods augmented with assignment information $(L^*_2)$,

$$\phi^r_{i+1} = \varphi^r_i, \ i = 1, 2, \cdots, K, \ r = B, I$$

$$\phi^r_{i,j} = \phi^r_{i,j-1} * (1-p^*_j) * \varphi^r_j, \ j = i+2, \cdots, K$$

(1.11)

Let $m^r_i = \sum_{j=i+1}^{K} m^r_{ij}$ and $\phi^r_i = 1 - \sum_{j=i+1}^{K} \phi^r_{i,j} * p^*_j$, we have

$$L^*_2(\varphi^B_i, \varphi^I_i, p^*_i | R^B_i, R^I_i, m^B_{ij}, m^I_{ij})$$

$$= \prod_{r=B,I} \left\{ \prod_{i=1}^{K-1} \frac{R^r_i!}{\prod_{j=i+1}^{K} m^r_{ij}!(R^r_i - m^r_{ij})!} \prod_{j=i+1}^{K-1} \left( \phi^r_{i,j} p^*_j \right)^{m^r_{ij}} * \left( \phi^r_i \right)^{R^r_i - m^r_i} \right\}$$

(1.12)

However, perfect assignment procedures are unlikely to exist in practice, which leads us to imperfect assignment procedures.

### 1.3.3 Super Population Capture-Recapture Model Integrated with Imperfect Assignment Information

As in the case where we had a perfect assignment procedure, any captured individual for which birth state is unknown is assigned a population of origin. However, now the assignment procedure is subject to uncertainty. Consider an animal genotyped at capture in period 2. The assignment procedure assigns the animal as an *in situ* birth with probability $\pi$ ($0 < \pi < 1$). We view $\pi$ as the probability that the animal is born *in situ*, and $1 - \pi$ as the probability that it is an immigrant. If the assignment test is a good one then when the animal is truly born *in situ*, $\pi$ will be close to one and $(1 - \pi)$ will be close to zero; whereas when the animal is truly an immigrant, $\pi$ will be close to zero and $(1 - \pi)$ will be close to one.

To generalize our model to the more realistic case of imperfect assignment, we will take a resampling approach similar to multiple imputation (Rubin, 1987) to impute
the perfect assignment data based on the imperfect information. Consider that each newly captured animal, \( j (j = 1, \cdots, N) \), has an assignment probability \( \pi_j, 0 < \pi_j < 1 \). We will impute a “pseudo perfect assignment data set” \( m \) times using the following procedures.

First, each newly captured animal will be randomly allocated as a uniform random variable \( U_j \in (0, 1) \). If \( U_j \in (0, \pi_j) \), we assign this animal as born \textit{in situ}, whereas if \( U_j \in (\pi_j, 1) \), we assign this animal as an immigrant. We will apply this assignment procedure to all newly captured animals in all primary sampling periods. Then we will split \( u_i \) into \( u_i^B \) and \( u_i^I \) as in the perfect assignment situation and get a pseudo perfect assignment data set. We repeat the imputation of this kind of pseudo perfect assignment data \( m \) times. For each of the \( m \) “pseudo perfect assignment data”, we treat it as super population capture-recapture model with perfect assignment to estimate the entry probability, survival probability, and capture probability for \textit{in situ} births and immigrants. Then we can take various distributional statistics (means, medians, standard deviation, and percentiles) of the estimators. By so doing, we obtain estimators that reflect the inherent uncertainty due to the assignment data being imperfect, in addition to the usual uncertainty associated with capture-recapture inference. So the likelihood will the be same as that of super population capture-recapture model added with perfect information.

There is uncertainty in the pseudo perfect data imputed by our resampling approach, which is caused by the assignment process. The pseudo perfect assignment data changes from one imputation to another imputation. So for the variance of the estimates, it is composed of within-imputation variance and between imputation variance, which is the same as that of the multiple imputation. For example, for the variance of the estimate of \( \beta_0^B \), it is calculated as follows.

\[
\hat{\text{var}}(\beta_0^B) = \frac{\sum_{j=1}^{m} \sigma_j}{m} + (1 + \frac{1}{m}) \cdot \frac{1}{m - 1} \sum_{j=1}^{m} ((\hat{\beta}_0^B)_j - (\bar{\hat{\beta}}_0^B)^2)
\]

where the first term is the natural variability inherent to the data and is calculated by the information matrix. \( \sigma_j \) is the variance of \( \beta_0^B \) as the generated data is the true data (no imputation needed). The second term is the between imputation variance.

### 1.3.4 Assumptions

The super population capture-recapture model with the Robust Design and assignment data is a generalization of the standard super population model, and therefore
the assumptions required are similar.

1. All members of the super population \( N \) entering the population at the same time from the same population of origin have the same entry probability (homogeneous entry probabilities).

2. Every animal has the same probability of surviving from one period to the next period (homogeneous survival probabilities).

3. Each marked animal present during sampling period \( i \) has the same probability of capture at sampling period \( i \) (homogeneous capture probabilities).

4. Marks do not affect the survival or behavior of the animal, are not lost, and are recorded correctly.

5. Each animal behaves independently with respect to survival, entry, and detection probabilities.

6. A valid assignment procedure has been carried out for all animals where population of origin is unknown. This implies that there are no errors in the genotyping used in the tests (however, there will be uncertainty in the assignments, regardless).

### 1.4 Simulation

Remember, if the imperfect assignment procedure is a valid one, given that the animal is born on site, its assignment probability should be close to 1, and the higher the value, the better the imperfect assignment procedure. Similarly, if the animal is an immigrant, its assignment probability should be close to zero, and the smaller the value, the better the assignment procedure. When the assignment probability is one (zero) given the animal is born in situ (is an immigrant), the assignment procedure is the perfect assignment procedure. In real data analysis, we have the specified imperfect assignment probability for each newly captured animal. To generate the required imperfect assignment probabilities for all newly captured unmarked animals, we use the following procedure. First, we generate a data set \( u_i^B \) and \( u_i^I \) based on the true parameter values, \( \beta_i^B, \beta_i^I, p_{ij}, \phi_i \). Then for the \( j \)th animal belonging to \( u_i^B \) we randomly give it a high assignment probability, for example
\( \pi_j \in U(0.95, 1), (j = 1, \cdots, u^B_i) \), and \( 1 - \pi_j \) will be the probability that the animal has immigrated from outside. Similarly for the \( j \)th animal belonging to \( u^I_i \) we randomly give it a small assignment probability, for example, \( \pi_j \in U(0, 0.05), (j = 1, \cdots, u^I_i) \), and \( 1 - \pi_j \) will be the probability that the animal has immigrated from outside.

In our simulations, we explored imperfect assignment procedures with four different levels of uncertainty. The first (perfect) case assumes the assignment probability to be one (zero) if the animal is an in situ birth (immigrant); the second assumes the assignment probability is uniform between (0.95, 1) if the animal is born in situ and uniform between (0, 0.05) if it is an immigrant; the third assumes the assignment probability is uniform between (0.9, 1) if the animal is born in situ and uniform between (0, 0.1) if it is an immigrant; and the fourth assumes the assignment probability is uniform between (0.8, 1) if the animal is born in situ and uniform between (0, 0.2) otherwise.

In our simulations, the sampling periods are fixed as five years with three secondary periods close together each year. The capture probabilities are the same for the three secondary periods within each year and the survival probabilities are the same for animals born in situ and immigrants. However, the entry, primary capture and survival probabilities are dependent on time. The parameters take different values with secondary capture probability taking values 0.5 and 0.2, survival probability taking values of 0.4 and 0.15, and super population size taking values 1000 and 10000. For the entry probabilities, we also consider two different cases. In one situation, the entry probabilities are assumed to be equal for all sampling periods and for all immigrants and in situ births. In the other situation, the entry probabilities are assumed to be unequal, with the first sampling period having more new recruits than later sampling periods, and there are always more in situ births than immigrants. This latter scenario is reasonable in practice because most animals stay in the same population throughout their lives. Following a complete factorial design, our simulations were done under sixteen different scenarios. Here we only report a representative subset of the simulation results.

First, we want to test the performance of the assignment data on the super population capture-recapture model. We compare the simulation results for the super population capture-recapture model with imperfect assignment data having \( \pi \in (0.95, 1) \) (Imperfect model) and the capture-recapture model without assignment data (Standard model). The parameters of standard model take values corresponding to those of the imperfect model
with $\beta_i = \beta_i^B + \beta_i^I$. Table 2.1 reports simulation results from 1000 repetitions with $N = 1000$ under four different situations. From Table 2.1, we can see that the imperfect model has low relative bias when the entry probability is really low with $\beta_i^I = 0.06$, $i = 1, 2, 3, 4$. For the relative bias of the survival probability estimators, the imperfect model has smaller relative bias than the standard one when the survival probability is big enough (0.4 here). For both models, the higher the parameter values, the better the estimators (lower relative bias and relative SE). One big advantage of adding assignment information is that the imperfect model can estimate $\beta_i^B$ and $\beta_i^I$ separately whereas this cannot be achieved by the standard model without such information. The bias of the imperfect model ranges from 2% to 3.69% and most of the sample standard errors are less than 20% even when the parameter values are as low as $\varphi_i = 0.15$ and $\beta_i^I = 0.06$. The imperfect model can also estimate $\varphi_i^B$ and $\varphi_i^I$ separately which we did not simulate here. We have similar results for the other eight scenarios.

Now we test how the uncertainty of the assignment procedure affects the estimators of the model. Here we only report the simulation results for parameter values $p_{ij} = 0.5, \varphi_i = 0.4, \beta_0^B = 0.2, \beta_i^B = 0.12, \beta_0^I = 0.1, \beta_i^I = 0.06, N = 1000$ as in Figure 1.1, Figure 1.2, and Figure 1.3.

From Figure 1.1, we can see that as the reliability of the assignment procedure increases from $\pi \in (0.8, 1)$ to $\pi \in (0.9, 1)$ to $\pi \in (0.95, 1)$ to $\pi = 1$ the relative bias of the estimates of the in situ entry probability and immigrant entry probability decrease dramatically. Estimates of survival probability and capture probability show no such trend. This is because the assignment data reflects the animal’s population of origin, which directly affects the estimate of the entry probability. Estimates of capture probability and survival probability are mainly based on the capture-recapture data. The relative standard errors of the estimates of the immigrant and in situ birth entry probabilities also decrease as the assignment procedure becomes more reliable Figure 1.2. Again, the relative standard error associated with capture and survival probabilities are not influenced by the uncertainty of the assignment procedure. Figure 1.1, for the estimates of the it birth entry probability and immigrant entry probability, the relative root of mean square error (RMSE) decrease quickly as the assignment procedure becomes better and better, which is similar as the relative bias. For all other parameter values, the simulation results suggest similar conclusions about the relative bias and the relative standard error. So we can say that the decrease of the RRMSE
Figure 1.1: Relative bias of 1000 simulations for the super population capture-recapture model augmented with different assignment procedures. $p_{ij} = 0.5, \varphi_i = 0.4, \beta^B_0 = 0.2, \beta^B_i = 0.12, \beta^l_0 = 0.1, \beta^l_i = 0.06, N = 1000$
Figure 1.2: Relative sample standard error of 1000 simulations for the super population capture-recapture model augmented with different assignment procedures. $p_{ij} = 0.5, \varphi_i = 0.4, \beta_0^B = 0.2, \beta_t^B = 0.12, \beta_0^I = 0.1, \beta_t^I = 0.06, N = 1000$
Figure 1.3: Relative RMSE of 1000 simulations for the super population capture-recapture model integrated with different assignment procedures. $p_{ij} = 0.5, \varphi_1 = 0.4, \beta_0^B = 0.2, \beta_i^B = 0.12, \beta_0^I = 0.1, \beta_i^I = 0.06, N = 1000$
Figure 1.4: Relative RMSE for the simulation results of the super population capture-recapture model integrated with four different levels of uncertainty of assignment information ($\pi = 1, \pi \in (0.95, 1), \pi \in (0.9, 1)$ and $\pi \in (0.8, 1)$). Parameters take a complete combinations of $p_{ij} = 0.5, 0.2, \varphi_i = 0.4, 0.15, N = 1000, 10000, \beta^B_0 = 0.4, \beta^I_0 = 0.1, \beta^B_i = 0.12, \beta^I_i = 0.06, i = 1, \cdots, K - 1$.

of the estimates are mainly caused by the decrease of the relative bias of the estimates.

Finally, we want to see the robustness of the super population capture-capture model to different uncertainty levels genetic tests and different parameter values. Figure 2.4 compares the RRMSE of the estimates of all the parameters for eight different parameter values, four different levels of assignment uncertainty, and the standard model. From Figure 2.4 we can see that more than 95% of the RRMSE are less than 10% when $N = 10000$ no matter what the uncertainty of the assignment procedure. When $N = 1000$, most RRMSE are less than 20% for all four uncertainty levels.

1.5 Example: Kangaroo Rats in Arizona

To illustrate the application of our models, we use data from a longterm capture-recapture study combined with assignment information from banner-tailed kangaroo rats *Dipodomys spectabilis* in southern Arizona, USA (Waser and Elliott, 1991; Skvarla *et al.*, 2004; Waser *et al.* in press). Animals were trapped in a set of 8 small habitat patches separated by 200-1000m of uninhabited terrain; we refer to these as “populations” although
they are linked by a small but measurable rate of movement. Here, we use trapping data collected using Robust Design methodology between 1994 and 2001. We use data from one primary period each year in March, near the end of the breeding season. During each primary period, all active dens in each population were trapped during three secondary periods. Each trapped animal was given a numbered ear tag and its trapping location, sex, and age (“juvenile” for animals less than one year old, “adult” for animals older than one year) were recorded.

In parallel with demographic data, approximately 90% of all individuals were genotyped at an average of 7.7 microsatellite loci (Waser et al. in press). We used a genetic assignment procedure implemented in program GENECLASS2 (Piry et al. 2004) to infer population of origin. To adapt the data to our model for one age group and one population, we used capture-recapture data only from the central kangaroo rat population and ignored age and sex. We used genotypes from the central population to define the genetic characteristics expected from individuals born in that population; genotypes from animals in the seven surrounding populations were used to define the expected genetic characteristics of immigrants.

We first used GENECLASS2 to estimate the probability that each individual’s genotype is drawn from the sampled population using the Bayesian procedure developed by Rannala and Mountain (1995), which has been shown Piry et al. 2004) to have superior properties. This gave us an assignment probability vector \((\pi_n, 1 - \pi_n), n = 1, \cdots, \sum_{i=1}^{K} u_i\). Since no genetic assignment procedure provides “perfect” assignment, we first modeled the “imperfect” case, where \(0 < \pi_n < 1\). We also handled the data as though it were “perfect” although this is not realistic. If \(\pi_n = \max(\pi_n, 1 - \pi_n), (n = 1, \cdots, \sum_{i=1}^{K} u_i)\), we assigned the animal to be born in situ by giving it \(\pi_n = 1\). Otherwise, we inferred that this animal was an immigrant from outside the population by assigning \(\pi_n = 0\).

To examine the effects of adding assignment information to capture-recapture data, we compared the results of the super population capture-recapture models with “perfect” and “imperfect” assignment information. We also compared those results to those of a standard super population model with the Robust Design, leaving out the genetic information. The estimates and standard error estimates of the parameters are listed in Table 2.2, Table 4.1, and Figure 1.6, and Figure 2.5. From Table 2.2, we can see that the perfect and imperfect assignment models can estimate \(\beta^B_i, \beta^I_i, \varphi^B_i, \varphi^I_i\) separately while the standard
model cannot. From Table 2.2 and Table 4.1, the estimates of the entry probability and capture probability under these three different models are similar to each other with $\hat{\beta}_i$ (from the standard model) $\approx \hat{\beta}_i^B$ (from the assignment models) $+ \beta_i^I$ (from the assignment models) and $\hat{p}_{ij}$ (from the standard model) $\approx \hat{p}_{ij}$ (from the genetic models). From Figure 1.6, the estimate of capture probability, and even the estimate of standard error of the estimates are very close between the “perfect” assignment model and the standard model. In Table 4.1, 12 out of the 14 survival probability estimates are close in value between these tests. Figure 2.5 indicates that from the standard model, $\beta_0 > \beta_i$ (from the standard model); from the assignment model $\beta_0^B > \beta_0^I, \beta_0^r > \beta_r^i, i = 1, \ldots, 7, r = B, I$. All these are what we should expect from biological facts.

In the kangaroo rat data set, the effective sample size (the number of times that individuals were caught and released at least once during primary sample periods) is 279. To choose the “best” model for the kangaroo rat data set, we considered 14 different models with the full model being the most general model with capture probability dependent on year, and entry probability and survival probability dependent on both time and origination. First we calculated the variance inflation vector $\hat{c} = \frac{\chi^2}{df}$ to correct for possible overdispersion.
Figure 1.6: Comparison of capture probabilities and the 95% confidence interval of its estimates from the kangaroo rat data set under three different models, super population capture-recapture model with perfect assignment information (Perfect genetic model), super population capture-recapture model with imperfect assignment information (Imperfect genetic model) and standard super population model capture-recapture model without assignment information (standard model).

in the data set, where $\chi^2$ is calculated by the Pearson goodness-of-fit chi-square statistic of the full model. When $c = 1$, there is no overdispersion; when $c > 1$ there is overdispersion in the data and the proper likelihood is $\log(L)/c$. The accepted range of overdispersion should be $1 \leq c \leq 4$. The kangaroo rat data set under our full model has little overdispersion with $\hat{c} = 2.60 (\chi^2 = 213.2096, df = 82)$. When $\hat{c} > 1$, the AIC model selection statistic can be modified to QAIC (Burnham and Anderson, 1998). $QAIC_c$ is a more general model selection criterion which adjusts for overdispersion and small sample size of capture-recapture data. We also used $\Delta QAIC_C$ to compare and rank the candidate models. We may ignore models with $\Delta QAIC > 10$ because these models cannot explain some important variation in the data (Burnham and Anderson, 2002). The BIC, $QAIC_C$, $\Delta QAIC_C$ and QAIC weight values for all fourteen models are listed in Table 1.5.

$$QAIC_C = -\{2 \ln(\frac{L(\hat{\theta})}{\hat{c}})\} + 2p + \frac{2p(p + 1)}{n - p - 1}$$ (1.14)
\[ \omega_i = \frac{\exp(-\Delta_i/2)}{\sum_r \exp(-\Delta_r/2)} \]  \hspace{1cm} (1.15)

where \( K \) is the number of parameters, \( n \) is the sample size. Models with lowest \( QAIC_c \) Anderson et al. (1994) proved that \( QAIC_c \) criteria performed well in product multinomial models of capture-recapture data when different levels of overdispersion exist. When \( c = 1 \), \( QAIC_c \) reduce to \( AIC_c \).

In Table 1.5, we can that the model \( p(c,c)\beta(t,o)\varphi(c,o) \) is the most suitable model for the kangaroo rate data set. This model assumes that capture probability is the same for all animals, that entry probability is dependent on time and population of origin, and that survival probability is depends only on the population of origin source. Now we analyze the Kangaroo rat data set under the “best” model \( p(c,c)\beta(t,o)\varphi(c,o) \) we selected and the results are in Table 4.1.

From Table 4.1, we can see that there appear to be many more \textit{in situ} births than immigrants for all periods. The \textit{in situ} entry probability for the initial period is 0.143, much higher than entry probabilities for later sampling periods or for immigrant entry probabilities during any entry period. Both results match expectations. Most kangaroo rats were first trapped when they were very young, so they were likely to have been born within the study population. During the initial sampling period, \textit{in situ} recruits included all animals that had previously entered the population and survived to the first sampling period. For immigrants, we expect no such trend. In addition, we see that survival probability for animals born \textit{in situ} is higher than that for immigrants, which makes since if individuals born in the population \textit{in situ} are more familiar with the local environment.

1.6 Discussion and Future Research

In this paper, we considered a single state super population capture-recapture model augmented with both perfect and imperfect assignment data. We can extend the model to a single population with two age groups and even to a more complex situation with multiple states and multiple age groups which was the original model motivated by the kangaroo rat study.

In the models, we assume homogeneous capture entry probability, capture probability, and survival probability among animals. However, in practice, these assumptions
may be violated. Inherent characteristics like gender and age can affect the capture probabilities (heterogeneity). Animals may become shy of traps after their initial capture or seek them out (trap response). We can introduce time, heterogeneity and trap response models into the last component of the likelihood. We have allowed the immigrants have different survival probability from the individuals born in situ and could also allow that these two sets of animals have different capture probabilities. However, the more general of the model, the more parameters would need to be estimated. The super population size needs to be big enough to provide enough recaptured animals to make the parameters estimable.

To validate the genetic assignment test, we assign the genetic assignment probability to a sample of animals with known origination, like young animals which are usually their natal place in their first year of their lives. If the genetic assignment test is a good one, the genetic assignment probability for the young animals should be very high, like bigger than 0.95, as the higher the better the genetic assignment test. However, when we use GENECLASS2 to analyze the genetic information of the sample of young animals, we found that the assignment methods “generally result in an excess of resident individuals being excluded”, i.e., scored as immigrants. Waser (personal communication) suspects that the biological reason for this is that immigrants tend to have alleles that are common in their birth population but locally rare in the population they move to. However they pass some of those alleles to their offspring, and their offsprings progeny, etc who then may be incorrectly judged to be immigrants too. Thus the assignment procedure is biased towards “immigration”. For example, when we run GENECLASS with data from the population we are interested in vs. everything else, it correctly assigns about 70% captured in population to population. A few percent may really come from outside, but the true number is surely far below 30%, as many of these individuals were first captured as small juveniles and/or have known parents in population. Waser suggested one way to deal with this problem might be to use assignment test results only when it is relatively unambiguous, for example for individuals that are 10x as likely to come from one population than the other(s), and to treat the other individuals as having no genetic information. We hope to explore this and other ideas in future work.

Based on genetic parentage determination rather than assignment tests, Waser (personal communication) found that for animals whose parents can be determined, only 7% live as adults in a different population than their parents. So in our real data analysis
here, there is only a very small proportion of tagged animals (at most a few percent) are estimated to disperse between populations (Skvarla and Nichols, 2004). Therefore the entry probabilities from outside should be really small. However, in the simulation part, we assume that the entry probabilities of the immigrants are reasonably large to achieve reasonable precision. If the genetic assignment test wrongly assigns the animal to immigrant when the animal is truly an *in situ* birth or if the genetic assignment test wrong assigns the animal to *in situ* birth when the animal is truly an immigrant, misclassification happens. Yoshizaki *et al.* (2009) discussed the effect of misclassification in dna finger printing closed population capture-recapture studies. We may be able to use similar ideas to estimate the effect of misclassification of genetic assignment tests to our open population model estimates here.
Table 1.1: Comparing simulation results of the super population capture-recapture model with imperfect assignment information (“Imperfect” columns) and the standard super population capture-recapture model (“Standard” columns). The uncertainty level of the “imperfect” assignment procedure is $0.95$, that is, $\pi_i \in (0.95, 1)$, $i = 1, \ldots, \sum_{i=1}^{K}(u_i^B + u_i^I)$. The super population size is $N = 1000$.

