



<http://researchspace.auckland.ac.nz>

ResearchSpace@Auckland

Copyright Statement

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

This thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognise the author's right to be identified as the author of this thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author's permission before publishing any material from their thesis.

To request permissions please use the Feedback form on our webpage.

<http://researchspace.auckland.ac.nz/feedback>

General copyright and disclaimer

In addition to the above conditions, authors give their consent for the digital copy of their work to be used subject to the conditions specified on the [Library Thesis Consent Form](#) and [Deposit Licence](#).

Note : Masters Theses

The digital copy of a masters thesis is as submitted for examination and contains no corrections. The print copy, usually available in the University Library, may contain alterations requested by the supervisor.

AN EXTENSION OF THE JOLLY-SEBER MODEL COMBINING TWO SOURCES
OF CAPTURE-RECAPTURE DATA

by

Bénédicte Madon

A dissertation submitted in partial fulfilment
of the requirements for the degree

of

Doctor of Philosophy

in

Statistics

UNIVERSITY OF AUCKLAND
New Zealand

January 2010

Abstract

I propose a modification of the Jolly-Seber model, the two-source Jolly-Seber (TSJS) model, to estimate population size by combining two sources of capture-recapture data of the same population where there might be an unknown overlap between two independent datasets. This is the case with recent surveys of whales and dolphins where researchers use individual identification records from both photo-identification and DNA profiling of skin biopsy samples. This sampling configuration results in two datasets that might contain the same individuals. This new approach enables the estimation of the overlap and the calculation the population size using capture-recapture information arising from both sampling methods. Monte Carlo simulations are used to assess the properties of the present estimator. When all the assumptions are met, the estimator seems to be unbiased as long as the occasion-specific simultaneous sampling probability is above 0.2. Simulation analyses also indicate that the proposed method performs better than existing closed-population estimators when there is little heterogeneity among individuals in capture probabilities and when the average capture probability is high. Alternatives have been explored and a two-source version of model M_0 has also been developed and compared to the TSJS estimator. Traditional closed-population estimators have been compared to the new approaches (TSJS and two-source M_0 models) when the population is open and the assumption of homogeneous capture probability is violated. Both procedures are finally applied to real data on the humpback whale *Megaptera novaeangliae*, on the wintering grounds of New Caledonia (South Pacific), where individuals have been sampled independently by skin sampling biopsy and photo-identification or simultaneously by both methods on a same capture occasion. The proposed methods hold great promise in monitoring by providing researchers and managers with a method allowing a diversity of sampling protocols. It could be more efficient in estimating population size, in terms of both precision and bias, than models based only on one type of data. And as it is important to control variation in a sampling design, this methodology could also provide a useful way to reduce variation by increasing the sample size and, hence, to enhance the estimator precision.

Acknowledgements

When I started this PhD in summer 2005, I was far from realizing that the next 4 years would so challenging. I could not have completed this work without the help of many people and I hope I will not forget anyone. If I do, please forgive me.

I would like to thank my supervisor, Brian McArdle, for always encouraging me, for supporting me to overcome the difficulties I faced, and for constantly pushing me to think critically and independently. I am also grateful to my co-supervisor, Scott Baker, for giving me the opportunity to work on such an interesting subject and to the South Pacific Whale Research Consortium for discussions and data on humpback whale populations. I acknowledge the contributions of Rochelle Constantine and Debbie Steel to photographic and genetic data management.

I would like to thank Claire Garrigue for letting me use her data and giving me the opportunity to join her team in their remote location in New Caledonia for field work. Working in the field in Prony for one month every year was probably my biggest motivation for all the PhD-required computer and office-based work. It also provided an invaluable insight into the data and into the management issues of a whale population.

I am forever indebted to my PhD roommates, especially Richard Umstätter and Christian Roever: I was very fortunate to be surrounded by such kind and clever people who generously and endlessly helped me with my computer coding difficulties. I would also like to thank all the other PhD students who contributed to the friendly and relax atmosphere in our PhD room: Heti, Derek, Kristy, Jenny, Debbie, Sarah, Sammie, Thomas, Steven. I will never forget our Yum Char lunches and game nights.

I am also very grateful to the department of statistics of the University of Auckland for funding part of my PhD and providing such a nice, friendly and stimulating work environment. I do not think I would have been able to work better anywhere else. I want to thank, particularly, Sharon Walker for her kindness and devotion in taking care of many of my administrative and financial issues. With Shoba, you were definitely for me the Heart of the department.

