A hierarchical Bayes approach to capture-recapture abundance estimation
by Tomoharu Eguchi

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of
Philosophy in Biological Sciences
Montana State University
© Copyright by Tomoharu Eguchi (2003)

Abstract:
A reanalysis of stranding data for bottlenose dolphins along the Atlantic coast of the United States
during the 1987-1988 epizootic indicated that the widely used mortality rate during the die-off of 53%
was an overestimate. Using the same dataset, I computed the mortality during the epizootic to be less
than 46%. Continuous and consistent effort in collecting stranding data and estimation of life history
parameters were shown to be important in using stranding data to find die-off of a population. To make
inferences on abundance of bottlenose dolphins in estuaries along the Atlantic coast of the United
States, a hierarchical Bayes approach for capture-mark-recapture data was proposed. The performance
of the proposed method was comparable to the maximum likelihood method. When sample sizes were
small, however, the proposed method performed better than the maximum likelihood method.
Simulation analyses indicated a frequent sampling and high capture probabilities increased the
precision of posterior distributions. The proposed method was applied to the existing photographic
identification datasets from eight locations along the coast. Opportunistically collected data were useful
in obtaining the joint prior distribution of the hyperparameters of sighting probabilities. Although
posterior distributions of abundance were computed for these datasets, potential violations of the
underlying assumptions and differences in the sampling protocols among studies made impossible to
compile all results. To obtain abundance estimations from capture-mark-recapture data, the existing
sampling protocols should be modified so that the assumptions of capture-recapture analyses are
satisfied.

A HIERARCHICAL BAYES APPROACH TO CAPTURE-RECAPTURE

ABUNDANCE ESTIMATION

by

Tomoharu Eguchi

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

of

Doctor of Philosophy

in

Biological Sciences

MONTANA STATE UNIVERSITY
Bozeman, Montana

April 2003

APPROVAL

of a dissertation submitted by

Tomoharu Eguchi

This dissertation has been read by each member of the dissertation committee and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the College of Graduate Studies.

Daniel Goodman
_____   4/21/03
(Signature)   (Date)

Approved for the Department of Ecology

Jay J. Rotella
_____   4/21/03
(Signature)   (Date)

Approved for the College of Graduate Studies

Bruce McLeod
_____   4-21-03
(Signature)   (Date)

STATEMENT OF PERMISSION TO USE

In presenting this dissertation in partial fulfillment of the requirements for a doctoral degree at Montana State University, I agree that the Library shall make it available to borrowers under rules of the Library. I further agree that copying of this dissertation is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for extensive copying or reproduction of this dissertation should be referred to Bell & Howell Information and Learning, 300 North Zeeb Road, Ann Arbor, Michigan 48106, to whom I have granted "the exclusive right to reproduce and distribute my dissertation in and from microform along with the non-exclusive right to reproduce and distribute my abstract in any format in whole or in part."

Signature _____

Date _____21 April 2003_____

ACKNOWLEDGEMENTS

v

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

A reanalysis of stranding data for bottlenose dolphins along the Atlantic coast of the United States during the 1987-1988 epizootic indicated that the widely used mortality rate during the die-off of 53% was an overestimate. Using the same dataset, I computed the mortality during the epizootic to be less than 46%. Continuous and consistent effort in collecting stranding data and estimation of life history parameters were shown to be important in using stranding data to find die-off of a population. To make inferences on abundance of bottlenose dolphins in estuaries along the Atlantic coast of the United States, a hierarchical Bayes approach for capture-mark-recapture data was proposed. The performance of the proposed method was comparable to the maximum likelihood method. When sample sizes were small, however, the proposed method performed better than the maximum likelihood method. Simulation analyses indicated a frequent sampling and high capture probabilities increased the precision of posterior distributions. The proposed method was applied to the existing photographic identification datasets from eight locations along the coast. Opportunistically collected data were useful in obtaining the joint prior distribution of the hyperparameters of sighting probabilities. Although posterior distributions of abundance were computed for these datasets, potential violations of the underlying assumptions and differences in the sampling protocols among studies made impossible to compile all results. To obtain abundance estimations from capture-mark-recapture data, the existing sampling protocols should be modified so that the assumptions of capture-recapture analyses are satisfied.