<table>
<thead>
<tr>
<th>Par.</th>
<th>Val</th>
<th>Rel bias</th>
<th>Rel SE</th>
<th>Rel bias</th>
<th>Rel SE</th>
<th>True Rel bias</th>
<th>Rel SE</th>
<th>Rel bias</th>
<th>Rel SE</th>
<th>Rel bias</th>
<th>Rel SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0^B$</td>
<td>0.20</td>
<td>-1.78</td>
<td>7.03</td>
<td>0.49</td>
<td>5.62</td>
<td>0.20</td>
<td>-0.79</td>
<td>7.03</td>
<td>-0.06</td>
<td>5.69</td>
<td></td>
</tr>
<tr>
<td>$\beta_0^I$</td>
<td>0.10</td>
<td>3.65</td>
<td>10.02</td>
<td>0.10</td>
<td>2.12</td>
<td>0.30</td>
<td>-1.44</td>
<td>9.55</td>
<td>-0.33</td>
<td>8.17</td>
<td></td>
</tr>
<tr>
<td>$\beta_1^B$</td>
<td>0.12</td>
<td>-0.35</td>
<td>10.64</td>
<td>0.12</td>
<td>1.42</td>
<td>9.61</td>
<td>0.12</td>
<td>-2.90</td>
<td>9.61</td>
<td>0.49</td>
<td>7.45</td>
</tr>
<tr>
<td>$\beta_1^I$</td>
<td>0.06</td>
<td>1.41</td>
<td>15.80</td>
<td>0.06</td>
<td>2.09</td>
<td>14.16</td>
<td>0.06</td>
<td>4.12</td>
<td>13.81</td>
<td>0.49</td>
<td>7.45</td>
</tr>
<tr>
<td>$\beta_2^B$</td>
<td>0.12</td>
<td>-1.58</td>
<td>9.99</td>
<td>-0.44</td>
<td>8.36</td>
<td>0.12</td>
<td>-1.29</td>
<td>9.61</td>
<td>0.49</td>
<td>7.45</td>
<td></td>
</tr>
<tr>
<td>$\beta_2^I$</td>
<td>0.06</td>
<td>2.68</td>
<td>14.85</td>
<td>-0.45</td>
<td>8.36</td>
<td>0.12</td>
<td>-1.39</td>
<td>9.10</td>
<td>0.49</td>
<td>7.45</td>
<td></td>
</tr>
<tr>
<td>$\beta_3^B$</td>
<td>0.12</td>
<td>-0.98</td>
<td>10.07</td>
<td>-0.16</td>
<td>7.77</td>
<td>0.12</td>
<td>-1.60</td>
<td>10.02</td>
<td>0.10</td>
<td>7.99</td>
<td></td>
</tr>
<tr>
<td>$\beta_3^I$</td>
<td>0.06</td>
<td>2.98</td>
<td>13.56</td>
<td>0.16</td>
<td>7.77</td>
<td>0.12</td>
<td>-2.10</td>
<td>10.02</td>
<td>0.10</td>
<td>7.99</td>
<td></td>
</tr>
<tr>
<td>$\beta_4^B$</td>
<td>0.12</td>
<td>-2.01</td>
<td>9.94</td>
<td>-0.30</td>
<td>7.97</td>
<td>0.12</td>
<td>-1.60</td>
<td>10.02</td>
<td>0.10</td>
<td>7.99</td>
<td></td>
</tr>
<tr>
<td>$\beta_4^I$</td>
<td>0.06</td>
<td>2.62</td>
<td>14.02</td>
<td>0.16</td>
<td>7.97</td>
<td>0.12</td>
<td>-2.10</td>
<td>10.02</td>
<td>0.10</td>
<td>7.99</td>
<td></td>
</tr>
<tr>
<td>$\phi_1$</td>
<td>0.40</td>
<td>-0.17</td>
<td>8.32</td>
<td>0.02</td>
<td>8.13</td>
<td>0.15</td>
<td>1.63</td>
<td>15.57</td>
<td>0.48</td>
<td>16.85</td>
<td></td>
</tr>
<tr>
<td>$\phi_2$</td>
<td>0.40</td>
<td>0.41</td>
<td>7.96</td>
<td>0.13</td>
<td>7.77</td>
<td>0.15</td>
<td>1.65</td>
<td>18.10</td>
<td>0.28</td>
<td>17.66</td>
<td></td>
</tr>
<tr>
<td>$\phi_3$</td>
<td>0.40</td>
<td>0.17</td>
<td>8.34</td>
<td>0.54</td>
<td>8.48</td>
<td>0.15</td>
<td>1.06</td>
<td>18.42</td>
<td>1.42</td>
<td>19.02</td>
<td></td>
</tr>
<tr>
<td>$\phi_4$</td>
<td>0.40</td>
<td>-0.09</td>
<td>8.85</td>
<td>-0.27</td>
<td>8.43</td>
<td>0.15</td>
<td>-0.99</td>
<td>20.23</td>
<td>0.48</td>
<td>19.52</td>
<td></td>
</tr>
</tbody>
</table>

$p_{ij} = 0.5$

<table>
<thead>
<tr>
<th>Par.</th>
<th>Val</th>
<th>Rel bias</th>
<th>Rel SE</th>
<th>Rel bias</th>
<th>Rel SE</th>
<th>Rel bias</th>
<th>Rel SE</th>
<th>Rel bias</th>
<th>Rel SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0^B$</td>
<td>0.20</td>
<td>-0.08</td>
<td>12.77</td>
<td>0.20</td>
<td>12.28</td>
<td>0.10</td>
<td>2.78</td>
<td>16.46</td>
<td>-0.66</td>
</tr>
<tr>
<td>$\beta_0^I$</td>
<td>0.10</td>
<td>3.44</td>
<td>16.40</td>
<td>0.10</td>
<td>2.78</td>
<td>16.46</td>
<td>-0.66</td>
<td>12.12</td>
<td></td>
</tr>
<tr>
<td>$\beta_1^B$</td>
<td>0.12</td>
<td>-3.37</td>
<td>21.69</td>
<td>-1.66</td>
<td>19.98</td>
<td>0.12</td>
<td>0.76</td>
<td>20.05</td>
<td>1.17</td>
</tr>
<tr>
<td>$\beta_1^I$</td>
<td>0.06</td>
<td>-0.06</td>
<td>26.79</td>
<td>0.06</td>
<td>2.39</td>
<td>23.40</td>
<td>0.10</td>
<td>7.99</td>
<td></td>
</tr>
<tr>
<td>$\beta_2^B$</td>
<td>0.12</td>
<td>-2.47</td>
<td>19.95</td>
<td>0.26</td>
<td>16.67</td>
<td>0.12</td>
<td>-1.27</td>
<td>19.05</td>
<td>0.94</td>
</tr>
<tr>
<td>$\beta_2^I$</td>
<td>0.06</td>
<td>2.92</td>
<td>27.27</td>
<td>0.06</td>
<td>0.94</td>
<td>24.24</td>
<td>0.10</td>
<td>7.99</td>
<td></td>
</tr>
<tr>
<td>$\beta_3^B$</td>
<td>0.12</td>
<td>-1.89</td>
<td>18.91</td>
<td>-0.58</td>
<td>16.63</td>
<td>0.12</td>
<td>-2.36</td>
<td>18.99</td>
<td>-1.15</td>
</tr>
<tr>
<td>$\beta_3^I$</td>
<td>0.06</td>
<td>2.33</td>
<td>24.07</td>
<td>0.06</td>
<td>2.48</td>
<td>24.08</td>
<td>-1.15</td>
<td>16.79</td>
<td></td>
</tr>
<tr>
<td>$\beta_4^B$</td>
<td>0.12</td>
<td>0.49</td>
<td>18.32</td>
<td>-0.02</td>
<td>15.88</td>
<td>0.12</td>
<td>-2.08</td>
<td>19.02</td>
<td>0.16</td>
</tr>
<tr>
<td>$\beta_4^I$</td>
<td>0.06</td>
<td>3.69</td>
<td>25.39</td>
<td>0.06</td>
<td>2.15</td>
<td>23.90</td>
<td>0.16</td>
<td>17.14</td>
<td></td>
</tr>
<tr>
<td>$\phi_1$</td>
<td>0.40</td>
<td>0.51</td>
<td>16.59</td>
<td>1.25</td>
<td>17.28</td>
<td>0.15</td>
<td>2.50</td>
<td>13.62</td>
<td>2.32</td>
</tr>
<tr>
<td>$\phi_2$</td>
<td>0.40</td>
<td>0.44</td>
<td>15.84</td>
<td>1.53</td>
<td>16.64</td>
<td>0.15</td>
<td>1.61</td>
<td>15.58</td>
<td>2.96</td>
</tr>
<tr>
<td>$\phi_3$</td>
<td>0.40</td>
<td>1.42</td>
<td>16.79</td>
<td>1.52</td>
<td>17.12</td>
<td>0.15</td>
<td>2.34</td>
<td>18.19</td>
<td>0.25</td>
</tr>
<tr>
<td>$\phi_4$</td>
<td>0.40</td>
<td>2.36</td>
<td>10.84</td>
<td>2.41</td>
<td>11.92</td>
<td>0.15</td>
<td>0.62</td>
<td>15.90</td>
<td>4.01</td>
</tr>
</tbody>
</table>
Table 1.2: Estimates of entry probabilities from the kangaroo rat data set under models with a “perfect” ($M_1$) population assignment procedure, an “imperfect” population assignment procedure ($M_2$), and without population assignment information ($M_3$).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Perfect model</th>
<th></th>
<th></th>
<th></th>
<th>Imperfect model</th>
<th></th>
<th></th>
<th></th>
<th>standard model</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimator</td>
<td>S.E.</td>
<td>Estimator</td>
<td>S.E.</td>
<td>Estimator</td>
<td>S.E.</td>
<td>Estimator</td>
<td>S.E.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta^B_0$</td>
<td>0.1747</td>
<td>0.0377</td>
<td>0.1591</td>
<td>0.0553</td>
<td>0.2288</td>
<td>0.0427</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta^I_0$</td>
<td>0.0514</td>
<td>0.0177</td>
<td>0.0447</td>
<td>0.0152</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta^B_1$</td>
<td>0.1067</td>
<td>0.0403</td>
<td>0.1228</td>
<td>0.0113</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta^I_1$</td>
<td>0.0176</td>
<td>0.0157</td>
<td>0.0217</td>
<td>0.0054</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta^B_2$</td>
<td>0.0607</td>
<td>0.0295</td>
<td>0.0704</td>
<td>0.0295</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta^I_2$</td>
<td>0.0168</td>
<td>0.0129</td>
<td>0.0020</td>
<td>0.0028</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta^B_3$</td>
<td>0.1131</td>
<td>0.0267</td>
<td>0.1234</td>
<td>0.0027</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta^I_3$</td>
<td>0.0491</td>
<td>0.0161</td>
<td>0.0442</td>
<td>0.0057</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta^B_4$</td>
<td>0.0507</td>
<td>0.0170</td>
<td>0.0421</td>
<td>0.0111</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta^I_4$</td>
<td>0.0165</td>
<td>0.0095</td>
<td>0.0181</td>
<td>0.0058</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta^B_5$</td>
<td>0.0666</td>
<td>0.0171</td>
<td>0.0771</td>
<td>0.0126</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta^I_5$</td>
<td>0.0268</td>
<td>0.0111</td>
<td>0.0295</td>
<td>0.0165</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta^B_6$</td>
<td>0.1046</td>
<td>0.0217</td>
<td>0.0936</td>
<td>0.0179</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta^I_6$</td>
<td>0.0615</td>
<td>0.0169</td>
<td>0.0624</td>
<td>0.0171</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta^B_7$</td>
<td>0.0341</td>
<td>0.0127</td>
<td>0.0397</td>
<td>0.0136</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta^I_7$</td>
<td>0.0491</td>
<td>0.0152</td>
<td>0.0492</td>
<td>0.0133</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 1.3: Estimates of survival probabilities from the kangaroo rat data set under models incorporating “perfect” assignment information (“Perfect model” columns), imperfect assignment information (“Imperfect model” columns), and without assignment information (“Standard” columns).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Perfect model</th>
<th>Imperfect model</th>
<th>Standard model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimator</td>
<td>S.E.</td>
<td>Estimator</td>
</tr>
<tr>
<td>$\varphi_1^B$</td>
<td>0.4053</td>
<td>0.0335</td>
<td>0.4037</td>
</tr>
<tr>
<td>$\varphi_1^I$</td>
<td>0.3764</td>
<td>0.0881</td>
<td>0.4009</td>
</tr>
<tr>
<td>$\varphi_2^B$</td>
<td>0.3999</td>
<td>0.0297</td>
<td>0.4015</td>
</tr>
<tr>
<td>$\varphi_2^I$</td>
<td>0.3727</td>
<td>0.0947</td>
<td>0.2996</td>
</tr>
<tr>
<td>$\varphi_3^B$</td>
<td>0.4015</td>
<td>0.0136</td>
<td>0.4009</td>
</tr>
<tr>
<td>$\varphi_3^I$</td>
<td>0.4002</td>
<td>0.0057</td>
<td>0.3999</td>
</tr>
<tr>
<td>$\varphi_4^B$</td>
<td>0.4012</td>
<td>0.0066</td>
<td>0.4008</td>
</tr>
<tr>
<td>$\varphi_4^I$</td>
<td>0.3944</td>
<td>0.0470</td>
<td>0.4000</td>
</tr>
<tr>
<td>$\varphi_5^B$</td>
<td>0.4005</td>
<td>0.0019</td>
<td>0.4005</td>
</tr>
<tr>
<td>$\varphi_5^I$</td>
<td>0.3981</td>
<td>0.0265</td>
<td>0.3998</td>
</tr>
<tr>
<td>$\varphi_6^B$</td>
<td>0.4007</td>
<td>0.0011</td>
<td>0.4006</td>
</tr>
<tr>
<td>$\varphi_6^I$</td>
<td>0.4003</td>
<td>0.0007</td>
<td>0.4003</td>
</tr>
<tr>
<td>$\varphi_7^B$</td>
<td>0.4008</td>
<td>0.0018</td>
<td>0.4007</td>
</tr>
<tr>
<td>$\varphi_7^I$</td>
<td>0.4004</td>
<td>0.0012</td>
<td>0.4003</td>
</tr>
</tbody>
</table>
Table 1.4: Kangaroo rat data set probability models, ranked from best (lowest QAICc score) to worst. \( p(c,c) \) represents the model where capture probability is constant over time and is the same for in situ birth and immigrant. The first letter of parenthesis indicates whether the parameter is dependent on time (“t”) or is constant over time (“c”). The second letter of parenthesis indicates whether the parameter is different for in situ birth and immigrant (“o”) or is the same for in situ birth and immigrant (“c”). \( \beta t, o \) means the model whose entry probability is dependent on time and is the same for in situ birth and immigrant. \( \varphi (c,c) \) means the model whose survival probability is constant over time and is the same for immigrant and in situ birth.

<table>
<thead>
<tr>
<th>Model name</th>
<th>BIC</th>
<th>QAICc</th>
<th>( \Delta )QAICc</th>
<th>QAICc weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p(c,c)\beta(t,o)\varphi(c,o) )</td>
<td>2007.94</td>
<td>872.94</td>
<td>0.00</td>
<td>0.31</td>
</tr>
<tr>
<td>( p(t,c)\beta(t,o)\varphi(t,o) )</td>
<td>2103.34</td>
<td>892.51</td>
<td>19.58</td>
<td>0.00</td>
</tr>
<tr>
<td>( p(t,c)\beta(t,o)\varphi(c,c) )</td>
<td>2078.15</td>
<td>895.15</td>
<td>2.64</td>
<td>0.08</td>
</tr>
<tr>
<td>( p(c,c)\beta(t,o)\varphi(c,c) )</td>
<td>2063.87</td>
<td>896.29</td>
<td>1.14</td>
<td>0.08</td>
</tr>
<tr>
<td>( p(t,c)\beta(t,o)\varphi(c,o) )</td>
<td>2092.65</td>
<td>905.25</td>
<td>5.27</td>
<td>0.05</td>
</tr>
<tr>
<td>( p(c,c)\beta(t,o)\varphi(t,o) )</td>
<td>2140.84</td>
<td>914.36</td>
<td>9.12</td>
<td>0.05</td>
</tr>
<tr>
<td>( p(c,c)\beta(c,o)\varphi(c,c) )</td>
<td>2174.99</td>
<td>953.46</td>
<td>16.84</td>
<td>0.03</td>
</tr>
<tr>
<td>( p(c,c)\beta(c,o)\varphi(t,o) )</td>
<td>2217.52</td>
<td>957.75</td>
<td>21.13</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Table 1.5: Estimate of Kangaroo rat data under the super population capture-recapture model added with imperfect genetic data (\( p(c,c)\beta(t,o)\varphi(c,o) \)), with capture probability be the same for all animals, entry probability is dependent on time and origination of animal, and survival probability is the same over time but different for in situ birth and immigrant.

<table>
<thead>
<tr>
<th>Par.</th>
<th>Est.</th>
<th>S.E.</th>
<th>Par.</th>
<th>Est.</th>
<th>S.E.</th>
<th>Par.</th>
<th>Est.</th>
<th>S.E.</th>
<th>Par.</th>
<th>Est.</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p )</td>
<td>0.553</td>
<td>0.023</td>
<td>( \varphi^B )</td>
<td>0.400</td>
<td>0.006</td>
<td>( \varphi^I )</td>
<td>0.352</td>
<td>0.130</td>
<td>N</td>
<td>207.6</td>
<td>2.354</td>
</tr>
<tr>
<td>( \beta_0^B )</td>
<td>0.143</td>
<td>0.026</td>
<td>( \beta_1^B )</td>
<td>0.085</td>
<td>0.009</td>
<td>( \beta_2^B )</td>
<td>0.061</td>
<td>0.018</td>
<td>( \beta_3^B )</td>
<td>0.125</td>
<td>0.024</td>
</tr>
<tr>
<td>( \beta_0^I )</td>
<td>0.042</td>
<td>0.015</td>
<td>( \beta_1^I )</td>
<td>0.015</td>
<td>0.009</td>
<td>( \beta_2^I )</td>
<td>0.015</td>
<td>0.009</td>
<td>( \beta_3^I )</td>
<td>0.052</td>
<td>0.016</td>
</tr>
<tr>
<td>( \beta_4^B )</td>
<td>0.059</td>
<td>0.018</td>
<td>( \beta_5^B )</td>
<td>0.077</td>
<td>0.020</td>
<td>( \beta_6^B )</td>
<td>0.114</td>
<td>0.024</td>
<td>( \beta_7^B )</td>
<td>0.056</td>
<td>0.017</td>
</tr>
<tr>
<td>( \beta_4^I )</td>
<td>0.020</td>
<td>0.011</td>
<td>( \beta_5^I )</td>
<td>0.031</td>
<td>0.013</td>
<td>( \beta_6^I )</td>
<td>0.068</td>
<td>0.018</td>
<td>( \beta_7^I )</td>
<td>0.038</td>
<td>0.015</td>
</tr>
</tbody>
</table>
Bibliography


Chapter 2

A Super Population Capture-Recapture Model for Single Population with Two Age Groups Augmented with Age Information or Genetic Data

**Summary** Here we consider a super population capture-recapture model for single population with two age groups augmented with age information and genetic data. The relative contribution of *in situ* reproduction versus immigration to the recruitment process is important to ecologists. We first use age information to estimate the entry probabilities of *in situ* reproduction and immigrants separately for all except the first period. Then we combine age and genetic assignment test information with the capture-recapture model, which enables us to estimate the entry probability of *in situ* birth and the entry probability of immigrants separately for all sampling periods. Further, this augmentation of age specific capture-recapture data with genetic data greatly improves the precision of estimates of parameters. We apply our new model to a capture-recapture study with genetic information for banner-tailed kangaroo rats in Southern Arizona.
2.1 Introduction

The super population model (Crosbie and Manly, 1985, Schwarz and Arnason, 1996) is a representation of the open population model, Jolly-Seber model (Jolly 1965, Seber, 1982). In super population model, all the individuals exposed to sampling during the sampling periods are treated as a part of a super population \( N \). The animal of the super population are assumed to enter the super population via entry probability \( \beta_i \) in the intervals between sampling occasion \( i \) and \( i+1 \) (where \( i = 0, 2, \cdots, K - 1 \)). Let \( B_i \) be the numbers of recruits that enter the population between period \( i \) and \( i+1 \), \( i = 0, 2, \cdots, K - 1 \), the super population is all the recruits from all the sampling periods \( N = \sum_{i=0}^{K-1} B_i \), where \( B_0 \) is the number of recruits that enter the population before the time of the initial sampling period and are alive at the first sampling period.

Robust design is an important capture-recapture scheme. It was first proposed by Pollock (1982), which combines the properties of open population and closed population. In Robust design, the sampling periods is composed of two time scales, with \( l_i \) (where \( l_i \geq 2, i = 1, 2, \cdots, K \)) shorter secondary sampling periods within the longer primary sampling periods \( i \) (where \( i = 1, 2, \cdots, K \)). The population is assumed to be closed within each primary period (Kendall and Nichols, 1995, Kendall et al., 1997) and open between primary periods. That is, there can be losses (deaths, emigration) or gains (births, immigration) between primary periods. With the robust design, we can estimate some parameters (such as the first and last period capture probabilities and population sizes) which can not be estimated under the general Jolly-Seber model (Pollock, 1982, Pollock et al., 1990, Kendall and Pollock, 1992). Here all our capture-recapture data is assumed to include Robust design.

Understanding the relative contribution of in situ reproduction versus immigration to the recruitment process is important for ecologists. Nichols and Pollock (1990) used a two age class capture-recapture model with the robust design to estimate recruitment from immigration versus in situ reproduction. Manly et al. (2003) also improved population size estimates by incorporating age information of animals. Here we propose a totally different approach to estimate the relative contribution of in situ birth and immigrant to the population growth. All our new models are developed on the basis of super population model.

Considering an open population with two age groups, the animal of the popula-
tion can only be born on site or originate from outside the population. To analyze the
components of the new recruits of each period, we assume we can distinguish the in situ
birth from immigrant for all newly captured animals. Correspondingly, we split the en-
try probability into a new on site recruitment entry probability and an immigrant entry
probability. After including age information into the super population capture-recapture
model, we prove that it is possible to estimate these two entry probabilities for all except
the first period separately. Genetic assignment test can directly assign the origination of
the animal with high certainty based on the genetic data of the captured animals. We then
combine age information and the genetic assignment test information with the standard
capture-recapture robust super population model, which can estimate the entry probability
and survival probability for in situ birth and immigrant for all sampling periods.

Genetic assignment tests have recently been developed to determine an animal’s
population of origin. Individuals are genotyped at multiple neutral loci, and the likelihood of
their genotypes in potential source populations is estimated assuming that those populations
are in Hardy-Weinberg equilibrium (Waser and Strobeck, 1998). Paetkau et al. (1995) and
Favre et al. (1997) first suggested that individuals with unusual genotypes might come
from elsewhere, and also that individuals whose genotypes were a priori more likely in a
neighboring population might be immigrants from that population. Since then, statistical
methodologies that assign a population of origin to an individual based on its genotype
have multiplied (Guinand et al., 2002, Manel et al., 2005).

A “perfect” assignment test means that each animal captured can be assigned
without error either as an animal born in the population or an immigrant from outside
the population. That is, a “birth state” (immigrant or in situ birth) is assigned to any
captured animal where this state is otherwise unknown. Imperfect assignment test is a
practical case. In imperfect genetic assignment test, the assignment of the origination of an
animal is subject to uncertainty.

We shall begin with the assumption that the genetic assignment test is perfect (i.e.
without uncertainty) and later generalize to a model for imperfect genetic assignment tests
(i.e., the uncertainty in the assignment is accounted for). The fusion of genetic data and
capture-recapture data will allow us to determine the separate contributions of immigration
and in situ reproduction to the growth of the population. Another advantage of adding
genetic information is that the accuracy of the estimates will greatly be improved.
2.2 Model Structure

We consider a standard super population capture-recapture model where there is one population and two age groups, young (0-1 year) and old (>= 1 year). Young animals (0-1 year old) naturally become old animals (>= 1 year old) one year after birth. We assume that the time interval between successive primary periods is one year, the same time required to make a transition from young to old. There are \( K \) primary periods and each primary period has \( l_i \) shorter secondary periods. Considering the origination sources of the animal, the new recruits can only come from born on site and the immigrant from outside.

2.2.1 Notation

Of all the following notations, \( r = Y \) (Young) or \( r = O \) (Old).

\( p_{ij}^r \): the secondary period capture probability which means the probability that an animal of age \( r \) is captured on secondary period \( j \) of primary period \( i \), \( q_{ij} = 1 - p_{ij} \).

\( p_{i}^{*r} \): the primary period capture probability which means the probability that an animal of age \( r \) is caught at least once in the \( l_i \) secondary periods of primary period \( i \), \( p_{i}^{*r} = 1 - \prod_{j=1}^{l_i} (1 - p_{ij}^r) \).

\( \varphi_i^r \): the probability the animal surviving from primary period \( i \) to primary period \( i + 1 \) given that the animal is of age \( r \) at primary period \( i \).

\( B_i^r \): the new recruits (in situ births or immigrants) of age \( r \) to the population between primary period \( i - 1 \) and primary period \( i \). \( B_0^r = N_1^r \) is the number of animals of age \( r \) in the population at the initial sampling period.

\( u_i^r \): the number of unmarked animals of age \( r \) caught on at least one secondary occasion within primary period \( i \). The newly captured unmarked young animals \( u_i^Y \) can only be the animals enter the super population at primary period \( i \) at age \( Y \). While, the newly captured unmarked adult animals \( u_i^O \) can be

- animals enter the super population before period \( i \) at either young age or old age, survive to period \( i \), and never be captured by period \( i \).
- animals enter the super population at period \( i \) at old age.
\( \beta_r^i \): the probability of an animal of the super population entering the super population at age \( r \) between primary period \( i \) and \( i+1 \), and surviving until sampling period \( i+1 \). \( \beta_0^r \) is the probability that a member of the super-population came into the population before the study started and survived to primary period 1 as age \( r \).

\( m_{ij}^r \): the number of adult animals caught in primary period \( j \) that were last caught in primary period \( i \) \((1 \leq i < j \leq K)\) given that the animals are of age \( r \) at primary period \( i \).

\( R_i^r \): the number of marked animals of age \( r \) that are released back into the population following primary period \( i \).

\( \chi_{i}^{\omega r} \): the number of animals of age \( r \) having capture history \( \omega \) over the \( l_i \) secondary periods of primary period \( i \) given that the animals are part of \( R_i \).

Assuming no tag mortality and all captured animals are released back to the population, we have \( R_i^Y = u_i^Y \), and \( R_i^O = u_i^O + \sum_{r=0}^{r=Y,O} \sum_{j=1}^{i-1} m_{j,i}^r \). If we consider the origination of the animal, we can decompose \( u_i^r = u_i^{rB} + u_i^{rI} \), \( \beta_i^r = \beta_i^{rB} + \beta_i^{rI} \), \( R_i^r = R_i^{rB} + R_i^{rI} \), and \( m_{ij}^r = m_{ij}^{rB} + m_{ij}^{rI} \), where

\( u_i^{rB} (u_i^{rI}) \): the number of newly caught animals of age \( r \) at primary period \( i \) that are \textit{in situ} births (immigrants).

\( \beta_i^{rB} (\beta_i^{rI}) \): the probability of the animal of age \( r \) of the super population entering the super population between sampling period \( i \) and \( i+1 \), and being a local birth (an immigrant).

\( R_i^{rB} (R_i^{rI}) \): the number of animals of age \( r \) released to the population in primary period \( i \) given that these animals are \textit{in situ} births (immigrants).

\( m_{ij}^{rB} (m_{ij}^{rI}) \): similar to \( m_{ij}^r \) except that the animals originate as new born of the population (immigrants from outside the population).

### 2.2.2 Standard Super Population Capture-Recapture Model for Single Population with Two Age Classes

In the standard super population capture-recapture model, the likelihood for this model can be write as

\[
L = L_1 \ast L_2 \ast L_3 \tag{2.1}
\]
where \( L_1 \) deals with the capture of unmarked animals of the primary periods (the super population part), \( L_2 \) is the probability distribution of the recapture data of primary periods conditional on the releases \( R_i \), and \( L_3 \) is the likelihood corresponding to the secondary period captures when using the Robust design.

Now we learn how the age information of the animal be used to estimate the relative contribution of the in situ birth and immigrant to the growth of the population. First, we split the entry probability \( \beta_i \) into \( \beta^Y_i \) and \( \beta^O_i \) according to the age of the animals when they enter the super-population. For any animal, it originates either from \textit{in situ} reproduction or from immigration. We assume the animal does not move early in its first year which is reasonable based on our experience. So if the animal is an \textit{in situ} birth, the animal enters the super-population as soon as it is born. If the animal is an immigrant, it will not enter the super-population until it survives to become an adult and moves to the super-population. When comes to the entry probability, it means that \( \beta^Y_i = \beta^Y_i, i = 0, 1, \cdots, K \) and \( \beta^O_i = \beta^O_i, i = 1, \cdots, K - 1 \). However, for the first sampling period of old animals, we have a more complex expression where \( \beta^O_0 = \beta^O_0 + \beta^O_0 \). This is because unmarked adults at the time the study starts will be composed of animals that were born on site and animals that were immigrants by the initial sampling period \( i \).