I want to thank my family for having supported my decision to come to New Zealand and enduring my living far from home. I do not think I could have done it without their love and encouragement. I am also forever grateful to Myriam for taking such good care of Palou: this means more than I will ever be able to express. I also want to thank all my friends for their unconditional support during the difficult moments...

I cannot adequately express my gratitude to Olivier Gimenez and Rachel Fewster for their invaluable help and support during the revision of my dissertation. I would like to thank the CEFÉ CNRS team in Montpellier for welcoming me so nicely, for their precious input and discussions in the corrections and Claire Garrigue, Isabelle Rombouts, Marc Oremus, Richard Umstätter, Christian Roever, Tomoharu Eguchi and Agnese Mancini for their useful comments on earlier draft of my chapters.

TABLE OF CONTENTS

ABSTRACT.....	i
ACKNOWLEDGEMENTS.....	ii
LIST OF TABLES.....	vi
LIST OF FIGURES.....	x
 INTRODUCTION.....	 1
 1. Chapter One.....	 11
1.1. Abstract	11
1.2. Introduction	12
1.3. The classic Jolly-Seber model.....	14
1.3.1. Model components	14
1.3.2. Abundance estimation	16
1.3.3. Assumptions	18
1.3.4. Derived models	18
1.4. The two-source Jolly-Seber model.....	22
1.4.1. Assumptions	23
1.4.2. Notation	23
1.4.3. Data structure	24
1.4.4. Estimation of the probability of single identity I_{id}	26
1.4.5. Estimating population size	33
1.5. Alternative models	36
1.5.1. The variance-weighted mean of the JS estimators	36
1.5.2. A closed form likelihood approach	37
1.6. Discussion	50
 2. Chapter Two	 55
2.1. Abstract	55
2.2. Introduction	56
2.3. Methods.....	57
2.3.1. Simulation Description.....	57
2.3.2. Relative error and measure of difference	64
2.3.3. Number of simulation runs.....	66
2.3.4. Analysis steps	69
2.4. Effect of the simultaneous sampling probability.....	69
2.4.1. With equal single sampling probabilities	70
2.4.2. With unequal single sampling probabilities	78
2.5. Effect of population size	84
2.6. Effect of number of capture occasions	93
2.7. Discussion	96
 3. Chapter Three	 100
3.1. Abstract	100
3.2. Introduction	101

3.3.	Heterogeneity in capture probabilities	104
3.3.1.	Materials and Methods	104
3.3.2.	Results	108
3.4.	Heterogeneity in sampling probabilities	115
3.4.1.	Materials and Methods	115
3.4.2.	Results	116
3.5.	Tag misreading	118
3.5.1.	Materials and Methods	118
3.5.2.	Results	119
3.6.	Discussion	122
4.	Chapter Four	126
4.1.	Abstract	126
4.2.	Introduction	127
4.3.	Materials and Methods	128
4.3.1.	Closed population models	128
4.3.2.	Heterogeneity cases	131
4.3.3.	Tag misreading	132
4.3.4.	Expected trends in population size estimates	132
4.4.	Results	134
4.4.1.	With heterogeneity in capture probability among individuals	134
4.4.2.	With heterogeneity in capture probability and unequal sampling probability	137
4.4.3.	With tag misreading	141
4.5.	Discussion	145
5.	Chapter Five	149
5.1.	Abstract	149
5.2.	Introduction	151
5.3.	Materials and Methods	152
5.3.1.	Field methods	152
5.3.2.	Matching process	153
5.3.3.	Datasets	154
5.3.4.	Capture-Recapture estimates	156
5.3.5.	Exploration of model selection and assumption violation	160
5.4.	Results	167
5.4.1.	Survey effort	167
5.4.2.	Abundance	169
5.4.3.	Model selection for closed-population models	179
5.4.4.	Capture probability	179
5.4.5.	Two-set estimation of Survival	184
5.4.6.	Temporary emigration	189
5.4.7.	Transience and trap-dependence	191
5.5.	Discussion	192
	CONCLUSION	199
	LITERATURE CITED	207
	APPENDICES	221