INTRODUCTION

Growing human activities around the globe have increased interactions between wildlife and humans. Many researchers have provided evidence that habitat degradation can cause population decline (Caley et al. 2001, Warren et al. 2001). Direct harvesting also has impacted many species (e.g., Steller's sea cows *Hydrodamalis gigas*, many baleen whales Balaenopteridae, northern elephant seals *Mirounga angustirostris*). In other situations, however, human activities unintentionally increase the mortality of a wildlife population, sometimes endangering the population. Studies have indicated that human activities, such as military activities, automobile and ship traffic, introduction of new species, and recreational beach activities, impact survival of wildlife (Gerrodette and Gilmartin 1990, Kraus 1990, Kenney and Kraus 1993, Doak et al. 1994, Marmontel et al. 1997, Burney et al. 2001).

In marine ecosystems, several kinds of fishing operations are known to catch untargeted species, such as marine mammals, sea birds, turtles, and fishes (National Research Council 1990, Mangel 1993, Crowder et al. 1994, Trippel et al. 1996, Bravington and Bisack 1996, Weimerskirch et al. 1997, Vinther 1999, Belda and Sánchez 2001, Romanov 2002). Because of the severity of incidental mortalities in some situations, various mitigation measures have been proposed to reduce accidental catches of non-target species in fisheries (e.g., acoustic deterrent devices, turtle-exclusion devices, streamer lines, and change in the timing of fishing; Crowder et al. 1994, Trippel et al. 1996, Melvin et al. 1999, Belda, and Sánchez 2001, Løkkeborg and Robertson 2002, Ryan and Watkins 2002). The effectiveness of these measures and compliance by fishers, however, often depend on regional and international regulations, cultural understandings, and enforcement (Crowder et al. 1994, Robertson 1998, Ryan and Boix

1998, Gilman 2001). In many instances, legal protection and enforcement are necessary for protecting wildlife from various human activities.

To protect marine mammals from anthropogenic mortality and disturbance, the United States enacted the Marine Mammal Protection Act in 1972 (MMPA, most recently amended 1994). The 1994 amendment explicitly incorporates a method of computing the number of individuals that can be killed incidentally for each marine mammal stock in U.S. waters. A "stock" of a marine mammal species is defined in the act as:

> *... a group of marine mammals of the same species or smaller taxa in a common spatial arrangement, that interbreed when mature.*

The number of individuals of a stock that can be killed incidentally is called potential biological removal (PBR) and defined as the following:

> *... the maximum number of animals, not including natural mortalities, that may be removed from a marine mammal stock while allowing that stock to reach or maintain its optimum sustainable population. The potential biological removal level is the product of the following factors:*
>
> *(A) The minimum population estimate of the stock.*
>
> *(B) One-half the maximum theoretical or estimated net productivity rate of the stock at a small population size.*
>
> *(C) A recovery factor of between 0.1 and 1.0.*

The PBR of a stock, therefore, can be computed as:

$$PBR = N_{min} \frac{R_{max}}{2} F_R, \tag{1.1}$$

where N_{min} = the minimum population estimate of the stock, R_{max} = the maximum theoretical or estimated net productivity rate of the stock at a small population size, and F_R = a recovery factor between 0.1 and 1.0. The goal of the PBR is to allow each stock to reach or maintain its optimum sustainable population (OSP):

The term "optimum sustainable population" means, with respect to any population stock, the number of animals which will result in the maximum productivity of the population or the species, keeping in mind the carrying capacity of the habitat and the health of the ecosystem of which they form a constituent element.

Because the act lacked a practical method for computing these quantities, Wade (1998) provided quantitative definitions for the three components in equation (1.1) such that one can evaluate the impact of known levels of anthropogenic mortality of marine mammals. Because of the relatively small variability in the life history parameters of marine mammals, default values of $R_{\max}$ and F_R may be determined theoretically. Consequently, the minimum abundance of a stock is the only necessary information for computing the PBR when data are limited. Although simple, the method provides a means to offer robust management plans that can prevent the depletion of marine mammal populations from known anthropogenic mortality (Wade 1998).