As a general case, we assume \( \{u^Y_1, u^Y_2, \cdots, u^Y_K, u^O_1, \cdots, u^O_K\} \) forms a multinomial distribution conditional on the super population size \( N \). The super population and the new estimable entry parameters are \( \{N, \beta^Y_0, \beta^Y_1, \cdots, \beta^Y_K, \beta^O_0, \beta^O_1, \cdots, \beta^O_{K-1}\} \). Notice that we do not decompose \( \beta^O_0 \) because the components turn out not to be estimable. Since all newly captured young animals are captured at their natal place, we have \( E(u^Y_i) = N \times \beta^Y_i \times p^Y_i, i = 1, \cdots, K \). However for old animals, the situation becomes complex because the captured old animals are composed of two groups, one group is the animals entering into the super population as young and not captured on entering but captured later at an older age. The other group is the animals entering the super population at an older age.
For the newly captured old animals entering the super-population at young age:

\[ \psi^Y_1 = 0 \]
\[ \psi^Y_2 = \beta_0^Y B \psi^Y_{i-1} + \beta_i^Y B \psi^Y_{i-2} \]
\[ \psi^Y_i = \psi^Y_{i-1} * \phi^Y_{i-1} + \beta_i^Y B \psi^Y_{i-2} * \phi^Y_{i-1} \]
\[ i = 3, 4, \cdots, K. \] 

For the newly captured old animals entering the super-population at old age:

\[ \psi^O_1 = \beta_0^O I \]
\[ \psi^O_i = \psi^O_{i-1} * \phi^O_{i-1} + \beta_i^O I \]
\[ i = 3, 4, \cdots, K. \] 

Let \( Pu^Y_i = \beta_i^Y B \psi^Y_i, Pu^O_i = (\psi^Y_i + \psi^O_i) * \phi^O_i, u. = \sum_{r,j} u^r_j, \text{ and } p. = \sum_{r,i} Pu^r_i, \) we have

\[ L_1(N, \beta_0^Y B, \beta_0^O, \beta_i^O I, \phi^Y_i | u^r_i) = \frac{N!}{u!(N - u)!} p^u (1 - p)^{N-u} \]
\[ \prod_{r=1}^{K} \prod_{i=1}^{Y,O} \frac{u^r_i!}{\left( u^r_i - \sum_{j=i+1}^{K} m^r_{ij} \right)!} \prod_{j=i+1}^{K} \left( \phi^r_{i,j} \psi^O_j \right) m^r_{ij} \left( \phi^r_i \right)^{R^r_i - m^r_i} \] 

Similarly, for recaptured animals of primary period \( L_2 \), define new parameters \( \phi^r_i \),

\[ \phi^r_i = \psi^O_{i} \] 
\[ \phi^r_{i,j} = \psi^O_{i} * (1 - p^O_j) * \psi^O_j, \] 
\[ j = \min(i + 2, K), \cdots, K \] 

Let \( m^r_i = \sum_{j=i+1}^{K} m^r_{ij} \) and \( \phi^r_i = 1 - \sum_{j=i+1}^{K} \phi^r_{i,j} * p^O_j \). Conditional on the releases of primary period \( i (R^r_i) \), all newly recaptured animals \( \{m^r_{i+1}, m^r_{i+2}, \cdots, m^r_{i,K}, m^r_i\} \) of the following sampling periods are assumed to follow a multinomial distribution. We also assume the animals released in one primary period are independent from the animals released in another different primary period. The likelihood of \( L_2 \) will be the product of the multinomial distributions conditional on \( R^r_1, R^r_2, \cdots, R^r_{K-1} \). Then we have the likelihood for the recapture part of the primary period is

\[ L_2(\phi^r_i, p^O_i | R^r_i, m^r_{ij}) = \prod_{r=Y,O}^{K-1} \prod_{i=1}^{K-1} \frac{R^r_i!}{m^r_{ij}!(R^r_i - m^r_i)!} \]
\[ \times \prod_{j=i+1}^{K-1} \left( \phi^r_{i,j} p^O_j \right)^{m^r_{ij}} \left( \phi^r_i \right)^{R^r_i - m^r_i} \]
Within every primary period, the population is closed. The animals can be captured at any of the secondary periods. It is reasonable to assume the capture-recapture animals is a multinomial distribution. Here for ease of presentation, we assume that there are two secondary periods within each primary period, but other numbers of secondary periods are also easily modeled.

\[
L_3(p_{ij}^r | x_1^r, R_i^r) = \prod_{r=Y, O} \prod_{i=1}^{K} \frac{R_i^r}{\chi_r^{101} \chi_r^{01} \chi_r^{11}} \cdot \left( p_{i1}^r q_{i2}^r \right)^{\chi_r^{10}} \cdot \left( \frac{q_{i1}^r p_{i2}^r}{p_{i1}^r} \right)^{\chi_r^{01}} \cdot \left( \frac{q_{i1}^r q_{i2}^r}{p_{i1}^r} \right)^{\chi_r^{11}} \tag{2.7}
\]

From the fact that \( \beta_i^Y = \beta_i^{YB} \) and \( \beta_i^O = \beta_i^{OI} \) (\( i \neq 0 \)), we conclude that for a single population with two age groups, after including age information, we can estimate nearly all \( \beta_i^{YB}, \beta_i^{OI} \) \( (i \neq 0) \). However, for the old animals of the initial sampling period, we can only estimate \( \beta_0^O \) and can not estimate \( \beta_0^{OB} \) and \( \beta_0^{OI} \) separately. This is because \( \beta_0^{OB} \) and \( \beta_0^{OI} \) always appear together in the term of \( \psi_0^O \) \( (\psi_0^O = \beta_0^{OB} + \beta_0^{OI}) \) in the likelihood. Further for \( u_0^O \), we cannot distinguish \( u_0^{OB} \) from \( u_0^{OI} \) based on our classical capture-recapture history. We also can not estimate the survival probability of in situ birth and immigrant, \( \phi_i^{OB} \) and \( \phi_i^{OI} \), separately because we can not split the newly captured unmarked animals into in situ birth and immigrant. The approach of the next section will solve these problems.

### 2.2.3 Super Population Capture-Recapture Model for a Single Population with Two Age Groups Augmented with Perfect Genetic Assignment Test

Now we consider incorporating additional genetic information into the super population model with one population and two age classes to aid in estimation of the separate contributions of in situ reproduction and immigration to the growth of the population. Based on the perfect genetic assignment information, we can split newly captured unmarked adult animals \( u_i^O = u_i^{OB} + u_i^{OI} \) according to their assigned populations of origin. Correspondingly, we can decompose the entry probability according to population of origin as \( \beta_i^O = \beta_i^{OB} + \beta_i^{OI} \). For young animals, we assume they are captured at their natal place with \( u_i^Y = u_i^{YB} \) and \( \beta_i^Y = \beta_i^{YB} \). Without generality, we can assume the \textit{in situ} birth and immigrant have different survival probability. The likelihood of the super population capture-recapture model integrated with perfect genetic assignment test now becomes

\[
L = L_1^* \cdot L_2^* \cdot L_3
\]
where \( L^*_1 \) is the likelihood for the newly captured in situ birth and immigrants incorporating information on population of origin, \( L^*_2 \) is the modified likelihood for the recapture component of primary periods conditional on the releases \( R^*_i \), \( R^*_i^{OB} \), and \( R^*_i^{OI} \), and \( L^*_3 \) is the likelihood for the secondary period component. \( L^*_3 \) is the same as the standard super population capture-recapture model if we assume the capture probability is the same for immigrant and in situ birth which is a reasonable assumption in practice.

Conditional on the super population \( N \), the newly captured unmarked in situ births and immigrants

\[
(\psi_{YB_1}, \psi_{YB_2}, \ldots, \psi_{YB_K}, \psi_{OB_1}, \psi_{OB_2}, \ldots, \psi_{OB_K}, \psi_{OI_1}, \psi_{OI_2}, \ldots, \psi_{OI_K}, N - \sum (\psi_{YB_i} + \psi_{OB_i} + \psi_{OI_i}))
\]

are assumed to follow a multinomial distribution.

Similar as the standard super population capture-recapture model, for newly captured young animals:

\[
\begin{align*}
\psi_{YB_i} &= \beta_{YB_i} \\
E(u_{YB_i}) &= N \psi_{YB_i} \cdot p^Y_i
\end{align*}
\]

The newly captured adult animals coming from on site born \( (\psi_{OB_i}) \) are composed of two parts. The first part is the animal enter the super population at young age before primary period \( i \), never captured by primary period \( i \), surviving and being captured at primary period \( i \) at older age.

\[
\begin{align*}
\psi_{OB_1} &= 0; \\
\psi_{OB_2} &= \beta_{OB_2} \cdot \phi_{OB_1}^{OB_2} \\
\psi_{OB_\nu} &= \psi_{OB_{\nu-1}} \cdot \phi_{OB_{\nu-1}}^{OB_\nu} + \beta_{OB_{\nu-2}} \phi_{OB_{\nu-1}}^{OB_\nu}
\end{align*}
\]

The second part is the animals enter the super population at the initial period 1 as adult animals, surviving and being captured at primary period \( i \) at older age.

\[
\begin{align*}
\psi_{OBO_1} &= \beta_{OBO_1} \\
\psi_{OBO_\nu} &= \psi_{OBO_{\nu-1}} \cdot \phi_{OBO_{\nu-1}}^{OBO_\nu}
\end{align*}
\]

Let \( \psi_{OB} = \psi_{OB_1} + \psi_{OBO_1} \), we have \( E(u_{OB_i}) = N \psi_{OB} \cdot p^O_i \). We also have

\[
E(u_{OI_i}) = N \psi_{OI_i} \cdot p^O_i
\]

for newly captured old animals originating from outside,
where

\[ \psi_{i'}^{O} = \beta_{0}^{O}, \]
\[ \psi_{i'}^{I} = \psi_{i'-1}^{O} \psi_{i'}^{I} + \beta_{i}^{I} \]

(2.12)

where \( i' = 2, \ldots, K, i = 1, \ldots, K \). Let \( u = \sum_{i=1}^{K} \left( u_{i}^{YB} + u_{i}^{OB} + u_{i}^{OI} \right) \), and

\[ p = \sum_{i=1}^{K} \psi_{i}^{YB} p_{i}^{Y} + \psi_{i}^{OB} p_{i}^{O} + \psi_{i}^{OI} p_{i}^{I}, \]

We have

\[ L_{1}^{*} (N, \beta_{1}^{YB}, \beta_{i}^{OB}, \beta_{i}^{OI}, p_{i}^{Y}, p_{i}^{O}, p_{i}^{I}, \varphi_{i}^{YB}, \varphi_{i}^{OB}, \varphi_{i}^{OI}, u_{i}^{YB}, u_{i}^{OB}, u_{i}^{OI}) = \frac{N!}{u!(N-u)!} \left( p_{i}^{Y} \right)^{u_{i}^{YB}} \left( p_{i}^{O} \right)^{u_{i}^{OB}} \left( p_{i}^{I} \right)^{u_{i}^{OI}} \prod_{i=1}^{K} \left( \psi_{i}^{YB} p_{i}^{Y} + \psi_{i}^{OB} p_{i}^{O} + \psi_{i}^{OI} p_{i}^{I} \right)^{u_{i}} \right. \]

\[ \left. \prod_{i=1}^{K} \left( \psi_{i}^{YB} p_{i}^{Y} + \psi_{i}^{OB} p_{i}^{O} + \psi_{i}^{OI} p_{i}^{I} \right)^{u_{i}} \right) \]

(2.13)

For the recapture part of primary periods added with genetic information \( (L_{2}^{*}) \), For animals originate as in situ birth:

\[ \phi_{i,i+1}^{B} = \varphi_{i}^{YB}, \quad i = 1, 2, \ldots, K, \quad r = Y, O \]
\[ \phi_{i,j}^{B} = \phi_{i,j}^{B} * (1 - p_{j}^{O}) * \varphi_{j}^{OB}, \quad j = i + 2, \ldots, K \]

(2.14)

For animals originate from outside:

\[ \phi_{i,i+1}^{O} = \varphi_{i}^{OI}, \quad i = 1, 2, \ldots, K, \]
\[ \phi_{i,j}^{O} = \phi_{i,j-1}^{O} * (1 - p_{j}^{O}) * \varphi_{j}^{OB}, \quad j = i + 2, \ldots, K \]

(2.15)

Let \( m_{i}^{B} = \sum_{j=i+1}^{K} m_{ij}^{B}, m_{i}^{O} = \sum_{j=i+1}^{K} m_{ij}^{O}, \phi_{i}^{B} = 1 - \sum_{j=i+1}^{K} \phi_{i,j}^{B} * p_{j}^{O}, \) and \( \phi_{i}^{O} = 1 - \sum_{j=i+1}^{K} \phi_{i,j}^{O} * p_{j}^{O} \), we have

\[ L_{2}^{*} \left( \varphi_{i}^{YB}, \varphi_{i}^{OB}, \varphi_{i}^{OI}, p_{i}^{Y}, p_{i}^{O}, p_{i}^{I}, R_{i}^{YB}, R_{i}^{OB}, R_{i}^{OI}, m_{ij}^{YB}, m_{ij}^{OB}, m_{ij}^{OI} \right) = \prod_{i=1}^{K-1} \prod_{r=1,1}^{R_{i}^{B1}} \prod_{j=i+1}^{K} (\phi_{i,j}^{B} p_{j}^{O})^{m_{ij}^{B}} * (\phi_{i,j}^{B} p_{j}^{B} - m_{ij}^{B}) \]

\[ \prod_{j=i+1}^{K} (\phi_{i,j}^{O} p_{j}^{O}) m_{ij}^{O} * (\phi_{i,j}^{O} R_{i}^{O} - m_{ij}^{O}) \]

(2.16)

The adding of perfect genetic assignment data enables the \( \beta_{0}^{O}, \beta_{0}^{I}, \varphi_{i}^{OB} \) and \( \varphi_{i}^{OI} \) to be estimated separately, and this cannot be achieved using the capture-recapture data alone. Perfect genetic data allow us to distinguish the in situ reproduction and immigration. We
can also construct a more general model by assuming that the immigrants have different survival probability from that of in situ birth animals. We can also assume the capture probabilities are different for immigrants and animals resulting from in situ reproduction, but we do not develop that generalization.

2.2.4 Super Population Capture-Recapture Model for a Single Population with Two Age Groups Augmented with Imperfect Genetic Assignment Test

As in the case where we had a perfect genetic assignment test, any animal captured for which the birth state is unknown has a genetic assignment test carried out so that a state of origin is assigned. However, now this genetic assignment information is subject to uncertainty, which is what occurs in practice. Consider a new adult (at least one year old) animal genotyped at capture in period 2. The assignment test assigns the animal as an \textit{in situ} birth with probability \( \pi \) (0 < \( \pi < 1 \)). We view \( \pi \) as the probability that the animal is an \textit{in situ} birth in the population, and 1 - \( \pi \) as the probability that it is an immigrant. If the assignment test is a good one then when the animal is truly an \textit{in situ} birth, \( \pi \) will be close to one and (1 - \( \pi \)) will be close to zero; whereas when the animal is truly an immigrant, \( \pi \) will be close to zero and (1 - \( \pi \)) will be close to one. Perfect assignment test is a special case of imperfect assignment test when \( \pi = 1 \) if the animal is truly an \textit{in situ} birth and \( \pi = 0 \) if the animal is truly an immigrant. Further, good assignment test information will be valuable and should improve our estimators of recruitment parameters.

Now we will generalize our model to the more realistic case of combining capture-recapture data with imperfect genetic assignment test data. Our objective is to utilize the existing model structure and computer code of the super population capture-recapture model with perfect genetic assignment tests. We will take a resampling approach similar to multiple imputation (Rubin’s, 1987) to impute the perfect genetic assignment test data based on the imperfect genetic information. We assume the young animals are captured at their natal places, so only old animals need to have their origination assigned genetically. Consider that each newly captured adult animal, \( j \) (\( j = 1, \cdots, \sum u_i^O \)), has a genetic assignment probability \( \pi_j, 0 < \pi_j < 1 \). We will impute a pseudo perfect assignment test data set \( m \) times using the following procedures. First, each newly captured adult animal will be randomly allocated as a uniform random variable \( U_j \in (0, 1) \). If \( U_j \in (0, \pi_j) \), we assign this
animal as an *in situ* birth, whereas if \( U_j \in (\pi_j, 1) \), we assign this animal as an immigrant. Apply this assignment procedure on all newly captured adult animals of all primary sampling periods. Then we will split \( u^O_i \) into \( u^{OB}_i \) and \( u^{OI}_i \) as in the perfect assignment test situation and get a “pseudo perfect assignment test data” set. We repeat the imputation of this kind of pseudo perfect assignment test data \( m \) times. For each of the \( m \) “pseudo perfect assignment test data”, we treat it as super population capture-recapture model with perfect genetic assignment test to estimate the entry probability, survival probability, and capture probability for *in situ* births and immigrants. Then we can take various distributional statistics (means, medians, standard deviation, and percentiles) of the estimators. Therefore, we obtain estimators that reflect the inherent uncertainty due to the genetic assignment test data being imperfect, in addition to the usual uncertainty associated with capture-recapture inference. So the likelihood will the be same as that of super population capture-recapture model added with perfect genetic assignment test.

There is uncertainty in the pseudo perfect data imputed by our resampling approach, which is caused by the assigning process. The pseudo perfect genetic data changes from one imputation to another imputation. So for the variance of the estimates, it is composed of within-imputation variance and between imputation variance, which is the same as that of the multiple imputation. For example, for the variance of the estimate of \( \beta^{OB}_0 \), it is calculated as follows.

\[
\hat{\text{var}}(\beta^{OB}_0) = \frac{\sum_{j=1}^{m} \sigma_j}{m} + \left(1 + \frac{1}{m}\right) \times \frac{1}{m - 1} \sum_{j=1}^{m} ((\beta^{OB}_0) - (\bar{\beta}^{OB}_0))^2
\]  

(2.18)

where the first term is the naturally variability inherent to the data and is calculated by the information matrix. \( \sigma_j \) is the variance of \( \beta^{B}_0 \) as the generated data is the true data (no imputation needed). The second term is the between imputation variance.

### 2.2.5 Assumptions

The super population capture-recapture model with the Robust Design and genetic assignment data is a generalization of the standard super population model, and therefore the assumptions required are similar.

1. All members of the super population \( N \) entering the population at the same time from the same origination source and of the same age have the same entry probability (homogeneous entry probabilities).
2. Every animal of the same age and same origination has the same probability of surviving from one period to the next period (homogeneous survival probabilities).

3. Each marked animal of same age presenting at sampling period $i$ has the same probability of capture at sampling period $i$ (homogeneous capture probabilities).

4. Marks do not affect the survival or behavior of the animal, are not lost, and are recorded correctly.

5. Each animal behaves independently with respect to survival, entry, and detection probabilities.

6. A valid genetic assignment test has been carried out for all animals where this origination is unknown. This implies that there are no errors in the genotyping used in the tests (however, there will be uncertainty in the genetic assignments, regardless).

### 2.3 Simulation

To generate the required imperfect assignment probabilities for all newly captured adult animals for simulation purpose, we use the following procedure. First, we generate a data set $u_i^{OB}$ and $u_i^{OI}$ based on the true parameter values, $\beta_i^{OB}, \beta_i^{OI}, \phi_i^r, \varphi_i^r, \varphi_i^O, r = Y,O$. Then for the $j$th animal belonging to $u_i^{OB}$ we randomly assign it a high genetic assignment probability, for example $\pi_j \in U(0.95, 1), (j = 1, \cdots, u_i^{OB})$, and $1 - \pi_j$ will be the genetic assignment probability the animal being from outside. Similarly for the $j$th animal belonging to $u_i^{OI}$ we randomly assign it a small genetic assignment probability, for example, $\pi_j \in U(0, 0.05), (j = 1, \cdots, u_i^{OI})$, and $1 - \pi_j$ will be the genetic assignment probability the animal being from outside.

In the simulations, we tried four different uncertain level of imperfect genetic assignment tests, the first as the perfect case assumes the genetic assignment probability to be one (zero) given the animal is in situ birth (immigrant); the second assumes the genetic assignment probability is uniform between $(0.95, 1)$ given the adult animal is in situ birth and the genetic assignment probability is uniform between $(0, 0.05)$ given the adult animal is immigrant; the third assumes the genetic assignment probability is uniform between $(0.9, 1)$ given the adult animal is in situ birth and the genetic assignment probability is uniform
between (0, 0.1) given the adult animal is immigrant; and the fourth assumes the genetic assignment probability is uniform between (0.8, 1) given the adult animal is *in situ* birth and the genetic assignment probability is uniform between (0, 0.2) given the adult animal is immigrant.

In the simulation, the sampling periods lasted five years with three secondary periods within each primary period. To reduce the number of parameters, we assume that all animals of the same age having the same captured probability within each year. This is a reasonable assumption based on experience. However, the entry probability, the primary capture probability and the survival probability are assumed to be dependent on the time. The entry probability and survival probability are also assumed to be different for *in situ* birth and immigrant. The parameters take different values with secondary capture probability taking values 0.5 and 0.2, survival probabilities are 0.3 and 0.15 for young animals and 0.35 and 0.55 for old animals, which reflects that young animals usually have lower survival probabilities, and super population size taking values 1000 and 10000. For the entry probabilities, the first sampling period always has more new recruits than later sampling periods, and there are always more *in situ* births than immigrants. This latter scenario is reasonable in practice because most animals stay in the same population throughout their lives. According to a complete factorization design, our simulations were done under eight different scenarios with parameters having any values listed above. Here we only report part of the results which can represent all simulation results.

First, we want to test the performance of the genetic data on the super population capture-recapture model. We compare the simulation results for the super population capture-recapture model with perfect genetic assignment data having $\pi \in (0.95, 1)$ (“Perf model”) and the capture-recapture model without genetic data (“Standard model”). Table 2.1 reports part of the simulation results from 1000 repetitions under four different situations. From Table 2.1, we can see the perfect model has lower RRMSE then the standard model for most of the parameters no matter what true values the parameters take. When the parameter take reasonable values like $N = 10000$, $\phi_{YB}^i = 0.3$, $\phi_{OB}^i = 0.35$, $\phi_{OI}^i = 0.35$, the estimators have very good precision and accuracy with maximum RRMSEs being 11.65%. Even when the parameters take very low value with $N = 1000$, $\phi_{YB}^i = 0.15$, $\phi_{OB}^i = 0.55$, $\phi_{OI}^i = 0.55$, 62% RRMSEs are lower than 23%.

One big advantage of adding genetic information is, the perfect model can estimate
Figure 2.1: Relative bias of 1000 simulations for the super population capture-recapture model integrated with different genetic assignment tests. $p^Y_{ij} = 0.3, p^O_{ij} = 0.27, \varphi_i^Y = 0.8, \varphi_i^O = 0.8, \beta_0^Y = 0.2, \beta_0^O = 0.2, \beta_0^{OB} = 0.2, \beta_0^{OI} = 0.1, \beta_i^{OI} = 0.05, , N = 1000$

$\beta^B, \beta^I, \varphi_i^{OB}, \varphi_i^{OI}$ separately whereas this can not be achieved by the standard model without genetic information.

Now we test how the uncertainty of the genetic assignment test affect the estimators of the model. Here we only report the simulation results for parameter values $p^Y_{ij} = 0.3, p^O_{ij} = 0.27, \varphi_i^Y = 0.8, \varphi_i^O = 0.8, \beta_0^Y = 0.2, \beta_0^O = 0.2, \beta_0^{OB} = 0.2, \beta_0^{OI} = 0.1, \beta_i^{OI} = 0.05, , N = 1000$ as in Figure 2.1, Figure 2.2, and Figure 2.3.

Figures 2.1 shows a decreasing relative bias following a increasing reliable genetic assignment test. There is no such trend in the relative S.E. (Figures 2.2). However, there is also decreasing relative RMSE with increasing certain level of genetic assignment
Figure 2.2: Relative sample standard error of 1000 simulations for the super population capture-recapture model integrated with different genetic assignment tests. $p_{ij}^Y = 0.3, p_{ij}^O = 0.27, \varphi_i^Y = 0.8, \varphi_i^O = 0.8, \beta_0^{YB} = 0.2, \beta_i^{YB} = 0.11, \beta_0^{OB} = 0.2, \beta_i^{OB} = 0.1, \beta_i^{OI} = 0.05, N = 1000$
Figure 2.3: Relative RMSE of 1000 simulations for the super population capture-recapture model integrated with different genetic assignment tests. 

\[
    p_{ij} = 0.3, p_{ij}^O = 0.27, \varphi_i^Y = 0.8, \varphi_i^O = 0.8, \beta_i^{YB} = 0.2, \beta_i^{YB} = 0.11, \beta_0^{OB} = 0.2, \beta_0^{OI} = 0.1, \beta_i^{OI} = 0.05, , N = 1000
\]
Figure 2.4: Relative RMSE for the simulation results of the super population capture-recapture model integrated with four different uncertainty levels of genetic assignment tests ($\pi \in (0.95, 1)$, $\pi \in (0.9, 1)$ and $\pi \in (0.8, 1)$). Parameters take a complete combinations of $p_{ij}^Y = 0.5$, $p_{ij}^O = 0.2$, $\varphi_i^Y = 0.3$, $\varphi_i^O = 0.35$, $\beta_{ij}^{YB} = 0.2$, $\beta_{ij}^{OB} = 0.11$, $\beta_{0j}^{OB} = 0.2$, $\beta_{0j}^{OI} = 0.1$, $\beta_{ij}^{OI} = 0.05$.

test(Figures 2.4), which means that the change of RRMSE is mainly due to the changing of the Relative bias. From Figures 2.4, we can see, no matter what the true values of the parameters, the RRMSEs of the capture probabilities and survival probabilities for the eight different cases are close to each other. When super population size is 10000, nearly 99% of the RRMSEs are less than 15%. When super population size 1000, the estimates are still reasonably good for all cases no matter what value the parameter takes. In this case, most RRMSEs are less than 25% and nearly half of the RRMSEs are smaller than 20%. And usually, for the estimates of capture probability and survival probability, the models with perfect genetic assignment data have smaller RRMSEs than the models with imperfect genetic assignment test data. However, the higher the genetic assignment probability, the better the estimates (with smaller RRMSE). In this case where the super population size is small at 1000, capture probability with low value of 0.2, survival probability is as low as 0.15, and some entry probabilities are just 0.06, the estimates are still reasonable. In this case, capture probability and survival probability have better estimates than entry probability due to the lower true values of the entry probabilities. Usually, the higher the
true value of the parameters, the smaller RRMSE of the estimates.

2.4 Example: Kangaroo Rats in Southern Arizona

We use data from a capture-recapture study with genetic assignment test information added for banner-tailed kangaroo rats, carried out in the San Simon and San Bernardino valleys in extreme SE Arizona by Waser and Dewoody of Purdue University. In their study, they sampled one population every year from 1994-2001 with an average of 3 consecutive nights within each year using a capture-recapture method. During each primary period, they trapped all active mounds in each population on three secondary periods. For each trapped animal, they marked the animal and recorded the trapping location and age (“juvenile” for animal less than one year old, “adult” for animals older than one yea). In parallel with their demographic data, they collected the genotype information of each newly captured animals. From their study, they found that most juveniles were still at their natal mounds. This shows that our assumption that young animals are captured at their natal place is a reasonable one for this study. They also noticed that the local populations were linked by occasional dispersal. Here we focus on the data from 1994 to 2001 and consider a model with two age groups and a single population.

We used GENECLASS2 (Piry et al. 2004) to analyze the genotype information to estimate the genetic assignment probabilities. For each adult animal with genetic information collected, we calculate its specific genetic assignment probability by GENECLASS2. The genetic assignment test implemented by GENECLASS2 is not perfect. We also handled the data as though they were “perfect” data although this is not realistic.