Appendix 1: Variances for Jolly-Seber parameters \hat{M}_i , $\hat{\phi}_i$, \hat{B}_i and \hat{p}_i .	221
Appendix 2: R codes simulating the Venn diagram in Fig.1-1 and calculating the approximation of the proportion of individuals in F (Fig.1-1).	223
Appendix 3: A linear model showing the effect of P_1 , P_2 , P_3 on the approximation of the proportion of individuals in F (Fig.1-1).	225
Appendix 4: Additional 2D views for figures 2-4, 2-5, 2-7, 2-9, 2-10, 2-11, 2-12 and 2-15.	227
Appendix 5: MRE, RMSRE, mean estimate and true value of the population size over the simulation runs, and standard errors, normal and log-normal CI coverage rates of the estimated values, at each capture occasion of a 10-year study for different capture probabilities with scenario 50-25-25 and a starting population of 500 individuals.	235
Appendix 6: MRE, RMSRE, mean estimate and true value of the population size over the simulation runs, and standard errors, normal and log-normal CI coverage rates of the estimated values, at each capture occasion of a 10-occasion study with a misread tag rate of 50%, homogeneous within the population and a starting population of 500 individuals.	239
Appendix 7: Simulation results for the two-source M_0 model (200 simulation runs).	242
Appendix 8: Sample of values of the two simulated JS estimates and corresponding weighted means under heterogeneity case 1, with scenario 20-40-40 and a starting population of 500 individuals.	244
Appendix 9: A USER'S GUIDE to the TWO-SOURCE JOLLY-SEBER MODEL and MODEL TWO-SOURCE M_0 .	246

LIST OF TABLES

Table 1-1- Latent histories i and recorded histories j for a $T=3$ -occasion experiment.....	46
Table 2-1- MRE, RMSRE, mean estimate and true value of the population size over the simulation runs, and standard errors of the estimated values and nominal and log-normal CI coverage rates at each capture occasion of a 10-year study for different capture probabilities with scenario 20-40-40 and a starting population of 500 individuals.....	74
Table 2-2- MRE, RMSRE, mean estimate and true value of the population size over the simulation runs, and standard errors of the estimated values and nominal and log-normal CI coverage rates at each capture occasion of a 10-year study for a 3:1 ratio in single sampling probabilities and with a simultaneous sampling probability of 0.5 (sampling scenario 50-37.5-12.5 or 50-12.5-37.5) and a starting population of 500 individuals.....	79
Table 2-3- MRE, RMSRE, mean estimate and true value of the population size over the simulation runs, and standard errors of the estimated values and nominal and log-normal CI coverage rates at each capture occasion of a 10-year study for sampling scenario 50-25-25 and a starting population of 50 individuals.....	86
Table 2-4- MRE, RMSRE, mean estimate and true value of the population size over the simulation runs, and standard errors of the estimated values and nominal and log-normal CI coverage rates at each capture occasion of a 5-capture-occasion study with sampling scenario 20-40-40 and a starting population of 500 individuals.....	94
Table 3-1- Capture probability for each class for every case of heterogeneity considered at an average capture probability of 0.2.....	107
Table 3-2- Capture probability for each class for every case of heterogeneity considered at an average capture probability of 0.7.....	107
Table 3-3- MRE, RMSRE, mean estimate and true value of the population size over the simulation runs, and standard errors of the estimated values and nominal and log-normal CI coverage rates at each capture occasion of a 10-occasion study for different cases of heterogeneity in capture probabilities with scenario 20-40-40 and a starting population of 500 individuals.....	108
Table 3-4- MRE, RMSRE, mean estimate and true value of the population size over the simulation runs, and standard errors of the estimated values and nominal and log-normal CI coverage rates at each capture occasion of a 10-occasion study for different cases of heterogeneity in capture probabilities with scenario 50-37.5-12.5 and a starting population of 500 individuals.....	112
Table 3-5- Sampling scenario for each class g for the four cases investigated.....	116