To estimate abundance of marine mammal populations, line-transect and capture-mark-recapture methods often are used. For line-transect and other distance methods, distances from the transect line or point are used to adjust the encounter rate for estimating the density of animals (Buckland et al. 1993). Distance methods provide an estimate of abundance of a closed population in a short time period. In some situations, however, the line-transect method is impractical or inappropriate. For example, in the North Atlantic Ocean, two types of bottlenose dolphins exist: offshore and inshore ecotypes. These two ecotypes are indistinguishable from the distance. Consequently, to estimate the abundance of either population, one needs to identify the species of each individual. Capture-mark-recapture (CMR) techniques use marking or identification of individuals for estimating abundance and other population parameters, such as birth, death, immigration, and emigration rates.

The CMR method is a widely used technique for estimating abundance and other vital rates of a population. Mark-recapture abundance estimation procedures are based on the following concept: if one marks a portion of a population of animals, allows them to mix with unmarked animals, and takes a random sample from the population, then the sample should contain approximately the same proportion of marked animals as the proportion of marked animals in the entire population. In an algebraic expression:

$$\frac{M}{N} \simeq \frac{m}{n}$$
(1.2)

where N = the total population size, M = the number of individuals with marks (i.e., the number of individuals in the first sample), n = the number of individuals in the second sample, and m = the number of individuals with marks in the second sample, i.e., recaptures.

In the last 50 years, many models have been proposed for analyzing mark-recapture data (Schwarz and Seber 1999). Recent developments in mark-recapture analyses have focused on multi-parameter models, in which survival, immigration and emigration, and movement rates are incorporated. Because the multi-parameter models are complex, the estimation procedure requires sophisticated mathematics or, more often, intense numerical approximations using advanced algorithms.

Although statisticians have built many complex models to deal with a variety of situations in CMR studies, successful applications of these models require a large amount of data from well-designed long-term studies. For many biologists, collecting enough data and finding an appropriate model for estimating a parameter of interest are difficult tasks. As Dhondt (2002) lamented, "… at the moment the statisticians are ahead of the biologists. They have developed the tools, but we have not often collected sufficient good quality data over a long enough time period to take full advantage of the tools."

Biologists have been relying on readily available software, such as CAPTURE, MARK, and POPAN, for analyzing mark-recapture data, sometimes without understanding the underlying assumptions and limitations of models (Dhondt 2002).

The existing CMR models can be separated into two broad categories: closed and open population models. These categories are based on the underlying assumptions about the population of interest. Closed models are based on the assumption that the population size is not affected by immigration, emigration, birth, and death, whereas open models lack this restriction (Seber 1982, Williams et al. 2001). Consequently, the number of parameters for open models is greater than for closed models. The main parameter in closed models often is the population size. Because the precision of estimates generally decreases as the number of parameters in a model increases, given the same amount of data, closed models provide better estimates of abundance than open models if assumptions are met.

Pollock (1982) introduced the "robust design," a sampling design that combined closed and open population models. This design is based on the assumption that the population of interest is closed over short time intervals (primary period), in which multiple samples are collected, but it is open when viewed from a long-term perspective across multiple primary periods (Figure 1.1). The abundance during each primary period is estimated from a closed model, whereas survival, immigration, and emigration rates are estimated between closed primary periods (Pollock 1982). The hierarchical assumption of the robust design is attractive for long-term studies because the approach allows one to obtain precise estimates of quantities in a simple sampling design.