Table 2.2 and Table 4.1 list out all the estimators and their estimate standard errors. Figure 2.5 focuses on the dynamic property of the entry probability and survival probability. From Table 2.2 and Table 4.1, we can see that the estimates for all the parameters of the two different genetic assignment tests are close to each other. As we expected, the model with genetic information can estimate $\beta^B_i$ and $\beta^I_i$ separately. It can also estimate $\varphi^{OB}_i$ and $\varphi^{OI}_i$ separately as in simulation which we did not do this here. Figure 2.5 shows that under the standard model, the young animals having lower survival rate than the old animals except for the last year. Under imperfect genetic model, the young animals also have lower or similar survival rate as the old animals except for the fourth and the
last year. For entry probability, there are more \textit{in situ} births than the immigrants for all sampling periods under both models with genetic data or not. This confirms the fact that only a small part of the population is coming from outside. Under both imperfect genetic model and standard model, the entry probability for old animals of initial sampling period is much higher than those of later periods. This should be what we expect because for the initial sampling, the new recruits are all the animals entering the population before the first sampling period and surviving to the first sampling period.

2.5 Discussion

In this paper, we considered a single population with two age groups super population model augmented with both perfect and imperfect genetic assignment data. In some species, more than two age classes can be specified. We can extend the model to a single population with more than two age groups in our future research.

For the kangaroo data analysis, we only constructed a full model where all parameters are different from each year. To improve the precision of the estimates, we can reduce the number of parameters by adding constraints to the parameters. Then we can
use AIC or BIC to choose a best model for the data set. The young kangaroo rats rarely move out of their natal mound at their first year “only once in several hundred trap nights (Waser, Personal communication).” To test the validation of the genetic assignment test, we assign the genetic assignment probability to a sample of young animals which are all in situ births. If the genetic assignment test is a good one, the genetic assignment probability should be very high, like bigger than 0.95, as the bigger the better the genetic assignment test. However, when we use GENECLASS2 to analyze the genetic information of the sample of young animals, we found that the assignment methods “generally result in an excess of resident individuals being excluded”, i.e., scored as immigrants. Waser (personal communication) suspects that the biological reason for this is that immigrants tend to have alleles that are common in their birth population but locally rare in the population they move to. However they pass some of those alleles to their offspring, and their offsprings progeny, etc who then may be incorrectly judged to be immigrants too. Thus the assignment procedure is biased towards “immigration”. For example, when we run GENECLASS with data from the population we are interested in vs. everything else, it correctly assigns about 70% of the animals first captured in population to population. A few percent may really come from outside, but the true number is surely far below 30%, as many of these individuals were first captured as small juveniles and/or have known parents in population. Waser suggested one way to deal with this problem might be to use assignment test results only when it is relatively unambiguous, for example for individuals that are >10x as likely to come from one population than the other(s), and to treat the other individuals as having no genetic information. We hope to explore this and other ideas in future work. Based on genetic parentage determination rather than assignment tests, Waser (personal communication) found that for animals whose parents can be determined, only 7% live as adults in a different population than their parents. So in our real data analysis here, there is only a very small proportion of tagged animals (at most a few percent) are estimated to disperse between populations (Skvarla and Nichols, 2004). Therefore the entry probabilities from outside should be really small. However, in the simulation part, we assume that the entry probabilities of the immigrants are reasonably large to achieve reasonable precision.
Table 2.1: Compare the simulation results of the super population capture-recapture model with perfect genetic assignment data ("P M" columns) and the standard super population capture-recapture model ("S M" columns). The simulation results are under $\beta^Y_0 = 0.06, \beta^O_0 = 0.2, \beta^I_0 = 0.1, \beta^Y_i = 0.05, \beta^I_i = 0.11$ and other parameter values are listed in the table.

<table>
<thead>
<tr>
<th>Par.</th>
<th>$p^Y_{ij} = 0.5, p^O_{ij} = 0.2, \varphi^Y_i = 0.3, \varphi^O_i = 0.35, N = 1000$</th>
<th>$p^Y_{ij} = 0.5, p^O_{ij} = 0.2, \varphi^Y_i = 0.15, \varphi^O_i = 0.55, N = 1000$</th>
<th>$p^Y_{ij} = 0.5, p^O_{ij} = 0.2, \varphi^Y_i = 0.3, \varphi^O_i = 0.35, N = 10000$</th>
<th>$p^Y_{ij} = 0.5, p^O_{ij} = 0.2, \varphi^Y_i = 0.15, \varphi^O_i = 0.55, N = 10000$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta^Y_0$</td>
<td>14.65</td>
<td>15.10</td>
<td>$\beta^O_0$</td>
<td>9.30</td>
</tr>
<tr>
<td>$\beta^Y_0$</td>
<td>14.13</td>
<td>14.95</td>
<td>$\beta^O_0$</td>
<td>14.46</td>
</tr>
<tr>
<td>$\beta^Y_2$</td>
<td>12.89</td>
<td>15.57</td>
<td>$\beta^O_3$</td>
<td>18.37</td>
</tr>
<tr>
<td>$\beta^Y_3$</td>
<td>14.12</td>
<td>15.00</td>
<td>$\beta^O_3$</td>
<td>18.29</td>
</tr>
<tr>
<td>$\beta^Y_4$</td>
<td>14.36</td>
<td>16.29</td>
<td>$\beta^O_4$</td>
<td>20.73</td>
</tr>
<tr>
<td>$N$</td>
<td>4.28</td>
<td>4.77</td>
<td>$\beta^O_4$</td>
<td>18.33</td>
</tr>
<tr>
<td>$p^Y_{ij} = 0.5, p^O_{ij} = 0.2, \varphi^Y_i = 0.3, \varphi^O_i = 0.35, N = 10000$</td>
<td>$\beta^Y_0$</td>
<td>13.51</td>
<td>13.77</td>
<td>$\beta^O_0$</td>
</tr>
<tr>
<td>$\beta^Y_0$</td>
<td>14.91</td>
<td>15.43</td>
<td>$\beta^O_0$</td>
<td>14.63</td>
</tr>
<tr>
<td>$\beta^Y_2$</td>
<td>14.54</td>
<td>14.41</td>
<td>$\beta^O_3$</td>
<td>20.00</td>
</tr>
<tr>
<td>$\beta^Y_3$</td>
<td>15.86</td>
<td>15.46</td>
<td>$\beta^O_3$</td>
<td>18.30</td>
</tr>
<tr>
<td>$\beta^Y_4$</td>
<td>12.51</td>
<td>15.66</td>
<td>$\beta^O_4$</td>
<td>19.99</td>
</tr>
<tr>
<td>$N$</td>
<td>3.23</td>
<td>3.81</td>
<td>$\beta^O_4$</td>
<td>18.68</td>
</tr>
<tr>
<td>$p^Y_{ij} = 0.5, p^O_{ij} = 0.2, \varphi^Y_i = 0.3, \varphi^O_i = 0.35, N = 10000$</td>
<td>$\beta^Y_0$</td>
<td>4.31</td>
<td>4.35</td>
<td>$\beta^O_0$</td>
</tr>
<tr>
<td>$\beta^Y_0$</td>
<td>4.75</td>
<td>4.42</td>
<td>$\beta^O_0$</td>
<td>4.52</td>
</tr>
<tr>
<td>$\beta^Y_2$</td>
<td>4.72</td>
<td>4.89</td>
<td>$\beta^O_3$</td>
<td>5.96</td>
</tr>
<tr>
<td>$\beta^Y_3$</td>
<td>4.46</td>
<td>4.61</td>
<td>$\beta^O_3$</td>
<td>5.73</td>
</tr>
<tr>
<td>$\beta^Y_4$</td>
<td>4.86</td>
<td>5.22</td>
<td>$\beta^O_4$</td>
<td>5.83</td>
</tr>
<tr>
<td>$N$</td>
<td>0.45</td>
<td>1.33</td>
<td>$\beta^O_4$</td>
<td>5.76</td>
</tr>
<tr>
<td>$p^Y_{ij} = 0.5, p^O_{ij} = 0.2, \varphi^Y_i = 0.15, \varphi^O_i = 0.55, N = 10000$</td>
<td>$\beta^Y_0$</td>
<td>4.68</td>
<td>4.74</td>
<td>$\beta^O_0$</td>
</tr>
<tr>
<td>$\beta^Y_0$</td>
<td>4.67</td>
<td>4.89</td>
<td>$\beta^O_0$</td>
<td>4.70</td>
</tr>
<tr>
<td>$\beta^Y_2$</td>
<td>4.97</td>
<td>4.97</td>
<td>$\beta^O_3$</td>
<td>6.12</td>
</tr>
<tr>
<td>$\beta^Y_3$</td>
<td>4.87</td>
<td>4.92</td>
<td>$\beta^O_3$</td>
<td>6.08</td>
</tr>
<tr>
<td>$\beta^Y_4$</td>
<td>4.81</td>
<td>4.96</td>
<td>$\beta^O_4$</td>
<td>5.76</td>
</tr>
<tr>
<td>$N$</td>
<td>3.23</td>
<td>3.81</td>
<td>$\beta^O_4$</td>
<td>5.89</td>
</tr>
</tbody>
</table>
Table 2.2: Estimate the entry probability of Kangaroo rat data set under three different super population capture-recapture models, model with “perfect” genetic assignment test (“Perfect model” columns), model with imperfect genetic assignment test (“Imperfect model” columns), and model without genetic assignment test (“Standard model”).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Perfect model</th>
<th>Imperfect model</th>
<th>Standard model</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_{0B}$</td>
<td>.1033 .0398</td>
<td>.1000 .0340</td>
<td>.0995 .0242</td>
</tr>
<tr>
<td>$\beta_{1B}$</td>
<td>.0998 .0239</td>
<td>.0948 .0282</td>
<td>.0857 .0195</td>
</tr>
<tr>
<td>$\beta_{2B}$</td>
<td>.0944 .0325</td>
<td>.0944 .0322</td>
<td>.0859 .0208</td>
</tr>
<tr>
<td>$\beta_{3B}$</td>
<td>.1935 .0385</td>
<td>.1939 .0426</td>
<td>.1486 .0287</td>
</tr>
<tr>
<td>$\beta_{4B}$</td>
<td>.0528 .0193</td>
<td>.0558 .0199</td>
<td>.0450 .0146</td>
</tr>
<tr>
<td>$\beta_{5B}$</td>
<td>.0825 .0208</td>
<td>.0824 .0196</td>
<td>.0721 .0170</td>
</tr>
<tr>
<td>$\beta_{6B}$</td>
<td>.1427 .0179</td>
<td>.1468 .0232</td>
<td>.1355 .0233</td>
</tr>
<tr>
<td>$\beta_{7B}$</td>
<td>.0692 .0165</td>
<td>.0696 .0159</td>
<td>.0640 .0154</td>
</tr>
<tr>
<td>$\beta_{0I}$</td>
<td>.1015 .0255</td>
<td>.1017 .0266</td>
<td>.1218 .0263</td>
</tr>
<tr>
<td>$\beta_{1I}$</td>
<td>.0232 .0114</td>
<td>.0248 .0103</td>
<td>.0211 .0147</td>
</tr>
<tr>
<td>$\beta_{2I}$</td>
<td>.0108 .0081</td>
<td>.0079 .0066</td>
<td>.0119 .0110</td>
</tr>
<tr>
<td>$\beta_{3I}$</td>
<td>.0100 .0001</td>
<td>.0010 .0001</td>
<td>.0146 .0110</td>
</tr>
<tr>
<td>$\beta_{4I}$</td>
<td>.0100 .0001</td>
<td>.0010 .0001</td>
<td>.0146 .0110</td>
</tr>
<tr>
<td>$\beta_{5I}$</td>
<td>.0050 .0045</td>
<td>.0053 .0042</td>
<td>.0204 .0129</td>
</tr>
<tr>
<td>$\beta_{6I}$</td>
<td>.0010 .0001</td>
<td>.0010 .0001</td>
<td>.0129 .0099</td>
</tr>
<tr>
<td>$\beta_{7I}$</td>
<td>.0088 .0060</td>
<td>.0099 .0075</td>
<td>.0216 .0096</td>
</tr>
</tbody>
</table>

| $\beta_{0}$ | .0094 .0059 | .0095 .0061 | .0116 .0071 |
Table 2.3: Estimate the survival probability of Kangaroo rat data set under three different super population capture-recapture models, model with “perfect” genetic assignment test (“Perfect model” columns), model with imperfect genetic assignment test (“Imperfect model” columns), and model without genetic assignment test (“Standard model”).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Perfect model</th>
<th>Imperfect model</th>
<th>Standard model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimator</td>
<td>S.E.</td>
<td>Estimator</td>
</tr>
<tr>
<td>$\phi_{Y1}$</td>
<td>.1415</td>
<td>.0807</td>
<td>.1372</td>
</tr>
<tr>
<td>$\phi_{Y2}$</td>
<td>.3402</td>
<td>.1154</td>
<td>.3538</td>
</tr>
<tr>
<td>$\phi_{Y3}$</td>
<td>.2967</td>
<td>.1183</td>
<td>.2954</td>
</tr>
<tr>
<td>$\phi_{Y4}$</td>
<td>.3235</td>
<td>.0789</td>
<td>.3128</td>
</tr>
<tr>
<td>$\phi_{Y5}$</td>
<td>.2816</td>
<td>.2228</td>
<td>.2892</td>
</tr>
<tr>
<td>$\phi_{Y6}$</td>
<td>.1601</td>
<td>.0457</td>
<td>.1776</td>
</tr>
<tr>
<td>$\phi_{Y7}$</td>
<td>.2009</td>
<td>.0414</td>
<td>.1951</td>
</tr>
<tr>
<td>$\phi_{O1}$</td>
<td>.3711</td>
<td>.1127</td>
<td>.3678</td>
</tr>
<tr>
<td>$\phi_{O2}$</td>
<td>.4461</td>
<td>.1366</td>
<td>.4869</td>
</tr>
<tr>
<td>$\phi_{O3}$</td>
<td>.3463</td>
<td>.1691</td>
<td>.3357</td>
</tr>
<tr>
<td>$\phi_{O4}$</td>
<td>.2762</td>
<td>.1092</td>
<td>.2715</td>
</tr>
<tr>
<td>$\phi_{O5}$</td>
<td>.4151</td>
<td>.2796</td>
<td>.3646</td>
</tr>
<tr>
<td>$\phi_{O6}$</td>
<td>.1459</td>
<td>.0916</td>
<td>.1569</td>
</tr>
<tr>
<td>$\phi_{O7}$</td>
<td>.0618</td>
<td>.0594</td>
<td>.0584</td>
</tr>
</tbody>
</table>
Bibliography


Chapter 3

Multi-State Super Population Capture-Recapture Model Integrated with Origination Assignment Data

**summary** We first extend the super population capture-recapture model to multi-state(location) and multi-age case. Then we augmented our multi-state super population capture-recapture model with origination assignment procedures. Origination assignment procedures are very useful approaches to distinguish the origination of the animals. By splitting the entry probability into *in situ* birth entry probability of each population and an immigrant entry probability and showing that it is possible to estimate these entry probabilities separately, we are able to estimate the relative contributions of *in situ* births of every population and the immigrants from outside to the growth of the meta population system. The fusion of origination information and capture-recapture data will allow us to improve the estimation of demographic parameters, like survival probability, entry probability, and especially movement rates. We apply our new model to a capture-recapture study with origination information for banner-tailed kangaroo rats in Southern Arizona.
3.1 Introduction

The Jolly-Seber (Jolly 1965, Seber 1965, Seber 1982) method is a very popular open population capture-recapture method. It was originally developed to estimate demographic parameters of open populations in which there is possibly death, recruitment, immigration, and permanent emigration. In the original Jolly-Seber models, the tagged animals are assumed to pertain to one homogeneous population segment. The multi-state model is a natural extension of the Jolly-Seber model, where animals are allowed to distribute across multiple states. Multi-state models were first considered by Arnason (1972, 1973), who conducted the capture-recapture studies with animals that could move among different sub populations (Schwartz 1993). Nichols et al. (1992, 1994) developed a model, where meadow voles made transitions between physiological states.

In multi-state models, every captured animal is assigned to a state at each time point. The animals move among a finite number of discrete states according to an unknown chance process. Although the transitions are stochastic, we assume that the state of an animal at time \( i + 1 \) is stochastically determined by its state at time \( i \). This is called a first-order Markovian model and the Arnason-Schwarz model is of this kind. Hestbeck et al. (1991) and Brownie et al. (1993) developed multi-state models with non-Markovian transitions but we do not consider them here. It is assumed that the true state of an encountered animal is known with certainty. Kendall et al. (2003) considered a more general model where this assumption is violated.

The Robust design is a very important sampling scheme which combines the features of both closed and open population models. It was first proposed by Pollock (1982). With the Robust design, we can estimate some parameters (like first year’s capture probability and last year’s population size) which cannot be estimated under the Jolly-Seber model (Pollock, 1982, Pollock et al., 1990, Kendall and Pollock, 1992). The robust design can also be used together with the multi-state model (Nichols and Coffman, 1999, Williams et al., 2002). In the robust design, each primary period consists of several shorter secondary sampling periods. The population is assumed to be closed among these shorter sampling periods of each primary period for each state.

The super population model is one representation of the Jolly-Seber model. It was first proposed by Crosbie and Manly (1985) and then developed further by Schwarz and
Arnason (1996). In single state super population model, it treats all the individuals exposed to sampling at any time of the study as a part of a super population \( (N) \). The animals of the super population are assumed to enter the super population during the sampling periods with certain entry probability. We can extend the super population model to multi-state situation. In the multi-state super population model, we treat all the individuals of all states exposed to sampling at any time of the study as a part of a super population \( (N) \). The animals of age \( r \) enter the super population with an entry probability \( \beta_{rh}^{i} \) between sampling occasion \( i-1 \) and \( i \) with specific state \( h \). That is, the super population is composed of all the new recruits to all the sates and of all ages from all the sampling periods \( N = \sum_{i,r,h} B_{rh}^{i} \), where \( B_{rh}^{i} \) is the numbers of new recruits that enter the super population at age \( r \) with state \( h \) between period \( i \) and \( i+1 \).

Throughout this paper, the multi-state means multiple populations, the system is composed of all the populations we are interested in, and there are movements between the populations (within the system). For a multi-population system, the new recruits of any population of the system can only come from three sources. The first possibility is the animal is an on site born of the population where it is captured, we defines this kind of animal as \textit{in situ} birth. The second possible origination is the animal is born in another population of the system however move to this population and being captured. We call this kind of animal as population immigrant. The last possible origination is the animals is coming from outside the system. We call these animals as system immigrant. Analyzing the components of the new recruits of each population of the system is of great interest to ecologists.

Origination assignment test means that the origination of one animal can be assigned with high certainty based on the animal’s genetic information, or other property. There are lots of approaches to provide the source of animals, like genetic assignment test, genetic parentage tests, and feather isotope analysis. A perfect origination assignment test means each animal captured of any population can be assigned perfectly either as an \textit{in situ} birth, or as a population immigrant, or as a system immigrant. That is, a “birth state” (immigrant, or an \textit{in situ} birth) is assigned to any captured animal where this state is otherwise unknown (i.e. where an animal has not been captured in its birth year).

Here we propose a new approach to combine the origination assignment test information with the multi-state super population model approach. We shall begin with the
assumption that the genetic assignment test is perfect (i.e. without uncertainty) and later
generalize to a model for imperfect origination assignment tests (i.e., the uncertainty in
the assignment is accounted for). The fusion of population origination data and capture-
recapture data will allow us to determine the separate contributions of population immi-
gration, system immigration, and \textit{in situ} reproduction to the growth of the each population
of the system. Further, the origination information will greatly help us to estimate the
dispersal process between populations which are difficult to observed based on standard
capture-recapture data.

3.2 Model Structure

Consider a model composed of \( L \) locations, \( K \) primary sampling period, \((l_i)\)
secondary sampling periods within each primary period, and two age groups, Y (young,
0-1 year) and O (old, \( \geq 1 \) year). Young animals (0 – 1 year old) naturally become old
animals (\( \geq 1 \) year old) one year after birth. The system is composed of all the young and
old animals in the \( L \) locations within the \( K \) sampling periods. The super population \( N \) is
all the animals that have been alive and potentially subject to capture efforts at least once
in our \( K \) primary periods. The new recruits enter the population through being an \textit{in situ}
birth of one of the \( L \) locations or as an immigrant from outside the system. Assume the
young animals are always captured at their natal place.

3.2.1 Notation

\( p_{ij}^{rh} \): the probability that an animal of age \( r \) and population \( h \) is captured on secondary
sampling period \( j \) of primary period \( i \), given that it is in the super population at that
period. \( q_{ij}^{rh} = 1 - p_{ij}^{rh} \).

\[
p_{i \cdot}^{r} = \left[ \begin{array}{ccc} p_{i \cdot}^{r1} & \cdots & p_{i \cdot}^{rh} & \cdots & p_{i \cdot}^{rL} \end{array} \right], \ i = 1, \ldots , K; \ r = Y, O, \text{with } p_{i \cdot}^{rh} \text{ be the}
\text{probability that an animal of age } r \text{ in population } h \text{ is being captured or observed}
\text{at least once in one secondary period of primary period } i. \ p_{i \cdot}^{rh} = 1 - \prod_{j=1}^{L} (1 - p_{ij}^{rh}).
\]

\[
q_{i \cdot}^{r} = \left[ \begin{array}{ccc} q_{i \cdot}^{r1} & \cdots & q_{i \cdot}^{rh} & \cdots & q_{i \cdot}^{rL} \end{array} \right]
\]

\[
\beta_{i \cdot}^{r} = \left[ \begin{array}{ccc} \beta_{i \cdot}^{r1} & \cdots & \beta_{i \cdot}^{rh} & \cdots & \beta_{i \cdot}^{rL} \end{array} \right], \ i = 0, 1, \ldots , K - 1; \ r = Y, O, \text{with } \beta_{i \cdot}^{rh} \text{ be the}
\text{probability of an animal of the super population entering population } h \text{ at primary}
period \(i\) as age \(r\).

\[ u^r_i = \begin{bmatrix} u^r_{i1} & \cdots & u^r_{ih} & \cdots & u^r_{iL} \end{bmatrix}, \quad i = 1, \cdots, K; r = Y, O, \] with \(u^r_i\) is the number of unmarked animals of age \(r\) captured on at least one secondary occasion within primary period \(i\) at population \(h\).

\[ Pu^r_i = \begin{bmatrix} Pu^r_{i1} & \cdots & Pu^r_{ih} & \cdots & Pu^r_{iL} \end{bmatrix}, \quad i = 1, \cdots, K; r = Y, O, \] with \(Pu^r_i\) is the probability the animal of age \(r\) is captured at primary period \(i\) at population \(h\) given that the animal has never been captured before.

\[ R^r_i = \begin{bmatrix} R^r_{i1} & \cdots & R^r_{ih} & \cdots & R^r_{iL} \end{bmatrix}, \quad i = 1, \cdots, K; r = Y, O, \] with \(R^r_i\) is the number of marked animals of age \(r\) that are released to population \(h\) at period \(i\)

\[ \phi^r_i = \begin{bmatrix} \phi^r_{i11} & \cdots & \phi^r_{ihq} & \cdots & \phi^r_{iLL} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \phi^r_{ih1} & \cdots & \phi^r_{ihq} & \cdots & \phi^r_{ihL} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \phi^r_{iL1} & \cdots & \phi^r_{iLq} & \cdots & \phi^r_{iLL} \end{bmatrix}, \quad i = 1, \cdots, K-1; r = Y, O, \] where \(\phi^r_{ihq}\) is the probability that an animal is in population \(q\) at primary period \(i + 1\), given that in primary period \(i\), the animal was in population \(h\) and was age \(r\).

\[ \psi^r_i = \begin{bmatrix} \psi^r_{i11} & \cdots & \psi^r_{ihq} & \cdots & \psi^r_{iLL} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \psi^r_{ih1} & \cdots & \psi^r_{ihq} & \cdots & \psi^r_{ihL} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \psi^r_{iL1} & \cdots & \psi^r_{iLq} & \cdots & \psi^r_{iLL} \end{bmatrix}, \quad i = 1, \cdots, K-1; r = Y, O, \] where \(\psi^r_{ihq}\) is the probability the animal survives to primary period \(i\) to population \(q\) given that the animal enter population \(h\) at age \(r\) before primary period \(i\) but never be captured by period \(i\).

\[ \phi^r_{i,j} = \begin{bmatrix} \phi^r_{i,j1} & \cdots & \phi^r_{i,jq} & \cdots & \phi^r_{i,jL} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \phi^r_{i,jh1} & \cdots & \phi^r_{i,jhq} & \cdots & \phi^r_{i,jhL} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \phi^r_{i,jL1} & \cdots & \phi^r_{i,jLq} & \cdots & \phi^r_{i,jLL} \end{bmatrix}, \quad i = 1, \cdots, K-1; j = i+1, \cdots, K, r = Y, O, \] where \(\phi^r_{i,jq}\) is the probability the animal survives to primary period \(j\) at population \(q\)
given that the animal last capture is happened at primary period $i$ at population $i$ of age $r$.

$$m_{i,j}^r = \begin{bmatrix}
  m_{i,j}^{11} & \cdots & m_{i,j}^{1q} & \cdots & m_{i,j}^{1L} \\
  \vdots & \ddots & \vdots & \ddots & \vdots \\
  m_{i,j}^{r1} & \cdots & m_{i,j}^{rq} & \cdots & m_{i,j}^{rL} \\
  \vdots & \ddots & \vdots & \ddots & \vdots \\
  m_{i,j}^{L1} & \cdots & m_{i,j}^{Lq} & \cdots & m_{i,j}^{LL}
\end{bmatrix}, \quad i = 1, \ldots, K-1; j = i+1, \ldots, K; r = Y, O,$$

where $m_{i,j}^{rhq}$ is the number of animals recaptured from population $h$ in primary period $j$, given that the animals were released to population $h$ at age $r$ at period $i$ and were not recaptured between primary period $i$ and $j$.

$$P_{i,j}^r = \begin{bmatrix}
P_{i,j}^{11} & \cdots & P_{i,j}^{1q} & \cdots & P_{i,j}^{1L} \\
  \vdots & \ddots & \vdots & \ddots & \vdots \\
  P_{i,j}^{r1} & \cdots & P_{i,j}^{rq} & \cdots & P_{i,j}^{rL} \\
  \vdots & \ddots & \vdots & \ddots & \vdots \\
  P_{i,j}^{L1} & \cdots & P_{i,j}^{Lq} & \cdots & P_{i,j}^{LL}
\end{bmatrix}, \quad i = 1, \ldots, K-1, j = i+1, \ldots, K, r = Y, O.$$

$P_{i,j}^{rhq}$ is the probability that an animal of age $r$ in population $h$ alive at period $i$ is next be recaptured from population $j$ in sampling period $j$.

$$m_j^O = \begin{bmatrix}
m_j^{O1} & \cdots & m_j^{Oh} & \cdots & m_j^{OL}
\end{bmatrix}, \quad j = 2, \ldots, K,$$

with $m_j^{Oh}$ is the number of marked animals recaptured at primary period $j$ from population $h$ of age $O$. We have

$$m_j^{Oh} = \sum_{r=Y,O} \sum_{i=1}^{j-1} \sum_{q=1}^{L} m_{ij}^{rqh}.$$

There is no term of $m_j^Y$.

$$n_i^r = \begin{bmatrix}
n_i^{r1} & \cdots & n_i^{rh} & \cdots & n_i^{rL}
\end{bmatrix}, \quad i = 1, \ldots, K; r = Y, O,$$

$n_i^r$ is the number of animals of age $r$ caught on at least one secondary occasion within primary period $i$ at population $h$. Usually, $n_i^r \geq R_i^r$ when release probability is less than 1. But here we assume all captured animals are released back, so $n_i^r = R_i^r$.