Table 3-6- MRE, RMSRE, mean estimate and true value of the population size over the simulation runs, and standard errors of the estimated values and nominal and log-normal CI coverage rates at each capture occasion of a 10-occasion study for 2 cases of heterogeneity in sampling probabilities between the 4 classes of individuals with capture probability of 0.2 and 0.5 and a starting population of 500 individuals.	116
Table 3-7- MRE, RMSRE, mean estimate and true value of the population size over the simulation runs, and standard errors of the estimated values and nominal and log-normal CI coverage rates at each capture occasion of a 10-occasion study with a tag misreading rate of 5% homogeneous within the population and a starting population of 500 individuals.	120
Table 4-1- Capture probability for each class for every case of heterogeneity considered at an average capture probability of 0.2.	131
Table 4-2- Capture probability for each subpopulation for every case of heterogeneity considered at an average capture probability of 0.7.	131
Table 4-3- Comparison of mean estimates under 6 cases of heterogeneity in capture probabilities of closed-population and JS models based on the dataset 1 and the weighted JS model, the two-source M_0 model, the TSJS model under scenario 20-40-40, over 10 occasions and with a starting population of 500 individuals.	135
Table 4-4- Comparison of mean estimates under 6 cases of heterogeneity in capture probabilities of closed-population and JS models based on the dataset 1 and the weighted JS model, the two-source M_0 model, the TSJS model under scenario 50-37.5-12.5, over 10 occasions and with a starting population of 500 individuals.	138
Table 4-5- Comparison of mean estimates, with a 5% tag misreading rate under constant capture probability, of closed-population and JS models based on the dataset 1 and the weighted JS model, the two-source M_0 model, the TSJS model under scenario 50-25-25, over 10 occasions and with a starting population of 500 individuals.	141
Table 5-1- Summary of photographic captures and recaptures by study site between 1999 and 2004.	165
Table 5-2- 2x2 contingency table for test 3.SR based on the individuals encountered at capture occasion i	165
Table 5-3- 2x2 contingency table for test 3.SR of expected numbers of individuals encountered at capture occasion i	166
Table 5-4- Sampling effort in the South lagoon of New Caledonia between 1999 and 2005.	168
Table 5-5- Population size, associated standard error and confidence intervals (nominal and log-normal) for closed population models based on the 1999-2005 QC photographic catalogue.	174

Table 5-6- Population size, associated standard error, confidence intervals (nominal and log-normal), growth rate and associated standard error with the Chapman’s estimator based on the 1999-2005 QC photographic catalogue.....	174
Table 5-7- Population size, associated standard error, confidence intervals (nominal and log-normal), growth rate and associated standard error with the JS model on the QC photographic catalogue 1999-2005.	174
Table 5-8- Population size, associated standard error and confidence intervals (nominal and log-normal) for closed population models based on the complete 1999-2005 photographic catalogue.....	175
Table 5-9- Population size, associated standard error, confidence intervals (nominal and log-normal), growth rate and associated standard error with the Chapman’s estimator based on the complete 1999-2005 photographic catalogue.	175
Table 5-10- JS estimates, associated standard error, confidence intervals (nominal and log-normal), growth rate and associated standard error based on the complete 1999-2005 photographic catalogue.	175
Table 5-11- Population size, associated standard error and confidence intervals (nominal and log-normal) for closed population models based on the 1999-2005 genetic dataset.	176
Table 5-12- Population size, associated standard error, confidence intervals (nominal and log-normal), growth rate and associated standard error with the Chapman’s estimator based on the 1999-2005 genetic dataset.	176
Table 5-13- JS estimates, associated standard error, confidence intervals (nominal and log-normal), growth rate and associated standard error based on the 1999-2005 genetic dataset.	176
Table 5-14- Population size, associated standard error, confidence intervals (nominal and log-normal), growth rate and associated standard error with the weighted mean of the JS model on the QC photographic catalogue and the genetic dataset 1999-2005.	177
Table 5-15- Population size, associated standard error, confidence intervals (nominal and log-normal), growth rate and associated standard error with the weighted mean of the JS model on the complete photographic catalogue and the genetic dataset 1999-2005.	177
Table 5-16- TSJS estimates, associated standard error, confidence intervals (nominal and log-normal), growth rate and associated standard error between 1999 and 2005 with the combined dataset under the photographic QC protocol.....	178
Table 5-17- TSJS estimates, associated standard error, confidence intervals (nominal and log-normal), growth rate and associated standard error between 1999 and 2005 with the combined dataset with all the photographs.	178
Table 5-18- Two-source M_0 estimates, associated standard errors, confidence intervals (nominal and log-normal) and capture probabilities between 1999 and 2005 with the combined datasets with all photographs and under the photographic QC protocol.....	178

Table 5-19- AIC for the CJS and related models for photographic catalogues 1999-2005 under the QC protocol (“Selected 1999-2005”), 1999-2005 with all photographs (“All 1999-2005”) and both without 2002/2004..... 181

Table 5-20- AIC for the CJS and related models for genetic data 1999-2005 and 1999-2005 without 2002/2004..... 183

Table 5-21- “Approximated” two-set CJS models and corresponding value of AICc and survival estimates with standard error and 95% CI lower and upper bounds with the genetic dataset and the dataset with all photographs. 186