The robust design has been used as an *ad hoc* approach and a likelihood-based approach. In an *ad hoc* approach, a combination of existing closed and open models is used. A closed model is used for estimating the abundance of a primary period, whereas

an open model is used for survival rates between primary periods. For the likelihood-based approach, the full likelihood function is constructed as a product of closed and open components of data. Several models have been proposed for estimating parameters using the hierarchical assumption (Kendall et al. 1995, Schwarz and Stobo 1997, Lindberg et al. 2001, Kendall and Bjorkland 2001).
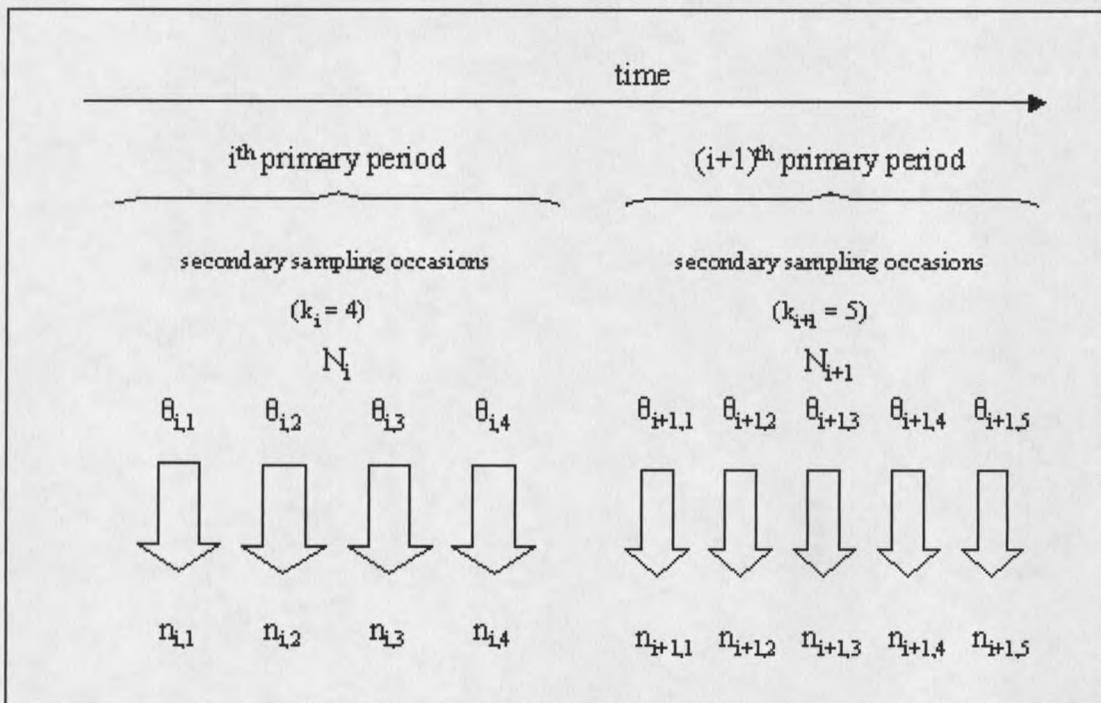
Figure 1.1. A schematic diagram of Pollock's robust design. Within each primary period, multiple samples are collected during secondary occasions with occasion-specific capture probabilities.

Kendall et al. (1995) first described a modeling framework and provided likelihood functions for a series of models for the robust design. The estimation procedure was conducted numerically. The models included capture probabilities and survival rates but the abundance was treated as a function of other parameters. Schwarz and Stobo (1997) and Kendall and Bjorkland (2001) extended the robust design so that

each primary period was considered open. In these models, the entire likelihood was considered a product of primary and secondary components. Parameters were estimated by sequentially maximizing the two parts. Lindberg et al. (2001) combined the robust design and band recovery for modeling the dispersal of Canvasback ducks (*Aythya valisineria*). Their model included temporary and permanent emigrations. Because the abundance was not a parameter of interest, it was not included in the model. In all these hierarchical models for the robust design, parameters were estimated using the maximum likelihood methods, most often via a numerical maximization of a likelihood function.