$Diag(x)$: a diagonal matrix with diagonal elements equal to the elements of $x$, given that $x$ is a vector. If $x$ is matrix, $Diag(x) = x$.

$Colvec(A)$: is a column vector function. Where $A$ is a $n$ by $m$ matrix, the $colvec$ function returns an $nm \times 1$ vector. It converts a matrix into a column vector in row-major order. The returned vector has 1 column and $nm$ rows. The first $m$ elements
in the vector correspond to the first row of the input matrix, the next \( m \) elements corresponding to the second row, and so on.

\( \text{sum}(x) \): is the summation of all the elements of \( x \), where \( x \) is a vector.

\( A(\mathbf{.}) \): is the summation of the rows for each column given \( A \) is a matrix.

### 3.2.2 Standard Multi-State Super Population Capture-Recapture Model

We assume the multi-state super population model using the robust design follows a product of multinomial distributions that includes the parameters of entry probabilities, transition probabilities, super population size, and secondary period capture probabilities. The overall likelihood for this model can be written as

\[
L = L_1 \times L_2 \times L_3
\]  

(3.1)

where \( L_1, L_2 \) and \( L_3 \) are the conditional likelihoods for the first captures component within the primary periods, the recaptures component within the primary periods, and the secondary periods capture-recapture component (robust design) respectively. Each part of is a product of multinomial distributions. We will now introduce how to get the likelihood components.

We assume all young animals captured at their natal place, we have

\[
\psi_{i}^{Yh} = \beta_{i-1}^{Yh} \quad E(u_{i}^{Yh}) = N \ast \psi_{i}^{Yh} \ast p_{i}^{Yh}
\]  

(3.2)

For old animals, the situation becomes complex because the captured old animals are composed of two groups, one group is the animals entering into the super population as young and not captured on entering but captured later at an older age. The other group is the animals entering the super population at an older age. For the newly captured old animals entering the super population at young age:

\[
\psi_{1}^{OY} = 0
\]

\[
\psi_{2}^{OY} = \text{Diag}(\beta_{0}^{Y}) \ast \text{Diag}(q_{1}^{Y}) \ast \varphi_{1}^{Y}
\]

\[
\psi_{i}^{OY} = \psi_{i-1}^{Y} \ast \text{Diag}(q_{i}^{OY}) + \text{Diag}(\beta_{i-1}^{Y}) \ast \text{Diag}(q_{i-1}^{Y}) \ast \varphi_{i-1}^{Y}
\]  

(3.3)

\[
i = 3, \cdots, K
\]  

(3.4)
For the newly captured old animals entering the super population at adult age:

\[
\psi_1^{OO} = \text{Diag}(\beta_0^O) \\
\psi_i^{OO} = \psi_{i-1}^{OO} \ast \text{Diag}(q_{i-1}^{OO}) + \text{Diag}(\beta_{i-1}^O), i = 2, \cdots, K
\]  

(3.5)

Assume the numbers of newly unmarked animals captured on each occasion, each population, and each age form a multinomial distribution which is decomposed into a marginal (first term) and conditional multinomial (second term). Then the likelihood function for the new capture components of primary periods in the multi-state case with the robust design is,

\[
L_1 = \frac{N}{u!(N-u)!}(Pu_u)^u(1-Pu_u)^{N-u} \\
\prod_{i=1}^{K-1} \prod_{h=1}^{L_{Yh}} \prod_{r=1}^{V_{Yr}} \prod_{i=1}^{K-1} \prod_{h=1}^{L_{Oh}} \prod_{r=1}^{V_{Or}} \frac{Pu_{rh}^h}{Pu_{rh}^h} \\
\text{where } Pu_u^Y = \psi_1^{Yh} \ast p_{iY}^Y, Pu_u^O = (\psi_1^{OY} + \psi_{i}^{OO}) \ast \text{Diag}(p_{i}^{*O})[+, 1]; Pu_u = \sum_i \sum_h (Pu_{Yh}^h + Pu_{Oh}^h)
\]

(3.6)

and \( u = \sum_i \sum_h (u_{1Y}^h + u_{1O}^h) \)

The recapture component of the primary periods is the conditional distribution of animals captured in primary periods where the conditioning is on the numbers of released animals at each primary period each population and each age \( R_{rh}^i \). We assume all captured animals are released back, that is, \( R_i = n_i^r \) here. We assume the recaptures conditional on the release number at the same age and population follow a multinomial distribution. The probability distribution of this multi-state model is a product of \((K-1)^*M*L\) conditional multinomial conditional on the releases \( R_{rh}^i \) respectively. For young released at young age,

\[
\phi_{Y}^{i+1} = \varphi_{iY}^Y = \varphi_{i,j-1}^{Y} \ast \text{Diag}(q_{j-1}^{O}) \ast \varphi_{j-1}^O \\
P_{i,j} = \varphi_{i,j}^{Y} \ast p_{j}^{*O}
\]  

(3.7)

For animals released at old age,

\[
\phi_{O}^{i+1} = \varphi_{i}^{O} = \varphi_{i,j-1}^{O} \ast \text{Diag}(q_{j-1}^{O}) \ast \varphi_{j-1}^O \\
P_{i,j} = \varphi_{i,j}^{O} \ast p_{j}^{*O}
\]  

(3.8)

where \( i = 1, \cdots, K - 1, j = i + 1, \cdots, K \). Let \( m_{rh}^i = \sum_{j=i+1}^{K} \sum_{q=1}^{L} m_{rhq}^i \). The animals of \( R_{rh}^i \) being recaptured for just once from primary period \( i + 1 \) to primary period \( j \).
\( P_{rh} = \sum_{j=i+1}^{K} \sum_{q=1}^{L} P_{iq,j}^{rhq} \): The probability that an animal is begin captured for just once from primary period \( i + 1 \) to primary period \( j \) given that the animal is part of \( R_{rh}^i \).

Now we get the likelihood of the multi-state model for the recapture part of the primary periods,

\[
L_2 \propto \prod_{i=1}^{K-1} \prod_{r=Y,O} \prod_{h=1}^{L} \prod_{q=1}^{L} (P_{iq,j}^{rhq})^{m_{iq,j}} \times (1 - P_{rh}^i)R_{rh}^i - m_{rh}^i
\]

(3.9)

Assume \( l_1 = l_2 = \cdots = l_{K+2+L} = 2 \), the likelihood for the robust design of the model we considered (\( K \) sampling periods, \( L \) locations, \( Y \) and \( O \) ages) is:

\[
L_3 = \prod_{i=1}^{K} \prod_{r=Y,O} \prod_{h=1}^{L} \chi_{rh\omega}^i \times (P_{rh}^{1p})^{\chi_{rh10}^i} \times (P_{rh}^{2p})^{\chi_{rh01}^i} \times (P_{rh}^{3p})^{\chi_{rh11}^i}
\]

(3.10)

where \( \chi_{rh\omega}^i \) is the total number of animals of age \( r \) and population \( h \) caught in primary period \( i \) that exhibit capture history \( \omega \) over \( l \) secondary periods. We can get the likelihood for the robust design for any \( l_i \geq 2 \) similarly. We can also allow different closed models (time, trap response etc) for different primary periods and different states. All survival rates are also identifiable when using the Robust Design.

### 3.2.3 Multi-State Super Population Model with Robust Design Augmented with Perfect Origination Assignment Information

Assume the young animals are captured at their natal place and consider the origination source of animals, we have,

\( \beta_i^Y = \text{Diag}(\beta_1^{Y1B}, \cdots, \beta_i^{YLB}, 0), i = 0, \cdots, K - 1 \). That is \( \beta_i^{YhBq}, h \neq q \) and \( \beta_i^{YhI} \) do not exist. There are also no immigrants into the super population as young because after their movement, they become old animals (and are no longer young). Correspondingly, for new captured young animals we have \( u_i^Y = \text{Diag}(u_1^{Y1B}, \cdots, u_i^{YLB}, 0), i = 1, \cdots, K \) since \( u_i^{Yh} = u_i^{YhBh} \).

\[
\beta_i^O = \begin{bmatrix}
0 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
\beta_i^{O1I} & \beta_i^{O2I} & \cdots & \beta_i^{OLI}
\end{bmatrix}
\]
since all entering old animals of the super population
are immigrants from outside the system, $\beta^{Oh}_i = \beta^{OhI}_i$ and $u^{Oh}_i = u^{OhI}_i$, $i = 1, \cdots, K-1$

$$
\beta^O_i = \begin{bmatrix}
\beta^{O1B1}_0 & \cdots & \beta^{OhB1}_0 & \cdots & \beta^{OLB1}_0 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
\beta^{O1B1}_0 & \cdots & \beta^{OhB1}_0 & \cdots & \beta^{OLB1}_0 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
\beta^{O1I}_0 & \cdots & \beta^{OhI}_0 & \cdots & \beta^{OLI}_0 \\
\end{bmatrix},
$$

which includes all the animals born before the first primary sampling period and entering into the super population at the first primary period. $\beta^{OhBq}_0$ is the probability of an adult animal born at population $h$ entering the super population at population $q$ at primary period 1. Similarly, $\beta^{OhI}_0$ is the probability of an adult immigrant entering the population $h$ at primary period 1.

$$
u^O_i = \begin{bmatrix}
u^{O1B1}_i & \cdots & \nu^{OhB1}_i & \cdots & \nu^{OLB1}_i \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
\nu^{O1I}_i & \cdots & \nu^{OhI}_i & \cdots & \nu^{OLI}_i \\
\end{bmatrix},
$$
i = 1, \cdots, K, j = 1, \cdots, K; r is age. $u^{rhBq}_i$

is the number of unmarked animals of age $r$ captured on at least one secondary occasion within primary period $i$ at population $h$ given the animal is an in situ birth of population $q$. Similarly, $u^{rhI}_i$ is the number of unmarked animals age $r$ captured on at least one secondary occasion within primary period $i$ at population $h$ given these animals originate outside the system.

Now we consider incorporating additional origination information into the super population model with multi states to aid in estimation of the separate contributions of in situ reproduction and immigration to the growth of the population. Origination assignment procedures are very useful approaches to distinguish the origination of the animals. In perfect origination assignment procedures, we can assign the origination of any captured animal with certainty. That is, we can split the new captured unmarked animals $u^{rh}_i = u^{rhB1}_i + u^{rhB2}_i + \cdots + u^{rhBl}_i + u^{rhI}_i$ accordingly. Correspondingly, we can decompose the entry probability according to their origination as $\beta^{rh}_i = \beta^{rhB1}_i + \beta^{rhB2}_i + \cdots + \beta^{rhBl}_i + \beta^{rhI}_i$

After included the origination assignment information, the likelihood of multi-state super population capture-recapture model becomes

$$L = L^*_1 \times L^*_2 \times L_3$$

(3.11)
where \( L_1^* \) is the likelihood for the newly captured \textit{in situ} birth and immigrants incorporating information on population of origin, \( L_2^* \) is the modified likelihood for the recapture component of primary periods conditional on the releases \( R_i^B \) and \( R_i^I \), and \( L_3 \) is the likelihood for the secondary period component. If we assume the system immigrants and the animals of the populations having the same transition rate, \( L_2^* \) and \( L_3 \) are the same as in the standard super population capture-recapture model if we assume that individuals born \textit{in situ} and immigrants have the same capture probabilities. When origination assignment information added, we can also assume the animals of the system and out of the system having different survival probability, which is not discussed here.

Conditional on the super population \( N \), we assume all the newly captured animals of all age all sampling period and all populations follow a multinomial distribution. For new captured young animals,

\[
P_{u_i}^{Yh} = \beta_{i-1} YhB \ast p_i^{Yh} \\
E(u_i^{Yh}) = N \ast P_{u_i}^{Yh} \tag{3.12}
\]

For newly captured old animals entering the super population at young age,

\[
\psi_1^{Y} = 0 \\
\psi_2^{Y} = \beta_{0}^Y \ast \text{Diag}(q_1^{Y}) \ast \varphi_1^{Y} \\
\psi_{i}^{Y} = \psi_{i-1}^{Y} \ast \text{Diag}(q_{i-1}^{O}) \varphi_{i-1}^{O} + \beta_{i-2}^Y \ast \text{Diag}(q_{i-1}^{Y}) \varphi_{i-1}^{Y} \\
i = 3, \cdots, K \tag{3.13}
\]

For newly captured old animals entering the super population at ode age,

\[
\psi_1^{O} = \beta_{0}^{O} \\
\psi_{i}^{O} = \psi_{i-1}^{O} \ast \text{Diag}(q_{i-1}^{O}) \varphi_{i-1}^{O} + \beta_{i-1}^{O}, i = 2, \cdots, K \tag{3.14}
\]

Let \( P_{u_i}^{O} = (\psi_1^{Y} + \psi_1^{O}) \ast \text{Diag}(p_i^{O}), P_{u_i}^{Ovec} = \text{colvec}(P_{u_i}^{O}), u_i^{Ovec} = \text{colvec}(u_i^{O}), P_u = \sum_i P_{u_i}^{Ovec} + \sum_{i,h} P_{u_i}^{Yh} \), and \( u. = \sum_i u_i^{Ovec} + \sum_{i,h} u_i^{Yh} \), we have

\[
L_1^* = \frac{N!}{u.!(N-u.!)^{(P_u.u)^u \ast (1-P_u..)^{N-u}}} \prod_{i=1}^{K} \prod_{h=1}^{L} u_i^{YhB1} u_i^{OhI1} u_i^{OhB1!} \\
\prod_{i=1}^{K} \prod_{h=1}^{L} \frac{P_{u_i}^{Yh}}{P_u} \prod_{i=1}^{K} \prod_{j=1}^{L} \frac{P_{u_i}^{Ovec[j]}}{P_u} \tag{3.15}
\]
3.2.4 Multi-State Super Population Capture-Recapture Model Augmented with Imperfect Origination Assignment Information

As in the case where we had a perfect origination assignment procedure, any animal captured for which the birth state is unknown has an origination assignment procedure carried out so that a state of origin is assigned. However, now this origination assignment information is subject to uncertainty which is what occurs in practice. To simplify the explanation, here the model is composed of two populations, 1 and 2, and two age groups, young and adult. The system is composed of all the animals of population 1 and 2 and of all ages. Consider a new captured adult (at least 1 year old) of population 2 has origination assignment procedure done. The origination assignment procedure assigns the animal an origination probabilities vector with \((\pi_1, \pi_2, \pi^I = 1 - \pi_1 - \pi_2)\), \((0 < \pi_1, \pi_2, \pi^I < 1)\). We view \(\pi_1\) as the probability that the animal is an in situ birth of population 1, \(\pi_2\) as the probability that the animal is an in situ birth of population 2, and \(\pi^I\) as the probability that it is an immigrant of the system. If the assignment test is a good one then when the animal is truly an in situ birth of population 1, \(\pi_1\) will be close to one and both \(\pi_2\) and \(\pi^I\) will be close to zero; if the animal is truly an in situ birth of population 2, \(\pi_2\) will be close to one and both \(\pi_1\) and \(\pi^I\) will be close to zero; whereas when the animal is truly an immigrant of the system, \(\pi_1\) and \(\pi_2\) will be close to zero and \(\pi^I\) will be close to one. Further, good assignment test information will be valuable and improve our estimators of recruitment parameters.

Consider that each newly captured adult animal has an origination assignment procedure vector \((\pi_1, \pi_2, \pi^I = 1 - \pi_1 - \pi_2)\), \((0 < \pi_1, \pi_2, \pi^I < 1)\). How to include the uncertainty of the assignment process into the likelihood. We will take a resampling approach similar to multiple imputation (Rubin’s, 1987) to “impute” the perfect assignment data based on the imperfect information. First, each newly captured adult animal will be randomly allocated a uniform random variable \(U_j \in (0, 1)\). If \(U_j \in (0, \pi_1)\) then we assign this animal as an in situ birth of population 1, if \(U_j \in (\pi_1, \pi_1 + \pi_2)\) then we assign this animal as an in situ birth of population 2, whereas if \(U_j \in (\pi_1 + \pi_2, 1)\), we assign this animal as an immigrant of the system. Do this assignment procedure for all newly captured adult animals, \(j = 1, \cdots, \sum_{i,h} u_{ih}^{Oh}\). Through this, we split \((u_i^{Oh})\) into \(u_i^{OhB1}, u_i^{OhB2}\), and \(u_i^{OhI}\) as in the perfect assignment test situation and get a “pseudo perfect assignment test data” set.
We treat the resulted data $u_i^{OhBq}$ as perfect origination assignment situation and combine the origination information with the capture-recapture data, we can estimate the *in situ* birth entry probability for each population, immigrant entry probability from outside, and transition rate. We repeat the assignment process m times and for each of the m “pseudo data sets” we estimate the parameters. Then we can take various distributional statistics (means, medians, standard deviation, and percentiles) of the estimators. Therefore, we obtain estimators that reflect the inherent uncertainty due to the genetic assignment test data being imperfect. A flow chart of this process is included (Figure 1.1 of chapter 1).

### 3.3 Assumption

1. Every animal of the same age of state $h$ living at period $i$ has the same probability of surviving to period $i+1$ and moving to state $q$ by period $i+1$ (homogeneous survival probabilities).

2. Each marked animal present in state $h$ at sampling period $i$ has the same capture probability during every sampling period (homogeneous capture assumption within each state).

3. All members of the super population $N$ entering the same population with the same age at the same time from the same origination have the same entry probability (homogeneous entry probabilities).

4. Marks do not affect the survival or behavior of the animal, are not lost, and are recorded correctly.

5. The state of each animal is assigned without error at each capture occasion.

6. Each animal is independent with respect to survival, transitions, and detection probability.

7. The state each animal at period $i + 1$ is only determined by its state at period $i$ (first-order Markov Process assumption)

8. The definition of survival probability assumes that the survival between two sampling periods depends only on the state at the first of the two periods, (this assumption is
reasonable for models where movement occurs at the end of the interval over which survival is estimated).

9. Here the states means populations and locations, so the multi-state model permits variation among animals in different states, but still assumes homogeneity within each state.

10. A valid genetic assignment test has been carried out for all animals where this origination is unknown. This implies that there are no errors in the genotyping used in the tests (However, there will be uncertainty in the genetic assignments).

### 3.4 Simulations

In our simulation, the study is assumed to be five years with three secondary periods close together each year, two populations with animals transition between these two populations, and two age groups (young and old). Young animals ($0 \text{– } 1$ year old) naturally becomes old animals ($>1\text{ year old}$) one year after their birth. Young animals are assumed to be captured at their natal place and every adult animal has a origination assignment vector. The parameter values take different true values in the simulation, with capture probability taking values 0.5 and 0.2. For the transition rate, the young animals are believed to have lower transition rate than old animals. We also believe that most animals stay in their population all the time, so the values of $\varphi_1^{12}$ and $\varphi_1^{21}$ are set to be very small in our simulation, which are just 0.05 and 0.03 for young animals and 0.07 and 0.05 for old animals. And $\varphi_i^{11}$ and $\varphi_i^{22}$ are 0.25 and 0.4 for young animals and 0.3 and 0.45 for old animals. For the entry probabilities, we also consider two different cases, the first sampling period is assumed to have more new recruits than later sampling periods and there are always more in situ births than immigrants. This is a reasonable assumption in practice because usually most animals stay in the same population throughout their lives. Here we only pasted part of the simulation results which can represent all the simulation results.

All simulation results are listed in Table 3.2, Figure 3.1, Figure 3.2, and Figure 3.3. We first want to test the performance of the multi-state super population capture-recapture model. Then we want to compare the multi-state super population capture-recapture model augmented with perfect origination information (perfect model) and the
multi-state super population capture-recapture model (standard model). From Table 3.2 we can see that nearly all relative RMSE of the entry probability under the standard model are less than 20% even when the parameter take values as low as $p_{ij}^{Yh} = 0.2, \varphi_{i}^{Yhh} = 0.25, \varphi_{i}^{Yhq} = 0.03, \varphi_{i}^{Ohh} = 0.2, \varphi_{i}^{Ohq} = 0.05, h \neq q$. When compared to the standard model, the perfect model can estimate $\beta_{0}^{O1B1}, \beta_{0}^{O1B2}, \beta_{0}^{O1I1}, \beta_{0}^{O2B1}, \beta_{0}^{O2I2}$ and $\beta_{0}^{O2I}$ separately with amazing accuracy (maximum relative RMSE is 7.87% under the four different situations listed in Table 3.2). Further, the perfect model has lower relative RMSE for all the estimates of $\beta_{i}^{OhI}, i = 1, 2, 3, 4, h = 1, 2$. The last segment of the lines in Figure 3.1 all have positive slope, which means that the perfect model has better transition probability estimators than standard model. For the estimate of the capture probability (Figure 3.2), for most captured probability, the perfect model has better or comparable good estimate as standard model. Under both perfect model and standard model, the maximum relative RMSE of captured probability is just 14.63%.

We are also interested how the estimates are affected by the reliable levels of the origination information. So we tried four different levels of uncertainty in the imperfect genetic assignment tests, the first as the perfect case assumes the origination assignment probability $\pi_{h}^{j} 1$ is one given the adult animal is an in situ birth of population $h$ and the genetic assignment probability $\pi_{j}^{I}$ is one given the adult animal is a system immigrant, the second assumes the origination assignment probability $\pi_{h}^{j}$ is uniform between (0.95, 1) given the adult animal is an in situ birth of population $h$ and the origination assignment probability $\pi_{j}^{I}$ is uniform between (0.95, 1) given the adult animal is a system immigrant, the third assumes the genetic assignment probability $\pi_{j}^{h}$ is uniform between (0.9, 1) given the adult animal is an in situ birth of population $h$ and the origination assignment probability $\pi_{j}^{I}$ is uniform between (0.9, 1) given the adult animal is a system immigrant, and the fourth assumes the origination assignment probability $\pi_{j}^{h}$ is uniform between (0.8, 1) given the adult animal is an in situ birth of population $h$ and the origination assignment probability $\pi_{j}^{I}$ is uniform between (0.8, 1) given the adult animal is a system immigrant. When one value of the origination assignment probability vector $(\pi_{j}^{1}, \pi_{j}^{2}, \pi_{j}^{I})$, like $\pi_{j}^{1}$ is determined, we can decide the values of the left two by letting $\pi_{j}^{2} \sim U(0, (1 - \pi_{j}^{1})/2)$ and $\pi_{j}^{I} = 1 - \pi_{j}^{1} - \pi_{j}^{2}$.

As the origination assignment procedures increase its certain level, the relative bias of the estimates have a decrease trend (Figure 3.4). While the standard errors of the estimates under four different uncertain assignment level models are very close to each
Figure 3.1: Relative root of mean square error (RMSE) of the estimates of the transition probability based on 1000 simulations. The upper left picture corresponds to parameter values of $p_{ij}^{Yh} = 0.5, p_{ij}^{Oh} = 0.2, \varphi_i^{Yhh} = 0.4, \varphi_i^{Yhq} = 0.05, \varphi_i^{Ohh} = 0.45, \varphi_i^{Yhq} = 0.07$, the upper right picture corresponds to parameter values of $p_{ij}^{Yh} = 0.5, p_{ij}^{Oh} = 0.2, \varphi_i^{Yhh} = 0.25, \varphi_i^{Yhq} = 0.03, \varphi_i^{Ohh} = 0.3, \varphi_i^{Yhq} = 0.05$, the lower left picture corresponds to parameter values of $p_{ij}^{Yh} = 0.2, p_{ij}^{Oh} = 0.5, \varphi_i^{Yhh} = 0.4, \varphi_i^{Yhq} = 0.05, \varphi_i^{Ohh} = 0.45, \varphi_i^{Yhq} = 0.07$, and the lower right corresponds to parameter values of $p_{ij}^{Yh} = 0.2, p_{ij}^{Oh} = 0.5, \varphi_i^{Yhh} = 0.25, \varphi_i^{Yhq} = 0.03, \varphi_i^{Ohh} = 0.3, \varphi_i^{Yhq} = 0.05$. Each line of the pictures represents the relative RMSE of a parameter under five different models. From left to right, the x-axis indicates the models are multi-state super population capture-recapture with imperfect origination assignment information and the uncertain levels are (0.8, 1), (0.9, 1), (0.95, 1), 1. The “standard model” corresponds to the multi-state super population capture-recapture model without origination information.
Figure 3.2: Relative root of mean square error (RMSE) of the estimates of the secondary capture probability based on 1000 simulations. The upper left picture corresponds to parameter values of $p_{ij}^Y = 0.5, p_{ij}^O = 0.2, \varphi_i^{Yhh} = 0.4, \varphi_i^{Yhq} = 0.05, \varphi_i^{Ohh} = 0.45, \varphi_i^{Yhq} = 0.07$, the upper right picture corresponds to parameter values of $p_{ij}^Y = 0.5, p_{ij}^O = 0.2, \varphi_i^{Yhh} = 0.25, \varphi_i^{Yhq} = 0.03, \varphi_i^{Ohh} = 0.3, \varphi_i^{Yhq} = 0.05$, the lower left picture corresponds to parameter values of $p_{ij}^Y = 0.2, p_{ij}^O = 0.5, \varphi_i^{Yhh} = 0.4, \varphi_i^{Yhq} = 0.05, \varphi_i^{Ohh} = 0.45, \varphi_i^{Yhq} = 0.07$, and the lower right corresponds to parameter values of $p_{ij}^Y = 0.2, p_{ij}^O = 0.5, \varphi_i^{Yhh} = 0.25, \varphi_i^{Yhq} = 0.03, \varphi_i^{Ohh} = 0.3, \varphi_i^{Yhq} = 0.05$. Each line of the pictures represents the relative RMSE of a parameter under five different models. From left to right, the x-axis indicates the models are multi-state super population capture-recapture with imperfect origination assignment information and the uncertain levels are $(0.8, 1), (0.9, 1), (0.95, 1), 1$. The “standard model” corresponds to the multi-state super population capture-recapture model without origination information.
other (Figure 3.5). This is reasonable. The in situ birth entry probability is higher than the immigrant entry probability. The more uncertain of the origination information, more in situ births are assigned to as immigrants then the immigrants assigned to in situ birth. However, the standard error should not depend on the uncertainly level of the origination information. From Figure 3.1, Figure 3.2, and Figure 3.3, we can see that as the origination information become more reliable, the estimates of the parameters become better with lower relative RMSE, which is mainly caused by the deceasing bias trend.