Table 5-22- “Approximated” two-set CJS models and corresponding value of AICc and survival estimates with standard error and 95% CI lower and upper bounds with the genetic dataset and the dataset with selected photographs (QC protocol)..... 187

Table 5-23- Survival estimates and corresponding standard error (se) and 95% CI lower and upper bounds under the TSJS model for the combined datasets 1999-2005 with all photographs and selected photographs..... 189

Table 5-24- Survival estimates or range of survival estimates and corresponding standard error with the best model(s) under the CJS, TSJS and two-set CJS approaches with the different datasets 1999-2005. 189

Table 5-25- AIC for multistate models using multi-site capture history data of Humpback whales between 1999 and 2004..... 190

Table 5-26- Estimates of the probabilities of transition of humpback whales between breeding grounds in the South Pacific, between 1999 and 2004, from the CAS model $\{F, \psi_{from+to}, P_{10*time}\}$ 190

LIST OF FIGURES

Fig. 1-1- A Venn diagram showing the possible types of capture histories (with examples) for captured animals for a $T=6$ study.	26
Fig.1-2- Value of the parameter \hat{I}_{id} over 100 steps of the iteration.	30
Fig.1-3- Root of the function $f(\hat{I}_{id})$ in scenario A, B, C and D.....	32
Fig.2-1- Structure of the birth matrix produced by the program.	58
Fig.2-2- Structure of the history matrix produced by the program.....	58
Fig.2-3- Venn Diagram illustrating the structure of the sampling process in the simulation.	59
Fig.2-4- MRE as a function of the capture probability and capture occasion over 10 years, with a starting population of 500 individuals for 200 replicates.....	67
Fig.2-5- MRE for 200 replicates as a function of the capture probability and capture occasion over 10 years, with a starting population of 500 individuals for 100 and 1000 replicates.....	68
Fig.2-6- MRE as a function of the capture probability and capture occasion over 10 years, with a starting population of 500 individuals for scenarios of sampling probabilities 50-25-25 and 20-40-40.	71
Fig.2-7- MRE as a function of the capture probability and capture occasion over 10 years, with a starting population of 500 individuals for scenarios of sampling probabilities 10-45-45 and 17-41.5-41.5.	72
Fig.2-8- Box plots of MRE averaged over the study period per capture probability for an experiment of 10 years, with a starting population of 500 individuals and simultaneous sampling probability of 0.2 and 0.5.....	73
Fig.2-9- MRE as a function of the capture probability and capture occasion over 10 years, with a starting population of 500 individuals for scenarios of sampling probabilities 50-37.5-12.5 and 20-60-20.	83
Fig.2-10- MRE as a function of the capture probability and capture occasion over 10 years, with a starting population of 500 individuals for scenarios of sampling probabilities 20-48-32 a ratio of 1.5:1.	84
Fig.2-11- MRE as a function of the capture probability and capture occasion over 10 years, with a starting population of 50 individuals for scenarios of sampling probabilities 20-40-40.	85
Fig.2-12- MRE as a function of the capture probability and capture occasion over 10 years, with a starting population of 1000 individuals for scenarios of sampling probabilities 20-40-40.	89

Fig.2-13- MRE as a function of the capture probability and capture occasion over 10 years, with a starting population of 1000 individuals for scenarios of sampling probabilities 20-60-20.	90
Fig.2-14- MRE as a function of the capture probability and capture occasion over 10 years, with a starting population of 1000 individuals for scenarios of sampling probabilities 50-10-40.	91
Fig.2-15- MRE as a function of the capture probability and capture occasion over 10 years, with a starting population of 2000 individuals for scenario of sampling probabilities 20-40-40.	92
Fig.2-16- MRE as a function of the capture probability and capture occasion over 5 years, with a starting population of 500 individuals for scenarios of sampling probabilities 50-25-25.	93
Fig.5-1- Sampling area in the South lagoon of New Caledonia.	153
Fig.5-2- Example of a photograph (on the left) removed under the quality-control protocol against a good-quality one (on the right).	154
Fig.5-3- Capture probability estimates and standard errors from model $\{p_t, \phi\}$ for complete photographic data 1999-2005 and for QC-photographic data 1999-2005.	180
Fig.5-4- Capture probability estimates for males and females from model $\{p_{t+g}, \phi\}$ using genetic data in 1999-2005.	182