Many statisticians advocate maximum likelihood methods because of the statistically attractive properties of maximum likelihood estimators, e.g., minimum variance and being asymptotically unbiased. Because these desirable properties are based on large sample sizes, however, application of the method may be inappropriate when data are scarce. Uncertainties about a maximum likelihood estimate are expressed by a confidence interval, which is the long-run inclusion frequency of the parameter. It has been stressed that a confidence interval does not provide the probability of the estimate (Ellison 1996). A C% confidence interval includes the true value in C% of an infinite number of repeated samples from the population. Problems with using maximum likelihood estimates and confidence intervals in ecology, conservation biology, and management of small populations have been discussed (Ellison 1996, Ludwig 1996, Wade 2001, Goodman 2002). These authors suggest Bayesian statistics as an alternative approach.

Bayesian statistics is based on the same probability axioms as traditional statistics, which often is called frequentist statistics. The inference procedure in a Bayesian analysis, however, is based only on a probability model, observed data, and existing information on the parameters of interest. Assumptions about long-run sampling are not

used in the Bayesian inference procedure. The end result of an analysis is the probability distribution of parameters, conditional on the model and observed data. Although controversial, Bayesian methods have been recommended for ecological studies and conservation and management issues (Reckhow 1990, Ellison 1996, Ludwig 1996, Hilborn et al. 1994, Johnson 1999, Wade 2000). In a Bayesian analysis, previous knowledge about the parameters (prior distribution on parameters) is incorporated mathematically into the analysis via Bayes's theorem. Thorough discussions of philosophical and applied Bayesian analyses can be found elsewhere (Berger 1985, Robert 1994, Gelman et al. 1995, Lee 1997, Carlin and Louis 2000).

Although Bayesian analyses of mark-recapture models have been proposed previously (Freeman 1972, Gaskel and George 1972, Casteldine 1981), computational difficulties limited the applications. Recent developments in computer technologies, however, enabled the implementation of Bayesian methods in mark-recapture analyses (Gazey and Staley 1986, Zucchini and Channing 1986, Best and Underhill 1990, Kinas and Bethlem 1998, King and Brooks 2001). To my knowledge, however, no Bayesian method has been proposed for making inferences on abundance of a population with data from the robust design. In this dissertation, I propose a Bayesian approach for estimating the abundance of a closed population with the robust design. The proposed method is applied to previously collected data for bottlenose dolphins along the Atlantic coast of the United States.

Because of the die-off of 1987-1988, coastal stocks of bottlenose dolphins (*Tursiops truncatus*) along the Atlantic coast of the United States were listed as depleted under the MMPA (Waring et al. 2002). Although methods were not reported, Scott et al. (1988) stated that approximately 53% of bottlenose dolphins of the presumed coastal migratory stock died during the epizootic. The value has been used widely in literature

and management plans without verification of the method (Geraci et al. 1999, Waring et al. 2001). In an attempt to verify this estimate, I reanalyzed the same dataset using a different method and found that the mortality during the epizootic may have been significantly less than previously reported (Eguchi 2002).

Recent genetic and stable isotope studies have indicated that the presumed coastal stock may consist of several regional and migratory groups of bottlenose dolphins (Waring et al. 2002, A. Hohn, P. Rosel, K. Urian, pers. comm.). The most recent stock assessment report from the National Marine Fisheries Service (NMFS) lists seven management units for summer: northern migratory, northern North Carolina, southern North Carolina, South Carolina, Georgia, northern Florida, and central Florida (Waring et al. 2002; Figure 1.2). During winter, the northern migratory, northern NC, and southern NC units overlap geographically.

The management of these stocks depends largely on minimizing the number of by-catches in coastal fishing activities. Entanglements in fishing gear of coastal fishing activities may kill a significant number of bottlenose dolphins in some areas (Waring et al. 2001, 2002). For example, Palka and Rossman (2001) reported 146 bottlenose dolphins were killed by interactions with fishing operations off North Carolina and Virginia during winter 2000. Waring et al. (2002) reported that more than 50% of stranded carcasses of bottlenose dolphins in the coastal water of North Carolina exhibited signs of human interactions. Consequently, management actions are necessary to reduce human-induced mortality of bottlenose dolphins along the Atlantic coast of the United States. Proper management, however, requires an estimate of abundance for each management unit.