3.5 Example: Kangaroo Rats in Southern Arizona

We use data from a capture-recapture study added with genetic assignment test information of banner-tailed kangaroo rats, carried out in the San Simon and San Bernardino valleys in extreme SE Arizona by Waser and Dewoody of Purdue University. Banner-tailed kangaroo rats construct and are entirely dependent on conspicuous mounds, 1 – 3m in diameter, in which they live (singly, for up to 6 years). This characteristic means that populations can be surveyed quickly and unambiguously, simply by locating mounds and assessing them for signs of activity. Mound distribution with the kangaroo-rat site is patchy; clusters of mounds within 5 – 30m of each other are separated by uninhabited stretches 200 – 1000m across. The biologist considers these as “populations”, however, because they are linked by a small but measurable rate of movement. In their study, they sampled a set of eight populations using a capture-recapture method utilizing the Robust Design for years 1994 – 2001. They report data from one primary period each year in March which is at the end of the breeding season. During each primary period, they trapped all active mounds in each population on three secondary periods. For each trapped animal, they marked the animal and recorded the trapping location, sex, and age (“juvenile” for animal less than one year old, “adult” for animals older than one yea). From their study, they found that most juveniles were still at their natal mounds. This shows that our assumption that young animals are captured at their natal place is a reasonable one. They also noticed that the local populations were linked by occasional dispersal. Here we focus on the data sampled at two of eight populations, population 1 and 2, with two age groups, and use data from 1994 to 2001. We consider a two age groups and two population model.

In their studies, they use genetic assignment test to get the origination assign-
Figure 3.3: Relative root of mean square error (RMSE) of the estimates of the entry probability based on 1000 simulations. The upper left picture corresponds to parameter values of $p_{ij}^{Yh} = 0.5$, $p_{ij}^{Oh} = 0.2$, $\psi_i^{Yhh} = 0.4$, $\psi_i^{Yhq} = 0.05$, $\psi_i^{Ohh} = 0.45$, $\psi_i^{Yhq} = 0.07$, the upper right picture corresponds to parameter values of $p_{ij}^{Yh} = 0.5$, $p_{ij}^{Oh} = 0.2$, $\psi_i^{Yhh} = 0.25$, $\psi_i^{Yhq} = 0.03$, $\psi_i^{Ohh} = 0.3$, $\psi_i^{Yhq} = 0.05$, the lower left picture corresponds to parameter values of $p_{ij}^{Yh} = 0.2$, $p_{ij}^{Oh} = 0.5$, $\psi_i^{Yhh} = 0.4$, $\psi_i^{Yhq} = 0.05$, $\psi_i^{Ohh} = 0.45$, $\psi_i^{Yhq} = 0.07$, and the lower right corresponds to parameter values of $p_{ij}^{Yh} = 0.2$, $p_{ij}^{Oh} = 0.5$, $\psi_i^{Yhh} = 0.25$, $\psi_i^{Yhq} = 0.03$, $\psi_i^{Ohh} = 0.3$, $\psi_i^{Yhq} = 0.05$. Each line of the pictures represents the relative RMSE of a parameter under four different models. From left to right, the x-axis indicates the models are multi-state super population capture-recapture with imperfect origination assignment information and the uncertain levels are (0.8, 1), (0.9, 1), (0.95, 1), 1.
Figure 3.4: Relative bias of the estimates of total 76 parameters based on 1000 simulations. The upper left picture corresponds to parameter values of $p_{ij}^Y = 0.5, p_{ij}^O = 0.2, \phi_i^{Yh} = 0.4, \phi_i^{Oh} = 0.05, \phi_i^{Yhq} = 0.45, \phi_i^{Yhh} = 0.07$, the upper right picture corresponds to parameter values of $p_{ij}^Y = 0.5, p_{ij}^O = 0.2, \phi_i^{Yh} = 0.25, \phi_i^{Oh} = 0.03, \phi_i^{Ohh} = 0.3, \phi_i^{Yhq} = 0.05$, the lower left picture corresponds to parameter values of $p_{ij}^Y = 0.2, p_{ij}^O = 0.5, \phi_i^{Yh} = 0.4, \phi_i^{Oh} = 0.05, \phi_i^{Yhq} = 0.45, \phi_i^{Yhh} = 0.07$, and the lower right corresponds to parameter values of $p_{ij}^Y = 0.2, p_{ij}^O = 0.5, \phi_i^{Yh} = 0.25, \phi_i^{Yhq} = 0.03, \phi_i^{Ohh} = 0.3, \phi_i^{Yhq} = 0.05$. Each line of the pictures represents the relative bias of a parameter under four different models. From left to right, the x-axis indicates the models are multi-state super population capture-recapture with imperfect origination assignment information and the uncertain levels are $(0.8, 1), (0.9, 1), (0.95, 1), 1.$
Figure 3.5: Relative standard error (S.E.) of the estimates of total 76 parameters based on 1000 simulations. The upper left picture corresponds to parameter values of $p_{ij}^{Yh} = 0.5, p_{ij}^{Oh} = 0.2, \varphi_i^{Yhh} = 0.4, \varphi_i^{Yhq} = 0.05, \varphi_i^{Ohh} = 0.45, \varphi_i^{Yhq} = 0.07$, the upper right picture corresponds to parameter values of $p_{ij}^{Yh} = 0.5, p_{ij}^{Oh} = 0.2, \varphi_i^{Yhh} = 0.25, \varphi_i^{Yhq} = 0.03, \varphi_i^{Ohh} = 0.3, \varphi_i^{Yhq} = 0.05$, the lower left picture corresponds to parameter values of $p_{ij}^{Yh} = 0.2, p_{ij}^{Oh} = 0.5, \varphi_i^{Yhh} = 0.4, \varphi_i^{Yhq} = 0.05, \varphi_i^{Ohh} = 0.45, \varphi_i^{Yhq} = 0.07$, and the lower right corresponds to parameter values of $p_{ij}^{Yh} = 0.2, p_{ij}^{Oh} = 0.5, \varphi_i^{Yhh} = 0.25, \varphi_i^{Yhq} = 0.03, \varphi_i^{Ohh} = 0.3, \varphi_i^{Yhq} = 0.05$. Each line of the pictures represents the relative S.E. of a parameter under four different models. From left to right, the x-axis indicates the models are multi-state super population capture-recapture with imperfect origination assignment information and the uncertain levels are (0.8, 1), (0.9, 1), (0.95, 1), 1.
ment information of the animals. In genetic assignment test, genotyping at autosomal and 2 sex-linked loci of approximate 95% of the trapped animals are completed. We use GENECLASS2 (Piry et al. 2004) to analyze the collected genotype information to estimate the origination assignment probabilities. For each adult animal with genetic information collected, we calculate its specific genetic assignment probability vector by GENECLASS2, $(\pi_{1j}, \pi_{2j}, \pi_{Ij} = 1 - \pi_{1j} - \pi_{2j}, j = 1, \cdots, \sum_{i,h} u_i^{Oh})$. The genetic assignment test implemented by GENECLASS2 is not perfect, with $0 < \pi_{1j}, \pi_{2j}, \pi_{Ij} < 1$. We also handled the data as though it were “perfect” data although this is not realistic. If $\pi_{hj} = \max(\pi_{1j}, \pi_{2j}, \pi_{Ij})$, we assign the adult animal originates as an in situ birth of population $h$ by assigning $\pi_{hj} = 1, h = 1, 2$. If $\pi_{Ij} = \max(\pi_{1j}, \pi_{2j}, \pi_{Ij})$, we assign this adult animal to be an system immigrant by assigning $\pi_{Ij} = 0$. We also consider capture-recapture data without including the genetic model, which is a standard multi-state super population capture-recapture model. All estimate results are in Figure 3.6, Figure 3.7, Figure 3.8, and Table 3.8.

From all these figures, we can see that the estimates for all the parameters of the three different models are close to each other. We can also see that the capture probability are high for most sampling periods (Figure 3.6). From Figure 3.7, the estimates of $\varphi_i^{rh}, r = Y, O; h = 1, 2$ are much higher than the estimates $\varphi_i^{rhq}, r = Y, O; h \neq q$. This proves the fact the most animals stay in their population and the disperse rate between the populations are very low. For the estimates of the entry probabilities (Figure 3.8), there are much more in situ births ($\varphi_i^{YhB}, h = 1, 2$) than immigrants ($\varphi_i^{OhI}, h = 1, 2$) for all periods. For the old animals, the entry probabilities of the initial period are $0.1092, 0.1109,$ and $0.1105$ for perfect assignment, imperfect assignment, and standard model respectively, which are much higher than those of all other sampling periods of whom maximum is $0.0219$. All these should be what we expect because for the initial sampling, the new recruits are all the animals entering the population before the first sampling period and survive to the first sampling period. From Table 3.8, we can also see that with the origination assignment information, we are able to estimate $\beta_0^{O1B1}, \beta_0^{O1B2}, \beta_0^{O1I}, \beta_0^{O2B1}, \beta_0^{O2B2}$ and $\beta_0^{O2I}$ separately, whereas this cannot be achieved by standard model.

To compare the $\beta_0^{Oh}, \tilde{H} = 1, 2$ under the models with origination information or not, we define $eta_0^{Oh}(\text{genetic model}) = \sum_{q=1,2} \beta_0^{OhBq} \text{ genetic model } + \beta_0^{OhI} \text{ genetic model }, h = 1, 2$. From Figure 3.8, we can see that $\beta_0^{Oh}(\text{perfect genetic model}) \approx \beta_0^{Oh}(\text{imperfect genetic model}) \approx \beta_0^{Oh}(\text{standard model}), h = 1, 2$. 
Figure 3.6: Compare capture probabilities and the 95% confidence interval of its estimates of the Kangaroo rat data set under three different models, multi-state super population capture-recapture models with perfect origination assignment information (perfect genetic), imperfect genetic assignment test (imperfect genetic), and without genetic assignment test (standard model).
Figure 3.7: Compare the transition probability of Kangaroo rat data set under three different models, multi-state super population capture-recapture models with perfect origination assignment information (perfect genetic), imperfect genetic assignment test (imperfect genetic), and without genetic assignment test (standard model). For the upper picture, the upper three lines are estimates of $\varphi_{Y22}^i$, the middle three lines are estimates of $\varphi_{Y11}^i$, and the bottom three lines for estimates of $\varphi_{Y12}^i$ and $\varphi_{Y21}^i$. For the bottom picture, the upper three lines are estimates of $\varphi_{O11}^i$, the middle three lines are estimates of $\varphi_{O22}^i$, and the bottom three lines for estimates of $\varphi_{O12}^i$ and $\varphi_{O21}^i$. 
Figure 3.8: Compare the entry probability of Kangaroo rat data set under three different models, multi-state super population capture-recapture models with perfect origination assignment information (perfect genetic), imperfect genetic assignment test (imperfect genetic), and without genetic assignment test (standard model). For the upper two pictures, the upper three lines are estimates of $\hat{\beta}_Y$, the middle three lines are estimates of $\hat{\phi}_e$. For the bottom picture, every three points in the same vertical line are the standard error of the estimates of entry probability under three different models. We define $\hat{\beta}_{Oh} = \sum_{q=1,2} \hat{\beta}_{0h} + \hat{\beta}_{Oh} h = 1, 2$. 
3.6 Discussion and Future Research

In this chapter, we considered a multi-state super population model augmented with both perfect and imperfect genetic assignment data. In the simulation and the real data analysis, we only use examples with two populations and two age classes. The approach and computer procedures can be easily applied to multiple populations and multiple age groups data set. In the models, we assume homogeneous capture probability among animals of the same age and location being captured at the same primary period. However, in practice, these assumptions may be violated. Inherent characteristics like gender and age can affect the secondary period capture probabilities (heterogeneity). Also trap response can affect future secondary capture probabilities (trap response). We can introduce time, heterogeneity and trap response models into the last component of the likelihood. We have allowed that the immigrants have different survival probability from the in situ births and could also allow that immigrants have different capture probability from the in situ births. However, the more general of the model, the more parameters needed to be estimated. The super population size needs to be big enough to provide enough recaptured animals to make the parameters estimable. For the kangaroo data analysis, we only constructed a full model where all parameters are different from each year. To improve the precision of the estimates, we can reduce the number of parameters by adding constraints to the parameters. Then we can use AIC or BIC to choose a best model for the data set. The young kangaroo rats rarely move out of their natal mound at their first year “only once in several hundred trap nights (Waser, Personal communication).” To test the validation of the genetic assignment test, we assign the genetic assignment probability to a sample of young animals which are all in situ births. If the genetic assignment test is a good one, the genetic assignment probability should be very high, like bigger than 0.95, as the bigger the better the genetic assignment test. However, when we use GENECLASS2 to analyze the genetic information of the sample of young animals, we found that the assignment methods “generally result in an excess of resident individuals being excluded”, i.e., scored as immigrants. Waser (personal communication) suspects that the biological reason for this is that immigrants tend to have alleles that are common in their birth population but locally rare in the population they move to. However they pass some of those alleles to their offspring, and their offsprings progeny, etc who then may be incorrectly judged to
be immigrants too. Thus the assignment procedure is biased towards “immigration”. For example, when we run GENECLASS with data from the population we are interested in vs. everything else, it correctly assigns about 70% of the animals first captured in population to population. A few percent may really come from outside, but the true number is surely far below 30%, as many of these individuals were first captured as small juveniles and/or have known parents in population. Waser suggested one way to deal with this problem might be to use assignment test results only when it is relatively unambiguous, for example for individuals that are \(10x\) as likely to come from one population than the other(s), and to treat the other individuals as having no genetic information. We hope to explore this and other ideas in future work. Based on genetic parentage determination rather than assignment tests, Waser (personal communication) found that for animals whose parents can be determined, only 7% live as adults in a different population than their parents. So in our real data analysis here, there is only a very small proportion of tagged animals (at most a few percent) are estimated to disperse between populations (Skvarla and Nichols, 2004). Therefore the entry probabilities from outside should be really small. However, in the simulation part, we assume that the entry probabilities of the immigrants are reasonably large to achieve reasonable precision. Another difficulty of analyzing the kangaroo data set is that only a very small proportion of tagged animals (at most a few percent) are estimated to disperse between populations (Skvarla and Nichols, 2004). So it is difficult to detect this dispersion process between populations with traditional capture-recapture tags. In the simulation part, the dispersal rates are small as 0.03 and 0.05 for young animals and 0.05 and 0.07 for old animals. The estimates of dispersal rates have the highest RRMES of all estimates, with most having RRMSE bigger than 90%. Ecologists are still interested in these dispersal processes. To detect these rare movements, we could use telemetry tags as illustrated in chapter 1 to trace the movement process of a sample of animals, which will improve the estimation of transition rate among populations. Telemetry allows estimation of the two components of the transition probability which are survival and movement rates. Using Otolith microchemistry on trace elements for fish species gives a similar augmentation to genotype data. The idea is that fisheries biologists can tell which natal stream an anadromous fish came from by changes in the trace element signatures. In fact by looking at different parts of the otolith it may be possible to know not only the natal stream but the whole history of every stream it has visited since birth (Ashford and Jones, 2007). If the
genetic assignment test wrongly assigns the animal to immigrant when the animal is truly an in situ birth or vice versa, misclassification happens. Yoshizaki et al. (2009) discussed the effect of misclassification in DNA finger printing closed population capture-recapture studies. We may be able to use similar ideas to estimate the effect of misclassification of genetic assignment tests to our open population model estimates here.
Table 3.1: Relative RMSE (%) of estimates of entry probabilities under multi-state super population capture-recapture model (standard model) and multi-state super population capture-recapture model augmented with perfect origination assignment information. The results are based on 500 simulations. “P” column corresponds to model with perfect origination information, “S” column corresponds standard model, and “Val” columns is the true parameter value.

<table>
<thead>
<tr>
<th>Par</th>
<th>Val</th>
<th>P</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p_{ij}^{yh} = 0.5, p_{ij}^{Oh} = 0.2, \varphi_{ih}^{YY} = 0.4, \varphi_{ih}^{Yh} = 0.05, \varphi_{ih}^{Ohh} = 0.45, \varphi_{ih}^{Yhq} = 0.07 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \beta_{1}^{1B} )</td>
<td>0.04</td>
<td>6.04</td>
<td>5.35</td>
</tr>
<tr>
<td>( \beta_{1}^{1B} )</td>
<td>0.02</td>
<td>14.93</td>
<td>22.23</td>
</tr>
</tbody>
</table>
Table 3.2: Relative RMSE (%) of estimates of entry probabilities under multi-state super population capture-recapture model (standard model) and multi-state super population capture-recapture model augmented with perfect origination assignment information. The results are based on 500 simulations.

<table>
<thead>
<tr>
<th>Par</th>
<th>Estimate</th>
<th>S.E.</th>
<th>Estimate</th>
<th>S.E.</th>
<th>Estimate</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_{O1B1}$</td>
<td>0.0502</td>
<td>0.0121</td>
<td>0.0537</td>
<td>0.0142</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_{O1B2}$</td>
<td>0.011</td>
<td>0.0061</td>
<td>0.01</td>
<td>0.0062</td>
<td>0.0716</td>
<td>0.0159</td>
</tr>
<tr>
<td>$\beta_{O2I}$</td>
<td>0.0065</td>
<td>0.0043</td>
<td>0.0069</td>
<td>0.0046</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_{O2B1}$</td>
<td>0.0082</td>
<td>0.004</td>
<td>0.0084</td>
<td>0.0049</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_{O2B2}$</td>
<td>0.0299</td>
<td>0.0094</td>
<td>0.029</td>
<td>0.0114</td>
<td>0.0389</td>
<td>0.0091</td>
</tr>
<tr>
<td>$\beta_{I2}$</td>
<td>0.0034</td>
<td>0.0028</td>
<td>0.0029</td>
<td>0.0023</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Bibliography


Chapter 4

Evaluation of Marine Reserves for Fisheries using Tag-Return, Tag-Recapture and Telemetry Models: Estimation of Patch-Specific Fishing Mortality, Natural Mortality and Movement Rates

Summary Currently there is great interest in the value of marine reserves for the conservation of important fisheries. Here we focus on tag-return models for two-site model with one area a marine reserve and one area a fishing area. The multi-state tag-return models developed by Joe (Ph.D. dissertation, 2001) are very useful provided that we have an estimate of tag-reporting rate and they also allow us to estimate patch specific fishing and natural mortality rates. However, these models are critically dependent on the tag-reporting rate. They also do not provide direct information about natural deaths. Telemetry tagging methods do provide direct information about natural deaths; (Hightower
et al., 2001, Pollock, et al., 2004) and there is no issue of tag-reporting for these special kinds of tags. In this paper, we will generalize the combination of telemetry and tag-return methods to multi-site populations. We will consider tag-return, tag-recapture and telemetry models in various combinations for two-site models where one area is a marine reserve and one is subject to regular fishing. Then we illustrate our methods with a comprehensive simulation study.

4.1 Introduction

Modern tag-return models for estimating survival rates and mortality rates were derived from the work of Seber (1970), Brownie (1973), Youngs and Robson (1975). An important synthesis was the monograph by Brownie et al. (1985). These models focused on the estimation of total mortality rates. Hoenig et al. (1998a,b) extended these models to fisheries applications and showed how it was possible to separately estimate fishing mortality rate (F) and natural mortality rate (M) if an external tag-reporting rate estimate (perhaps from using a high-reward tagging study or a planted tags study) was available. Multi-state fishery tag-return models are used for fisheries that have multiple sites or patches with movement possible between sites. These methods allow the estimation of patch-specific natural and fishing mortality rates and movement rates between patches based on methods of separating survival and movement rates in multi-state tag-return and capture-recapture models. Important work was developed in Joe (2001) and Joe and Pollock (2002). The models can be viewed as multiple site extensions of the single site tag-return models in Hoenig et al. (1998 a, b) (once- a-year tagging) and Hearn et al. (1998) (twice-a-year tagging). They can also be viewed as extensions of models for multiple patch systems using tag-return and tag-recapture methods (Chapman and Junge, 1956, Darroch, 1962, Arnason, 1972, Schwarz et al., 1993). The work in Joe (2001) first focused on tag-return models for a two-patch system where fishing and hence tag-returns can occur in both patches. However, there is currently great interest in the value of marine reserves for the conservation of important fisheries and therefore Joe (2001) also considered a two-site model with one area a marine reserve and one area a fishing area. She assumed that capture-recapture catch and release sampling was used in a marine reserve closed to regular fishing and the usual tag-return sampling in the area subject to regular fishing. She considered pulse fisheries and fisheries
that operate over part of the year with the assumption that movement occurs outside the fishing season of each year. Based on simulations most mortality and movement estimates have reasonable precision ($< 30\%$ Relative Standard Error) when movement occurs at a high rate, whereas precision of patch-specific natural mortality and reporting rates are often of poor precision when movement rates are low. Since there are more capture histories when using capture-recapture sampling in the reserve, she found that precision of estimates of the two-site model with capture-recapture and tag-return sampling model are better than in a two-site tag-return sampling model for the same values of the parameters. The multi-state tag-return models developed are very useful provided that we have an estimate of tag-reporting rate and they also allow us to estimate patch specific fishing and natural mortality rates. However, these models are critically dependent on the tag-reporting rate. They also do not provide direct information about natural deaths. Telemetry tagging methods do provide direct information about natural deaths; (Hightower et al., 2001, Pollock, et al., 2004) and there is no issue of tag-reporting for these special kinds of tags. Pollock et al. (2004) combined the telemetry tagging method with traditional tag-return models. They showed that in the combination, the precision of the estimates were much improved relative to those of the tag return and telemetry methods separately. Bacheler et al. (2008) applied these methods to a red drum fishing in North Carolina. In this paper, we will generalize these combined telemetry and tag-return methods to multi-site populations. We will consider tag-return, tag-recapture and telemetry models in various combinations for two-site models where one area is a marine reserve and one is subject to regular fishing. The specific models we develop and compare are:

Model 1: Tagged fish are released in the marine reserve but the only way these fish are returned is if they move out of the reserve into the fishing area so that they provide regular fishing tag-returns. Tagged fish are also released in the fishing area where they are subject to regular fishing tag-returns. (We will show how this model has serious limitations which motivates why we consider the other models).

Model 2: Tagged fish are released in the marine reserve and these fish may be returned is if they move out of the reserve into the fishing area so that they are subject to regular fishing pressure. Any tagged fish in the marine reserve will also be subject to capture-recapture from special research sampling. Tagged fish are also released in the fishing
area where they are subject to regular fishing tag-returns. And these tagged fish can also be recaptured in marine reserve given that they move out of the fishing area into the marine reserve.

Model 3: The fishing area and the marine reserve both have special telemetry tags used on fish so that detailed information on mortality and movement can be obtained by monitoring these telemetry tagged fish.

Model 4: The fishing area and the marine reserve now have special telemetry tagged fish subject to monitoring and regular tagged fish subject to tag-returns (model 4a) or subject to tag-returns and special recapture (model 4b). Therefore models 4 can be viewed as a combination of models 1 and 3 (model 4a) or a combination of model 2 and 3 (model 4b).

We develop likelihood and maximum likelihood estimators for each model developed and illustrate our methods with a comprehensive simulation study. We conclude with a general discussion and suggestion for future work.

### 4.2 Notation

\[ N_r^i \] = number of tagged fish released at the beginning (pre-season) of the \( i^{th} \) year in area \( r \).

\[ N_{r,post}^i \] = number of tagged fish released at the second tagging event (post-season) time of the \( i^{th} \) year in area \( r \).

\[ R_{r,ij}^p \] = number of fish released in area \( r \) in year \( i \) at pre-season tagging event that are recovered (or recaptured) in area \( s \) in year \( j \) \((i \leq j)\).

\[ R_{r,post}^{p,s} \] = number of fish released in area \( r \) in year \( i \) at post-season tagging event that are recovered (or recaptured) in area \( s \) in year \( j \) \((i \leq j)\).

\[ T \] = the length of the fishing season \((0 < T < 1)\), here we assume we have a continuous fishery from the beginning of each year \((0)\) to time \( T \).

\[ t \] = time in year \( i \) at which time point the fishing migration happens, we assume that the fish move at just one fixed time of the year with \( t \geq T \).
\( F_i^r \) = fishing mortality rate of fishing season \([0, T]\) in area \( r \) and year \( i \), for our model here, we have \( F_i^A = 0 \) in the marine reserve and \( F_i^B > 0 \).

\( M^r \) = instantaneous natural mortality rate of fish in area \( r \) in any year.

\( \lambda^r \) = tag reporting rate given that a tagged fish is harvested in area \( r \).

\( \psi_i^{rs} \) = the probability that a fish moves from area \( r \) to \( s \) given that it is alive in year \( i \) in area \( r \), with \( \psi_i^{rr} + \psi_i^{rs} = 1 \) and \( \psi_i^s + \psi_i^{sr} = 1 \).

\( p_i^r \) = the probability of being captured in year \( i \) given that a tagged fish is alive in area \( r \), in year \( i \), for two areas, \( p_i = \left[ \begin{array}{l} p_i^1 \\ p_i^2 \end{array} \right] \).

\( (s_i^r)^t \) = the probability that the fish survives from time \( t = 0 \) to time \( t \) in area \( r \) given that it is alive in area \( r \) in time period \( i \).

\( u_i^r \) = the probability that the fish harvested in area \( r \) in year \( i \) given that it is alive at the beginning of year \( i \) in area \( r \) and there is a fishing season in area \( r \).

\( \phi_i^{rs} \) = the probability of being alive in area \( s \) in year \( i + 1 \) given that it alive and in area \( r \) in year \( i \).

\( NF_i^A = 1 - e^{-M_i^A*T} \) be the probability that the fish in area \( A \) died naturally in time period \([0, T]\) given that there is no fishing season in the reserve.

\( S^B = e^{-M^B*T} \) : the probability that the fish in area \( B \) surviving from natural \([0, T]\) mortality in time interval.

\( NF_i^B = 1 - e^{-M^B*T} \); the probability for the fish in area \( B \) died naturally in time period \([0, T]\) given no fishing pressure in this time period in area \( B \).

\( S_i^{BF} = e^{-(F_i^B + M^B*T)} \) : Survival probability for the fishing season in area \( B \).

\( NF_i^{BF} = (1 - e^{-(F_i^B + M^B*T)}) * \frac{M^B*T}{F_i^B + M^B*T} \) : the natural mortality of the \( i^{th} \) fishing season in area \( B \).
4.3 Model Development

We assume that there are two areas area $A$ and $B$. The fish can move from $A$ to $B$ and vice versa. Area $A$ as a marine reserve area, has no fishing pressure ($F_A = 0$), however area $B$ is under regular fishing pressure ($F_B > 0$). We consider four different samplings models (model 1 to model 4) which have various combinations of tag-returns, tag-recaptures, and telemetry. We now present the details for each model.

4.3.1 Model 1

In model 1, for area $A$, tagged fish are released in the reserve area $A$ but the only way these fish can be returned is if they move out of the reserve to area $B$ so that they are subject to regular fishing pressure in area $B$. For area $B$, tagged fish can be returned as usual by tag-returns during the fishing season which lasts from the beginning of each year to $T$. Movement in both areas is assumed to take place one time each year and this movement happens after the fishing season. The number of fish that are released in both area $A$ and area $B$ in year 1 and recovered in area $B$ in future fishing seasons can be displayed as shown in table 1. This is for an example with pre and post season tagging in the first year and pre season tagging in the second year with recovery in both years. For this example, tag and release happen at time $t = 0, T, 1 + T$ in both areas $A$ and $B$, and tag-return happens between $(0, T)$ and $(1, 1 + T)$ in area $B$. From table 1, we can see, the tagged fish released in area $A$ can be returned if they move out to area $B$ and are recovered. For example, is the number of fish released in area $A$ at first season in year 1, survived in area $A$ from time zero to time $t$, then migrated to area $B$, survived in area $B$ from time $t$ to the end of year 1, and finally harvested in area $B$ at the second fishing season with the tag be reported.

In this model, we must assume $M_A = M_B = M$ to make all parameters estimable. We can get the estimates of all these parameters separately by maximizing the likelihood function above. But since there is no fishing pressure on area $A$, the only way the tagged fish released in area $A$ can be recovered is through moving to area $B$ and be fished, which means we have little information about the fish in area $A$. We shall see later in our simulations that this also means that the precision of estimates is poor in many scenarios. To estimate the patch specific natural mortality, $M_A$ and $M_B$ separately, we next apply special research capture-recapture sampling on the tagged fish of area $A$ which leads to model 2.
Table 4.1: Symbolic representation for model 1 with 3 tagging periods and two fishing seasons.

<table>
<thead>
<tr>
<th>Area release released</th>
<th>Year released</th>
<th>Season released</th>
<th>Number released</th>
<th>Year 1, area A</th>
<th>Year 2, Area B</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>Pre-season</td>
<td>$N_1^A$</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Post-season</td>
<td>$N_1^{Apost}$</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>Pre-season</td>
<td>$N_1^B$</td>
<td>$R_{11}^{BB}$</td>
<td>$R_{12}^{BB}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Post-season</td>
<td>$N_1^{Bpost}$</td>
<td>-</td>
<td>$R_{12}^{Bpost}$</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Pre-season</td>
<td>$N_2^A$</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

4.3.2 Model 2

This model is an extension of models in Schwarz et al. (1993) and Joe (2001). Schwarz et al. (1993) considered the band-recovery model for two states or sub-populations. Joe (2001) considered band-recovery model in two areas, one area is a marine reserve and the other is under regular fishing pressure. In this paper, we revisit Joe’s models and reformulate them in a slightly different way. She used continuous capture-recapture in area A, while we use a pulse capture-recapture model in area A, that is recapture in area A only happens directly after the tagged fish are released in the marine reserve and these fish may also be returned if they move out of the reserve to area B so that they are subject to regular fishing pressure. Any tagged fish in the marine reserve will also be subject to recapture sampling if they stay in area A until the next recapture period. For area B tagged fish can be returned as usual by harvest. Tagged fish in area B can also be recaptured from area A if they move out from area B to area A. The movement between area A and area B is assumed to be just once a year, which happens after the fishing season of the year. This model is more complex than model 1 because the fish released in area A can either be recaptured from area A by subsequent recapture events or be recovered from area B in future fishing seasons. However, it is a better model than model 1 because it provides more information for parameter estimation.

Consider an example with three tagging periods, and 3 recapture events in area
A and two fishing seasons in area B, where fish are tagged and released at time \( t = 0, T \), and 1, recapture events in area A happen at time \( t = T, 1 \), and \( 1 + T \), the fishing seasons in area B are \([0, T]\) and \([1, 1 + T]\) as shown in figure 1, and the movement happens once a year between \((T, 1]\) as in figure 1.

Furthermore, condition on \( t \), we assume the numbers of tags returned (recaptured) from both areas A and B, given they are released to area A at the pre-season of first year, follow a multinomial distribution. That is, for fish released, the fish can have been returned or recaptured with any of these recovery histories, AAAA, AAOA, AOAA, AAAO, AOOA, AAOA, AOBO, AAB, and others. “O” means the tag is not observed at the recapture event (fishing season) and “others” means the tagged fish not be recaptured or returned by the end of the study. Conditional on the tag number \( N^A_1 \), data \( \{R^{AAAA}, R^{AAOA}, \ldots, R^{AAB}, R^{Other}\} \) follow a multinomial distribution.

Now we consider the cell expression for \( P^{AAB}_1 \), which is the probability of the fish released in area A in pre-season of year 1, recaptured at time \( t = T \) in area A, then returned from area B at second fishing season, with \( E(R^{AAB}) = N^A_1 * e^{-M^A t} * e^{-M^B (1-t)} * P^A_1 * \lambda^B * u^B_{1} * \psi^{AB}_1 \). Similarly, conditional on the tag number \( N^B_1 \), the recovery data \( \{R^{BB}, R^{BOA}, R^{AOA}, R^{BOB}, R^{BOAO}, R^{BOther}\} \) follows a multinomial distribution. And the recovery data \( \{R^{BB}, R^{BAA}, R^{BOA}, R^{BAO}, R^{BOther}\} \) also represent a multinomial distribution when conditional on the tag number \( N^{B^\text{post}}_1 \). All the fish released in the different events \((N^A_1, N^{A^\text{post}}_1, N^A_2, N^{B_1, N^{B^\text{post}}_1, N^B_2}) \) lead to six different multinomial distributions and the likelihood for this model is the product of these six multinomial distributions. Compared to model 1, all the parameters are estimable in model 2 without the assumption of \( M^A = M^B = M \). In model 2, we can also estimate the season specific capture probability of area A, \( P^A_1, P^A_2 \) and \( P^A_2 \). The complete cell probabilities for each multinomial distribution are in appendix A. In both model 1 and model 2, we use traditional tags on fish, which can not provide direct mortality information about the fish, and also can not provide more detail information about the movements of the fish. In the next models 3 and 4, we consider more expensive telemetry tags, on the fish to provide detail information about mortality and movements of the fish.
Figure 4.1: The depiction of the fishing season and tagging period of area A and area B.
4.3.3 Model 3 (Telemetry Model)

Hightower et al. (2001) and then Pollock et al. (2004) have applied telemetry methods to the estimation of mortality rates of fish populations. Here we extend these models to a multi-state fish population model. There are two areas, area A (marine reserve) and area B subject to fishing. There are movements between these two areas and the movements are assumed to be either at the end of the sixth month of each year (model 3a) or at the end of each month of each year (model 3b). We assume the fishing season of area B is the first month of each year for both model 3a and model 3b. 50 telemetry tagged fish are released to area A and area B respectively at the beginning of each year. For model 3a and model 3b, we assume the relocation probability is certain. That is, if telemetry tagged fish were alive at the beginning of month, it will definitely be detected alive or dead by the telemetry signal by the end of month. Also the movement of the fish has the Markov property, that is, the number of fish migrating from area A to area B (or from area B to area A) only depends on the fish alive in the previous month.

Suppose that we consider $I$ years of tracking telemetered fish and relocate the fish monthly. “If a fish is first relocated alive (i.e., movement was detected between searches) at month $j$ for $j = 1, 2, \cdots, 12*I - 1$, it becomes part of a new (virtual) release at time $j + 1$.” (Pollock, et al. 2004). If for consecutive searches, the fish were found to be in the same location, they will be considered to have dead naturally (Pollock, et al. 2004). Telemetry tags are relocated by telemetry signal at the end of every month. Any movement of the fish (migration of the fish from A to B or from B to A) can be detected by the telemetry method, so we consider two models. Let

\[ N_r[i] = A, B \] be the number of fish relocated alive in area r by the end of the $i^{th}$ month given the fish having telemetry tags.

\[ N_r[i] = A, B \] be the number of telemetry tagged fish living in month $i$ in area $r$.

\[ d_r[i] = A, B \] be the number of telemetry tagged fish relocated dead in area r by the end of the $i^{th}$ month,

\[ R_r[i] = A, B \] be the number of fish released with telemetry tags at the beginning of the $i^{th}$ year,

\[ n_{rs}^i = r, s = A \text{ or } B \] be the number of fish relocated alive at month $j$ in area S, given
that is alive in area \( r \) in the month \( j-1 \).

\[ d_j^r=r, s = A \text{ or } B \text{ be the number of fish relocated dead at month } j \text{ in area } S, \text{ given that it is alive in area } r \text{ in the month } j-1. \]

\[ n_j^r=r, s = A \text{ or } B \text{ be the number of fish neither relocated alive nor dead at month } j \text{ given that it is alive in area } r \text{ in the month } j-1. \]

We have \( N^r[j] = \begin{cases} R^r[1] & j = 1 \\ n^r[12 + i] + R^r[i + 1] & j = 12 \cdot i \\ n^r[j] & j = \text{other} \end{cases} \)

**Model 3a: Migration of fish fixed at the end of the sixth month of each year**

We present the cell probability structure in general and then provide a 13 month example useful for the simulation, where \( P \) is the vector of cell probabilities of a multinomial distribution and \( T = 1 \) month.

\[ n_{rs}^i \text{=the number of fish detected alive in area } s \text{ by the end of } (2 \cdot i - 1) \cdot 6 + 1 \text{ month given that is alive in area } r \text{ by the end of } (2 \cdot i - 1) \cdot 6^{th} \text{ month, } r, s = A \text{ or } B. \]

There is only natural mortality in area \( A \), so for \( j = 12(i-1) + 1, \ldots, 12(i-1) + 6, i = 1, \ldots, I \), we have

\[ (n^A[j], d^A[j]) \sim \text{multinomial}(N^A[j-1], P = (e^{-M^A \cdot T}, 1 - e^{-M^A \cdot T})). \quad (4.1) \]

In area \( B \) there is fishing pressure during the first month of each year, so the multinomial distribution has three cells, with \( n_{j}^{B0} \) be the number of fish not relocated alive or dead at month given that it is alive in area \( B \) at \( j^{th} \) month. Therefore, for \( j = 12 \cdot (i-1) + 1, i = 1, \ldots, I \), we have

\[ (n^B[j], d^B[j], n_{j}^{B0}) \sim \text{multinomial}(N^B[j-1], P = (e^{-M^B \cdot T}, 1 - e^{-M^B \cdot T})) \quad (4.2) \]

Now comes to the sixth month, where migrations of fish happen immediately after the end of the sixth month. Thus for \( J = 6 \cdot (2 \cdot I - 1) + 1, I = 1, \ldots, I \),

\[ (n_i^{AA}, n_i^{AB}, d_i^{AA}, d_i^{AB}, n_i^{A0}) \sim \text{multinomial}(n^A[6 \cdot (2 \cdot i - 1)], P = (P^A, 1 - P^A)) \quad (4.3) \]

where \( P^A = (s_i^{BA} \ast S^A, s_i^{BB} \ast S^B, s_i^{BB} \ast NF^A, s_i^{BB} \ast NF^b) \), the fish relocated alive in the month followed with the month when the movements happen should composed of the fish
migrating from another area or the fish from the same area. So
\[ n^A[j] = n_i^{AA} + n_i^{BA} \]  
\[ n^B[j] = n_i^{AB} + n_i^{BB} \]  

Based on similar logic, we have, for \( J = 12 \ast (i - 1) + 8, 12(I - 1) + 12, i = 1, \cdots, I, \)
\[ (n^A[j], d^A[j]) \sim \text{multinomial}(N^A[j - 1], P = (e^{-M^A \ast T}, 1 - e^{-M^A \ast T})) \]  
\[ (n^B[j], d^B[j]) \sim \text{multinomial}(N^B[j - 1], P = (e^{-M^B \ast T}, 1 - e^{-M^B \ast T})) \]  

Now consider a specific example with 13 months long. For area \( A \) in the first month,
\[ (n^A[1], d^A[1]) \sim \text{multinomial}(R^A[1], P = (S^A_1, 1 - S^A_1)) \]  
For area \( B \) for first month, there is fishing mortality,
\[ (n^B[1], d^B[1], n^B_0) \sim \text{multinomial}(R^B[1], P = (S^B_1F, NF_F, S^B_0)) \]  
with \( S^B_0 = 1 - e^{-(F^B + M^B \ast T)} - (1 - e^{-(F^B + M^B \ast T)}) \ast \frac{M^B \ast T}{F^B + M^B \ast T} \) From \( i = 2, \cdots, 6 \) months, there are only natural deaths in both areas \( A \) and \( B \) since the fishing season lasts for the first month each year, the number of fish alive and dead are conditioned on the fish alive in the pervious month,
\[ (n^A[i], d^A[i]) \sim \text{multinomial}(N^A[i - 1], P = (e^{-M^A \ast T}, 1 - e^{-M^A \ast T})) \]  
\[ (n^B[i], d^B[i]) \sim \text{multinomial}(N^B[i - 1], P = (e^{-M^B \ast T}, 1 - e^{-M^B \ast T})) \]  

At the end of the sixth month the fish in area \( A \) can migrate to area \( B \) and vice versa and the fish relocated alive by the end of the seventh month in area \( A \) and \( B, n^A[7] \) and \( n^B[7] \) come from the fish movement from \( n^A[6] \) and \( n^B[6] \). That is, for \( i = 7, \)
\[ (n^A_6, n^AB_6, d^AA_6, d^AB_6, n^A_{other}) \sim \text{multinomial}(N^A[6], P = (P^A, 1 - P^A)) \]  
\[ (n^B_6, n^BA_6, d^BB_6, d^BA_6, n^B_{other}) \sim \text{multinomial}(N^B[6], P = (P^B, 1 - P^B)) \]
\[ P^A = (\psi^{AA} e(-M^A T), \psi^{AB} e(-M^B T), \psi^{AA} (1 - e(-M^A T)), \psi^{AB} (1 - e(-M^B T))) \] (4.14)

\[ P^B = (\psi^{BB} e(-M^B T), \psi^{BA} e(-M^A T), \psi^{BB} (1 - e(-M^B T)), \psi^{BA} (1 - e(-M^A T))) \] (4.15)

Therefore,

\[ n^A[7] = n_6^{AA} + n_6^{BA} \] (4.16)

\[ n^B[7] = n_6^{AB} + n_6^{BB} \] (4.17)

\[ d^A[7] = d_6^{AA} + d_6^{BA} \] (4.18)

\[ d^B[7] = d_6^{AB} + d_6^{BB} \] (4.19)

and similarly, we get the cell probability structure for the other months. For \( i = 8, \cdots, 13 \),

\[ (n^A[i], d^A[i]) \sim \text{multinomial}(N^A[i - 1], P = (e^{-M^A T}, 1 - e^{-M^A T})) \]

For \( i = 8, \cdots, 12 \), \((n^B[i], d^B[i]) \sim \text{multinomial}(N^B[i - 1], P = (e^{-M^B T}, 1 - e^{-M^B T}))\)

For \( i = 13 \), \((n^B[i], d^B[i], n^0B[i]) \sim \text{multinomial}(N^B[i - 1], P = (S_{2B}^F, NF_{2B}^F, 1 - S_{2B}^F - NF_{2B}^F))\)

Model 3b: Migration between area A and B happens by the end of each month

In practice, the migrations of fish might happen each month. Compared to model 3a, this model includes more information about the movement rate. Therefore model 3b should have more precise estimates for migration rates then model 3a. General case: There is always only natural mortality in area A but fishing pressure in area B at the first month of each year. We also let \( P \) denote the vector cell probabilities of a multinomial distribution. So, for the first month of every year, \( j = 12(i - 2) + 1, i = 1 \cdots, I \),

\[ (n^A_j, n^B_j, d^A_j, d^B_j, n^0_A) \sim \text{multinomial}(N^A[j - 1], P = (P^A, 1 - P^A)) \]

where \( P^A = (\psi_1^{AAS^A}, \psi_1^{ABS^B}, \psi_1^{AANF^A}, \psi_1^{ABNF^B}). \)

\[ (n^A_j, n^B_j, d^A_j, d^B_j, n^0_A) \sim \text{multinomial}(N^B[j - 1], P = (P^B, 1 - P^B)) \]

where \( P^B = (\psi_1^{BAS^A}, \psi_1^{BS^B}, \psi_1^{BANF^A}, \psi_1^{BBNF^B}). \)

For all other months (from 2 to 12 month) of each year, there is only natural
mortality in both areas A and B. So for \( j = 12(i - 2) + 2, \cdots, 12(i - 1) + 12 \),

\[
(n_j^{AA}, n_j^{AB}, d_j^{AA}, d_j^{AB}) \sim \text{multinomial}(N^{A[j - 1]}),
\]

\[
P = (\psi_i^{AA}S^A, \psi_i^{AB}S_B^i, \psi_i^{AA}NF_A^i, \psi_i^{AB}NF_B^i)
\]

\[
(n_j^{BA}, n_j^{BB}, d_j^{BA}, d_j^{BB}) \sim \text{multinomial}(N^{B[j - 1]}),
\]

\[
P = (\psi_i^{BA}S^A, \psi_i^{BB}S_B^i, \psi_i^{BA}NF_A^i, \psi_i^{BB}NF_B^i)
\]

Differently from model 3a, for all \( j \geq 2 \), \( n^A[j] \) is composed of the fish migration from area B (given the fish alive in area B at month \( j - 1 \)) and the fish remain in area A (given the fish alive in area A at month \( j - 1 \)) . Similarly, \( n^B[j] \) also is composed of the fish from area A or from area B. So, for \( j \geq 2 \),

\[
n^A[j] = n^A_{j-1} + n^B_{j-1}
\]

\[
n^B[j] = n^A_{j-1} + n^B_{j-1}
\]

\[
d^A[j] = d^A_{j-1} + d^B_{j-1}
\]

\[
d^B[j] = d^A_{j-1} + d^B_{j-1}
\]

Now consider a specific example with the telemetry study lasting 13 months. For \( i = 1 \) , there are fishing mortality in area B but not in area A,

\[
(n_i^{AA}, n_i^{AB}, d_i^{AA}, d_i^{AB}, n_i^{A0}) \sim \text{multinomial}(R_i^A, P = (P_i^A, 1 - P_i^A))
\]

\[
(n_i^{BB}, n_i^{BA}, d_i^{BB}, d_i^{BA}, n_i^{B0}) \sim \text{multinomial}(R_i^B, P = (P_i^B, 1 - P_i^B))
\]

where

\[
P_i^A = (\psi_i^{AA}S^A, \psi_i^{AB}S_B^i, \psi_i^{AA}NF_A^i, \psi_i^{AB}NF_B^i)
\]

\[
P_i^B = (\psi_i^{BB}S^B, \psi_i^{BA}S_A^i, \psi_i^{BB}NF_A^i, \psi_i^{BA}NF_B^i)
\]

For \( i = 2, \cdots, 12 \), there are only natural mortality in both areas, so

\[
(n_i^{AA}, n_i^{AB}, d_i^{AA}, d_i^{AB}) \sim \text{multinomial}(n^A[i - 1], p_i^A)
\]

\[
(n_i^{BB}, n_i^{BA}, d_i^{BB}, d_i^{BA}) \sim \text{multinomial}(n^B[i - 1], p_i^B)
\]

where \( p_i^A \) and \( p_i^B \) can easily be calculated as in previous.
Similarly to the first month, when $i = 13$,

$$(n^{AA}[13], n^{AB}[13], d^{AA}[13], d^{AB}[13], n^{A0}[13]) \sim \text{multinomial}(P^A + n^A[12]),$$

$$P = (P^A_{13}, 1 - P^A_{13}) \quad (4.32)$$

$$(n^{BB}[13], n^{BA}[13], d^{BB}[13], d^{BA}[13], n^{B0}[13]) \sim \text{multinomial}(P^B + n^B[12]),$$

$$P = (P^B_{13}, 1 - P^B_{13}) \quad (4.33)$$

where

$$P^A_{13} = (\psi^{AA} \ast S^A, \psi^{AB} \ast S^{BF}, \psi^{AA} \ast NFA, \psi^{AB} NFBF) \quad (4.34)$$

$$P^B_{13} = (\psi^{BB} \ast S^{BF}, \psi^{BA} \ast S^A, \psi^{BB} \ast NFBF, \psi^{BA} NFA) \quad (4.35)$$

In model 3, all parameters are estimable with desirable precision. But since telemetry tags are expensive compared to traditional tags, the sample can not be too large. Therefore we can combine both kinds of traditional tags and telemetry tags to get more information and improve the precision of the estimates, which is the goal of model 4.

### 4.3.4 Model 4 (Combines Regular Tags and Telemetry Tags)

In this model 4, both types of tags (regular tags and telemetry tags) are used. Here we only consider two cases: (1) combine model 1 with model 3a, 3b, (2) combine model 2 with model 3b. Here, we assume the released fish with different type of tags are independent with respect to capture, survival, mortality and migration. The likelihood is the product of the likelihood for model 1 (model 2) and the likelihood for model 3a because of the independence of the two sets of data of the two models. We assume the movement of the fish is fixed at $t=0.5$ no matter what types of tags they have and the fishing season of area B is from the beginning of each year to time $T$ ($T \leq 1/2$ year). We can also assume the movement of the telemetry tagged fish move at the end of every month and the traditional tagged fish move at $t=0.5$ of each year, which is the combination of model 2 and model 3b. But we do not discuss this case here.

**Model 4a:** This model combines model 1 and model 3a. In this model, certain numbers of regularly tagged fish are released to both area A and area B, at certain time point. The regularly tagged fish can migration from area A to area B and vice versa. For model 1, the regularly tagged fish can be returned by being fished from area B at following
fishing seasons. At the same time, certain number of fish with telemetry tags are released to both area A and B at the beginning of each year, and the fish can also move between area A and B. The telemetry tagged fish can be relocated at the end of each month in both areas A and B.

**Model 4b:** This model combines model 2 and model 3a. Similar to model 4a, certain numbers of regularly tagged fish are released to both area A and area B but regularly tags can be returned by either being recaptured from area A or being fished from area B in the following fishing seasons. The telemetry tags work the same way as in model 4a.

### 4.3.5 Assumptions

All the assumptions listed by Brownie et al. (1985), Nichols et al. (1982), Pollock (1991), Pollock and Raveling (1982), Hearn et al. (1998), Schwarz et al. (2003) for tag-recovery model and capture-recapture model are applicable here. Assumptions for both tag-recovery and capture-recapture method include: (1) the tagging sample is a representative of the population interested in, (2) there is no tag loss, (3) survival rates are not influenced by the tagging process, (4) the year and the location of the tag recovery (returned) is correctly tabulated, (5) the fate of each tagged fish is independent of the fate of other tagged fish, (6) all tagged fish in the same stratum have the same survival, migration and tag-recovery rates, (7) fish do not die after escaping a hook, (8) natural mortality rate is constant through the year, (9) “no animals migrate temporarily to a stratum where recoveries and releases do not occur, returning after one or more years of absence (animals may migrate out of the sampled strata permanently; such losses are indistinguishable from mortality)”, (10) the current migration route of the fish is independent of previous migration choices, (11) the fish may move directly between any populations or may stay in the same population, (12) the meta-population under study must be 'closed', that is, no immigrations or emigrations outside of the system, (13) fish must have equal 'catchability' in each population. The assumptions about the telemetry model listed in Pollock et al. (2004) are applicable here: (1) all tagged fish (whether alive or dead) have the same probability (p) of begin relocated, (2) all tagged fish alive at month have the same survival rate, mortality rate, and movement rate to the next time and the other area, (3) the probability of transmitter failure can be ignored, (4) tagged fish and untagged fish have same survival rate, mortality rate and movement rate, (5) all fish’s fate of capture and survival are independent of each other, (6)
natural mortality happens immediately prior to the first relocation, (7) no tagged fish not be first relocated for up to six periods when relocate probability is reasonable large, (8) when use both regular tags and telemetry tags, both kinds of release are randomly chosen to assure both tag types mix well throughout the population. I also made the following extra assumptions in the simulations for models 1, 2, 3a, and 4a (1) the movement time of the fish is fixed at time $t = 0.5$ of the year, (2) the fishing season of area B begins at the beginning of each year and lasts one month.

### 4.4 Evaluation of Likelihood Methods Using Simulation

For regular tagging data, we considered the most limited case of three tagging periods, two continuous fishing seasons in area B, and three recapture events in area A (for model 2) only. The fishing season is assumed to be the first month of each year. This was because the precision would be worst in this most limited situation. For telemetry data, in one scenario the time period is 13 months, which has the same length as the regular tagging method. In the regular tagging data (including model 1, model 2, model 4a and model 4b), 1000 fish were regularly tagged and released to area A and B respectively at time $t = 0, T, 1$, and $1 + T$, then for model 1, model 3, and model 4a the regular tags may be will be returned from area B at the fishing season of the first year and the second year, for model 2 and model 4b, the regular tags maybe be recaptured from area A at time point $t = T, 11 + T$, and returned from area B during the fishing seasons of year 1 and year 2. However in telemetry data (including model 3 and model 4), only 50 new fish were telemetered at the beginning of each year because of their expense. “It was also assumed that surviving telemetred fish had functioning tags for the entire length of the study” (Pollock et al. 2004). The regular tags were twice-a-year tagging, while the telemetry relocations were monthly. We used 1000 replicate simulations for each run for all models. We considered the bias of the estimators and the relative standard errors (RSE) of the estimates.

In model 1, we must assume $M^A = M^B = M$ to obtain estimability of all parameters, however in model 2, tags can be returned from both areas A and B, leading to more information about survival rates. In model 2, we can estimate and separately. Through this whole paper, we assume the natural mortality rate is the same for all years. Especially, in telemetry models, model 3a and model 3b, direct information about natural mortality
can be obtained, so $M^A$ and $M^B$ can also be estimated separately. Thus theoretically, the estimates of model 2 and model 3 should be more precise than those of model 1. And the estimates of model 4 should be better than those of model 2 and 3 because it has the benefits of both regular tag returns and telemetry tags.

The likelihoods for the models above are all products of multinomial distributions similar to Brownie et al. (1985). We ran the simulation under the 8 different cases where the values of the parameters are assigned by one of the following values $M^A = 0.2, 0.4, M^B = 0.3, 0.6, F^B_1 = 0.3, 0.5, F^B_2 = 0.2, 0.4, \psi^{AA}_1 = 0.8, 0.6, \lambda_B = 0.8, \psi^{BA}_1 = 0.2, 0.4, \psi^{AA}_2 = 0.7, \psi^{BA}_2 = 0.4, p^A_1 = 0.2, p^A_2 = 0.2$ and $p^A_2 = 0.7$.

In some cases, it is possible to get the explicit maximum likelihood estimators, whereas in other cases with parameter restrictions, iterative numerical methods should be used. Here we used L-BFGS-B method of the optimal package in R software for the computation of estimators and their variance.

We ran simulations for eight different cases. Table 4.2-4.9 are the relative RRMSE of all the estimates of model 1 to model 4b under different parameter values. Model 3a and model 4a are considered with the assumption of (column 5 and column 8 of the tables) or without this assumption (column 6 and column 9 of the tables). The relative bias and relative SEs of the simulation results are listed in appendix B. Here I assume the natural mortality M is the same for all years.

From table 4.2-4.9, under the assumption of $M^A = M^B = M$, telemetry model (model 3a) and combination model (model 4a) obtain precise estimates for all parameters. However, the estimates of model 1 have very high RRMSE. Especially for $M$, the RRMSE is as high as 177.54%. The estimates of the movement rate from combination model (model 4a) appeared to be essentially unbiased (Table 1 of appendix B). For combination model 4a, the RRMSEs of all parameters are smaller than those of model 1 and model 3a, who only use either traditional tag or telemetry tag. We can also see that the combination model does somewhat better than the telemetry model and the regular tag model for all common parameters. Specifically, the combination model does moderately better than model 3a in estimating M and F and substantially better in estimating movement rate. Without the assumption of $M^A = M^B = M$, model 4a is worse in estimating natural mortality rate than with this assumption. However it has similar precise estimators as model 4a with assumption of $M^A = M^B = M$ for all other parameters. However, the combination model
with assumption $M^A \neq M^B$ still works much better in estimating all parameters when compared to model 1 and model 3a. So, we can conclude that no matter what assumptions we have made the combination model (model 4a) always work much better than the separate models.

When compared to model 1, model 2 has much more precise estimates, especially for mortality rate and movement rate from area B to A. For simulation value 1, the RRMSE of movement rate were changed from 61.51% in model 1 to 7.53% in model 2. The RRMSE of were improved from 177.54% in model 1 to 36.04% of and 30.83% of in model 2. So even without the assumption of , the estimates of natural mortality rates in model 2 are much better than those of model 1. We can see, the estimates of model 2 have high precision, with most parameters are nearly unbiased (table 1-8 appendix B). The highest RRMSE of estimators is 42.2% and the lowest RSE is just 2.98%, with most RRMSEs are between 6% and 23%. All these parameters are also estimated reliably with the combination method (model 4b). For movement rates and recapture probabilities, the combination models have similar but a litter better estimates than model 2. For the left estimates, the combination models have apparently improved RRMSE than model 2, with the lowest improvement is 7%. Combination models are also better than single telemetry model (model 3a) with smaller RRMSE.

Of all the models listed in table 2-9, model 1 apparently has the worst estimates, with higher bias and RRMSE, it is especially bad for estimating parameter $M$ and $\psi_{1}^{AA}, \psi_{1}^{BA}$. The precision of $\psi_{1}^{AA}$ is higher than $\psi_{1}^{BA}$ because $\psi_{1}^{AA}$ has higher value than $\psi_{1}^{BA}$, resulting in more tag returns from area A. Combination model 4b is better than combination model 4a because model 2 is better than model 1. The combination models are always better than single regular tag return models and single telemetry models. The monthly relocation of telemetry method provides much more information than regular tags, especially because they obtain direct information of natural death. However, the telemetry method is very expensive and the sample is usually cannot be very big. When we combine the regular method with the telemetry method, it can take advantage of all the strengths of both methods, leading to better estimates.

In telemetry method, when the tagged fish are relocated monthly and the movement of the fish between area A and B can be detected monthly more precise estimates for migration rate should be available. In this model, we considered an example of 13 months.
But the estimates of the movement rate of the second year should be much reliable for a 14, 15, 24 months model than those of 13 months model. The reason is the more months of the second year included in the model 3b, the more migration information for second year is provided by the telemetry tags. Compared to model 3a, for value 8, the RRMSE of the estimates of the movement rate of model 3b improved from 12.81% to 4.03%, and from 22.93% to 5.92%.

Table 4.2: Simulation Result for parameters take value in column 2, comparison of the Relative Root Mean Square Error (%) of estimates for model 1 to model 4b. The first model 3a (4a) assumes \( M^A \neq M^B \), while the second model 3a (4a) assumes \( M^A = M^B \).

<table>
<thead>
<tr>
<th>Par</th>
<th>Value</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3a</th>
<th>Model 3a</th>
<th>Model 3b</th>
<th>Model 4a</th>
<th>Model 4a</th>
<th>Model 4b</th>
</tr>
</thead>
<tbody>
<tr>
<td>( M^A )</td>
<td>0.2</td>
<td>177.54</td>
<td>36.04</td>
<td>24.43</td>
<td>32.47</td>
<td>33.42</td>
<td>23.66</td>
<td>32.13</td>
<td>23.94</td>
</tr>
<tr>
<td>( M^B )</td>
<td>0.3</td>
<td>-</td>
<td>30.92</td>
<td>-</td>
<td>30.88</td>
<td>29.44</td>
<td>-</td>
<td>27.87</td>
<td>20.93</td>
</tr>
<tr>
<td>( F^B )</td>
<td>0.5</td>
<td>24.97</td>
<td>19.58</td>
<td>24.05</td>
<td>22.58</td>
<td>23.18</td>
<td>13.86</td>
<td>13.82</td>
<td>12.67</td>
</tr>
<tr>
<td>( F^B )</td>
<td>0.4</td>
<td>23.89</td>
<td>19.39</td>
<td>19.38</td>
<td>20.60</td>
<td>21.47</td>
<td>12.88</td>
<td>13.36</td>
<td>12.88</td>
</tr>
<tr>
<td>( \psi^A_{1A} )</td>
<td>0.6</td>
<td>28.75</td>
<td>5.14</td>
<td>12.61</td>
<td>12.44</td>
<td>3.81</td>
<td>5.84</td>
<td>5.99</td>
<td>4.45</td>
</tr>
<tr>
<td>( \psi^B_{1A} )</td>
<td>0.4</td>
<td>61.51</td>
<td>7.53</td>
<td>23.13</td>
<td>23.82</td>
<td>5.91</td>
<td>12.58</td>
<td>14.58</td>
<td>6.80</td>
</tr>
<tr>
<td>( \lambda_B )</td>
<td>0.8</td>
<td>17.77</td>
<td>14.39</td>
<td>-</td>
<td>-</td>
<td>10.38</td>
<td>10.56</td>
<td>9.60</td>
<td></td>
</tr>
<tr>
<td>( p^1_A )</td>
<td>0.2</td>
<td>-</td>
<td>6.65</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>6.63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( p^2_A )</td>
<td>0.2</td>
<td>-</td>
<td>8.17</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>7.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \psi^A_{1A} )</td>
<td>0.7</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>7.59</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>( \psi^B_{2A} )</td>
<td>0.4</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>13.38</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.3: Simulation Result for parameters take value in column 2, comparison of the Relative Root Mean Square Error (%) of estimates for model 1 to model 4b. The first model 3a (4a) assumes \( M^A \neq M^B \), while the second model 3a (4a) assumes \( M^A = M^B \).

<table>
<thead>
<tr>
<th>Par</th>
<th>Value</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3a</th>
<th>Model 3a</th>
<th>Model 3b</th>
<th>Model 4a</th>
<th>Model 4a</th>
<th>Model 4b</th>
</tr>
</thead>
<tbody>
<tr>
<td>( M^A )</td>
<td>0.2</td>
<td>105.01</td>
<td>33.88</td>
<td>23.11</td>
<td>32.83</td>
<td>32.87</td>
<td>23.17</td>
<td>30.65</td>
<td>23.52</td>
</tr>
<tr>
<td>( M^B )</td>
<td>0.3</td>
<td>-</td>
<td>30.24</td>
<td>-</td>
<td>32.91</td>
<td>29.60</td>
<td>-</td>
<td>25.78</td>
<td>22.19</td>
</tr>
<tr>
<td>( F^B )</td>
<td>0.5</td>
<td>21.49</td>
<td>19.20</td>
<td>23.03</td>
<td>23.37</td>
<td>22.88</td>
<td>12.90</td>
<td>13.16</td>
<td>12.72</td>
</tr>
<tr>
<td>( F^B )</td>
<td>0.4</td>
<td>21.28</td>
<td>18.51</td>
<td>20.64</td>
<td>22.17</td>
<td>21.90</td>
<td>13.12</td>
<td>13.40</td>
<td>12.59</td>
</tr>
<tr>
<td>( \psi^A_{1A} )</td>
<td>0.8</td>
<td>6.00</td>
<td>2.83</td>
<td>7.63</td>
<td>10.66</td>
<td>2.38</td>
<td>2.79</td>
<td>2.92</td>
<td>2.58</td>
</tr>
<tr>
<td>( \psi^B_{1A} )</td>
<td>0.2</td>
<td>80.78</td>
<td>10.48</td>
<td>36.83</td>
<td>39.67</td>
<td>9.51</td>
<td>27.24</td>
<td>30.31</td>
<td>10.06</td>
</tr>
<tr>
<td>( \lambda_B )</td>
<td>0.8</td>
<td>15.84</td>
<td>14.29</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>9.99</td>
<td>10.14</td>
<td>9.93</td>
</tr>
<tr>
<td>( p^1_A )</td>
<td>0.2</td>
<td>-</td>
<td>6.43</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>6.41</td>
<td></td>
</tr>
<tr>
<td>( p^2_A )</td>
<td>0.2</td>
<td>-</td>
<td>8.16</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>7.06</td>
<td></td>
</tr>
<tr>
<td>( p^3_A )</td>
<td>0.2</td>
<td>-</td>
<td>5.99</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>5.20</td>
<td></td>
</tr>
<tr>
<td>( \psi^A_{1A} )</td>
<td>0.7</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>7.27</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>( \psi^B_{2A} )</td>
<td>0.4</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>12.96</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Table 4.4: Simulation Result for parameters take value in column 2, comparison of the Relative Root Mean Square Error (%) of estimates for model 1 to model 4b. The first model 3a (4a) assumes $M^A \neq M^B$, while the second model 3a (4a) assumes $M^A = M^B$.

<table>
<thead>
<tr>
<th>Par</th>
<th>True Value</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3a</th>
<th>Model 3a</th>
<th>Model 3b</th>
<th>Model 4a</th>
<th>Model 4a</th>
<th>Model 4b</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M^A$</td>
<td>0.2</td>
<td>162.51</td>
<td>42.2</td>
<td>22.98</td>
<td>31.83</td>
<td>32.87</td>
<td>22.63</td>
<td>31.19</td>
<td></td>
</tr>
<tr>
<td>$M^B$</td>
<td>-</td>
<td>39.05</td>
<td>-</td>
<td>28.92</td>
<td>27.96</td>
<td>-</td>
<td>-</td>
<td>26.42</td>
<td></td>
</tr>
<tr>
<td>$F^b_1$</td>
<td>0.3</td>
<td>44.82</td>
<td>30.46</td>
<td>27.94</td>
<td>27.41</td>
<td>28.16</td>
<td>18.1</td>
<td>18.75</td>
<td></td>
</tr>
<tr>
<td>$F^b_2$</td>
<td>0.3</td>
<td>41.43</td>
<td>41.43</td>
<td>29.07</td>
<td>25.02</td>
<td>27.14</td>
<td>28.14</td>
<td>16.87</td>
<td>17.99</td>
</tr>
<tr>
<td>$\psi^A_1$</td>
<td>0.6</td>
<td>27</td>
<td>6.74</td>
<td>11.98</td>
<td>12.43</td>
<td>3.67</td>
<td>6.81</td>
<td>7.19</td>
<td></td>
</tr>
<tr>
<td>$\psi^A_2$</td>
<td>0.4</td>
<td>57.72</td>
<td>8.37</td>
<td>21.3</td>
<td>21.77</td>
<td>5.63</td>
<td>14.43</td>
<td>15.3</td>
<td></td>
</tr>
<tr>
<td>$\lambda_B$</td>
<td>0.8</td>
<td>25.48</td>
<td>20.82</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>14.43</td>
<td>14.96</td>
<td></td>
</tr>
<tr>
<td>$p^A_1$</td>
<td>0.2</td>
<td>-</td>
<td>6.55</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>$p^A_2$</td>
<td>0.2</td>
<td>-</td>
<td>7.86</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>$p^B_1$</td>
<td>0.2</td>
<td>-</td>
<td>5.72</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>$\psi^A_1$</td>
<td>0.7</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>7.37</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>$\psi^B_2$</td>
<td>0.4</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>13.32</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

4.5 Discussion and Future Research

In this chapter, we generalized the combination of telemetry and tag-return methods (Pollock, et al., 2004) to a two-site model with one area a marine reserve and one area a fishing area. These new models can estimate patch specific and year specific capture probability, transition rate, and tag reporting rate with good precision. When compared to the traditional tag capture-recapture model, the adding of telemetry tags contributes a lot to reduce the RRMSE of the estimates for all parameters, especially for movement rate. This is because the telemetry tags can provide us direct and detail movement information of the animals. Moreover, unlike the multi-state tag-return models developed by Joe (2001) critically depending on the tag-reporting rate, our new models can still provide reasonably good estimates even when the transition rate and tag reporting rate are low. Here we discussed four different combination methods of using telemetry tags and traditional tags in two-state models, for example, the marine reserve only has tag-recapture and the fishing area has tag-returns from a fishery such as both the marine reserve area and the fishing area have tag-recapture traditional and telemetry tags and the fishing area still has tag-returns from the fishery..

Here we only considered a full model where capture probability, survival probability, and movement rate are different for each year and each area. To improve the precision of
Table 4.5: Simulation Result for parameters take value in column 2, comparison of the Relative Root Mean Square Error (%) of estimates for model 1 to model 4b. The first model 3a (4a) assumes $M^A \neq M^B$, while the second model 3a (4a) assumes $M^A = M^B$.

<table>
<thead>
<tr>
<th>Par</th>
<th>True Value</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3a</th>
<th>Model 3a</th>
<th>Model 3b</th>
<th>Model 4a</th>
<th>Model 4a</th>
<th>Model 4b</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M^A$</td>
<td>0.2</td>
<td>98.93</td>
<td>36.99</td>
<td>22.79</td>
<td>31.88</td>
<td>32.75</td>
<td>22.62</td>
<td>31.84</td>
<td>23.38</td>
</tr>
<tr>
<td>$M^B$</td>
<td>0.3</td>
<td>-</td>
<td>38.4</td>
<td>-</td>
<td>30</td>
<td>27.63</td>
<td>-</td>
<td>25.71</td>
<td>22.32</td>
</tr>
<tr>
<td>$F^B_1$</td>
<td>0.3</td>
<td>40.03</td>
<td>31.5</td>
<td>28.87</td>
<td>27.36</td>
<td>27.94</td>
<td>17.46</td>
<td>17.65</td>
<td>16.84</td>
</tr>
<tr>
<td>$F^B_2$</td>
<td>0.2</td>
<td>37.15</td>
<td>29.76</td>
<td>26</td>
<td>27.6</td>
<td>27.53</td>
<td>17.42</td>
<td>17.48</td>
<td>16.89</td>
</tr>
<tr>
<td>$\psi^{BA}_1$</td>
<td>0.8</td>
<td>5.85</td>
<td>3.64</td>
<td>7.3</td>
<td>7.97</td>
<td>2.27</td>
<td>3.44</td>
<td>3.61</td>
<td>3.16</td>
</tr>
<tr>
<td>$\psi^{BA}_2$</td>
<td>0.2</td>
<td>78.59</td>
<td>10.72</td>
<td>35.17</td>
<td>35.11</td>
<td>9.06</td>
<td>29.03</td>
<td>30.39</td>
<td>9.73</td>
</tr>
<tr>
<td>$\lambda_B$</td>
<td>0.8</td>
<td>24.19</td>
<td>21.22</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>14.36</td>
<td>14.65</td>
<td>14.25</td>
</tr>
<tr>
<td>$p^A_1$</td>
<td>0.2</td>
<td>-</td>
<td>6.46</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>6.44</td>
<td></td>
</tr>
<tr>
<td>$p^A_2$</td>
<td>0.2</td>
<td>-</td>
<td>8.17</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>6.98</td>
<td></td>
</tr>
<tr>
<td>$p^A_b$</td>
<td>0.2</td>
<td>-</td>
<td>6.13</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>5.23</td>
<td></td>
</tr>
<tr>
<td>$\psi^{AA}_2$</td>
<td>0.7</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>7.29</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>$\psi^{BA}_2$</td>
<td>0.4</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>13.52</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

the estimates, we can add constraints to the parameters, like assuming the capture probability is constant through time. For the movement mechanism, we assume that the fish move just one time at the middle of each year. We could also generalize to the case where the fish can move between the populations at any time of the year as Joe (Ph.D. dissertation, 2001) did. We can also generalize our model to more states where some areas are reserves and other areas are fishing areas. The telemetry tag can provide us useful information about the detail movements of the fish. Based on this information, we could apply a prior distribution to the movement rate and using Bayesian approaches to estimate the parameters.
Table 4.6: Simulation Result for parameters take value in column 2, comparison of the Relative Root Mean Square Error (%) of estimates for model 1 to model 4b. The first model 3a (4a) assumes $M^A \neq M^B$, while the second model 3a (4a) assumes $M^A = M^B$.

<table>
<thead>
<tr>
<th>Par Value</th>
<th>True Model 1</th>
<th>Model 2</th>
<th>Model 3a</th>
<th>Model 3a</th>
<th>Model 3b</th>
<th>Model 4a</th>
<th>Model 4a</th>
<th>Model 4b</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M^A$ 0.4 (0.5)</td>
<td>90.14</td>
<td>19.95</td>
<td>16.12</td>
<td>23.43</td>
<td>25.11</td>
<td>14.48</td>
<td>22.78</td>
<td>14.96</td>
</tr>
<tr>
<td>$M^B$ 0.6</td>
<td>-</td>
<td>17.38</td>
<td>-</td>
<td>23.34</td>
<td>21.31</td>
<td>-</td>
<td>18.45</td>
<td>13.42</td>
</tr>
<tr>
<td>$F_1^B$ 0.5</td>
<td>30.16</td>
<td>21.48</td>
<td>23.81</td>
<td>23.87</td>
<td>23.1</td>
<td>14.76</td>
<td>14.87</td>
<td>13.34</td>
</tr>
<tr>
<td>$F_2^B$ 0.4</td>
<td>29.18</td>
<td>20.93</td>
<td>21.3</td>
<td>21.61</td>
<td>22.9</td>
<td>14.26</td>
<td>14.32</td>
<td>13.02</td>
</tr>
<tr>
<td>$\psi_{1A}$ 0.6</td>
<td>29.67</td>
<td>5.96</td>
<td>13.33</td>
<td>13</td>
<td>4.13</td>
<td>6.8</td>
<td>6.8</td>
<td>5.03</td>
</tr>
<tr>
<td>$\psi_{1A}$ 0.4</td>
<td>64.8</td>
<td>9.07</td>
<td>25.53</td>
<td>25.95</td>
<td>6.27</td>
<td>15.05</td>
<td>18</td>
<td>8.21</td>
</tr>
<tr>
<td>$\lambda_B$ 0.8</td>
<td>19.69</td>
<td>15.49</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>11.05</td>
<td>11.07</td>
<td>10.32</td>
</tr>
<tr>
<td>$p_{1A}$ 0.2</td>
<td>-</td>
<td>6.41</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>6.4</td>
</tr>
<tr>
<td>$p_{1A}$ 0.2</td>
<td>-</td>
<td>9</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>8.32</td>
</tr>
<tr>
<td>$p_{2A}$ 0.2</td>
<td>-</td>
<td>5.81</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>5.4</td>
</tr>
<tr>
<td>$\psi_{2A}$ 0.7</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>7.46</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\psi_{2A}$ 0.4</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>14.31</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 4.7: Simulation Result for parameters take value in column 2, comparison of the Relative Root Mean Square Error (%) of estimates for model 1 to model 4b. The first model 3a (4a) assumes $M^A \neq M^B$, while the second model 3a (4a) assumes $M^A = M^B$.

<table>
<thead>
<tr>
<th>Par Value</th>
<th>True Model 1</th>
<th>Model 2</th>
<th>Model 3a</th>
<th>Model 3a</th>
<th>Model 3b</th>
<th>Model 4a</th>
<th>Model 4a</th>
<th>Model 4b</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M$ 0.5</td>
<td>74.56</td>
<td>-</td>
<td>16.48</td>
<td>-</td>
<td>-</td>
<td>13.9</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$M^A$ 0.4</td>
<td>-</td>
<td>18.38</td>
<td>-</td>
<td>23.93</td>
<td>25.35</td>
<td>-</td>
<td>23.11</td>
<td>14.6</td>
</tr>
<tr>
<td>$M^B$ 0.6</td>
<td>-</td>
<td>16.78</td>
<td>-</td>
<td>25.45</td>
<td>21.8</td>
<td>-</td>
<td>16.61</td>
<td>13.08</td>
</tr>
<tr>
<td>$F_1^B$ 0.5</td>
<td>26.32</td>
<td>23.58</td>
<td>23.66</td>
<td>25.23</td>
<td>23.46</td>
<td>13.85</td>
<td>14.56</td>
<td>14</td>
</tr>
<tr>
<td>$F_2^B$ 0.4</td>
<td>25.05</td>
<td>23.18</td>
<td>21.67</td>
<td>23.2</td>
<td>23.27</td>
<td>13.84</td>
<td>13.72</td>
<td>13.65</td>
</tr>
<tr>
<td>$\psi_{1A}$ 0.8</td>
<td>8.27</td>
<td>2.98</td>
<td>7.9</td>
<td>11.32</td>
<td>2.43</td>
<td>3.31</td>
<td>3.15</td>
<td>2.75</td>
</tr>
<tr>
<td>$\psi_{1A}$ 0.2</td>
<td>120.49</td>
<td>11.88</td>
<td>42.72</td>
<td>41.22</td>
<td>10.44</td>
<td>30.08</td>
<td>34.49</td>
<td>11.27</td>
</tr>
<tr>
<td>$\lambda_B$ 0.8</td>
<td>17.7</td>
<td>16.57</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>10.54</td>
<td>10.92</td>
<td>10.74</td>
</tr>
<tr>
<td>$p_{1A}$ 0.2</td>
<td>-</td>
<td>6.48</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>6.48</td>
</tr>
<tr>
<td>$p_{1A}$ 0.2</td>
<td>-</td>
<td>8.8</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>8.12</td>
</tr>
<tr>
<td>$p_{2A}$ 0.2</td>
<td>-</td>
<td>5.9</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>5.49</td>
</tr>
<tr>
<td>$\psi_{2A}$ 0.7</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>7.68</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\psi_{2A}$ 0.4</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>13.97</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Table 4.8: Simulation Result for parameters take value in column 2, comparison of the Relative Root Mean Square Error (%) of estimates for model 1 to model 4b. The first model 3a (4a) assumes $M^A \neq M^B$, while the second model 3a (4a) assumes $M^A = M^B$.

<table>
<thead>
<tr>
<th>Par</th>
<th>True Value</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3a</th>
<th>Model 3b</th>
<th>Model 4a</th>
<th>Model 4b</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M^A$</td>
<td>0.4</td>
<td>89.37</td>
<td>23.73</td>
<td>15.65</td>
<td>22.92</td>
<td>24.59</td>
<td>14.95</td>
</tr>
<tr>
<td>$M^B$</td>
<td>0.6</td>
<td>-</td>
<td>22.15</td>
<td>-</td>
<td>21.39</td>
<td>20.51</td>
<td>-</td>
</tr>
<tr>
<td>$F^B_1$</td>
<td>0.3</td>
<td>54.37</td>
<td>35.99</td>
<td>27.47</td>
<td>27.36</td>
<td>28.04</td>
<td>19.23</td>
</tr>
<tr>
<td>$F^B_2$</td>
<td>0.2</td>
<td>49.19</td>
<td>34.36</td>
<td>26.4</td>
<td>27.9</td>
<td>31.36</td>
<td>17.58</td>
</tr>
<tr>
<td>$\psi^B_{1A}$</td>
<td>0.6</td>
<td>29.45</td>
<td>8.16</td>
<td>13.04</td>
<td>12.81</td>
<td>4.03</td>
<td>10.65</td>
</tr>
<tr>
<td>$\psi^B_{2A}$</td>
<td>0.4</td>
<td>64.41</td>
<td>9.26</td>
<td>22.78</td>
<td>22.93</td>
<td>5.92</td>
<td>16</td>
</tr>
<tr>
<td>$\lambda_B$</td>
<td>0.8</td>
<td>27.41</td>
<td>22.54</td>
<td>-</td>
<td>-</td>
<td>14.72</td>
<td>14.9</td>
</tr>
<tr>
<td>$p^A_1$</td>
<td>0.2</td>
<td>-</td>
<td>6.47</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$p^A_{2A}$</td>
<td>0.2</td>
<td>-</td>
<td>9.38</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\psi^2_{1A}$</td>
<td>0.7</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>8.07</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\psi^2_{2A}$</td>
<td>0.4</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>13.97</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 4.9: Simulation Result for parameters take value in column 2, comparison of the Relative Root Mean Square Error (%) of estimates for model 1 to model 4b. The first model 3a (4a) assumes $M^A \neq M^B$, while the second model 3a (4a) assumes $M^A = M^B$.

<table>
<thead>
<tr>
<th>Par</th>
<th>True Value</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3a</th>
<th>Model 3b</th>
<th>Model 4a</th>
<th>Model 4b</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M^A$</td>
<td>0.4</td>
<td>76.84</td>
<td>20.27</td>
<td>18.59</td>
<td>23.42</td>
<td>24.66</td>
<td>14.46</td>
</tr>
<tr>
<td>$M^B$</td>
<td>0.6</td>
<td>-</td>
<td>20.9</td>
<td>-</td>
<td>22.32</td>
<td>20.53</td>
<td>-</td>
</tr>
<tr>
<td>$F^B_1$</td>
<td>0.3</td>
<td>46.77</td>
<td>35.97</td>
<td>29.29</td>
<td>27.59</td>
<td>28.59</td>
<td>17.96</td>
</tr>
<tr>
<td>$F^B_2$</td>
<td>0.2</td>
<td>44.59</td>
<td>34.95</td>
<td>28.67</td>
<td>28.64</td>
<td>30.35</td>
<td>17.84</td>
</tr>
<tr>
<td>$\psi^B_{1A}$</td>
<td>0.8</td>
<td>8.44</td>
<td>4.05</td>
<td>12.62</td>
<td>8.92</td>
<td>2.38</td>
<td>3.82</td>
</tr>
<tr>
<td>$\psi^B_{2A}$</td>
<td>0.2</td>
<td>123.4</td>
<td>12.17</td>
<td>39.22</td>
<td>37.67</td>
<td>10.14</td>
<td>30.95</td>
</tr>
<tr>
<td>$\lambda_B$</td>
<td>0.8</td>
<td>26.09</td>
<td>22.94</td>
<td>-</td>
<td>-</td>
<td>14.77</td>
<td>14.88</td>
</tr>
<tr>
<td>$p^A_1$</td>
<td>0.2</td>
<td>-</td>
<td>6.54</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$p^A_{2A}$</td>
<td>0.2</td>
<td>-</td>
<td>8.88</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\psi^2_{1A}$</td>
<td>0.7</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>7.38</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\psi^2_{2A}$</td>
<td>0.4</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>14.17</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Bibliography


