Incorporating model selection and decision analysis into population dynamics modeling

Eric John Ward

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

Doctor of Philosophy

University of Washington

2006

Program Authorized to Offer Degree:
School of Aquatic and Fishery Sciences

University of Washington
Graduate School

This is to certify that I have examined this copy of a doctoral dissertation by

Eric John Ward

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Chair of the Supervisory Committee:

Ray Hilborn

Reading Committee:

Tim Essington

Ray Hilborn

Andre Punt

Date: 12/01/06

Signature

Date 12/06/06

University of Washington

Abstract

Incorporating model selection and decision analysis into population dynamics modeling

Eric John Ward

Chair of the Supervisory Committee:
Professor Ray Hilborn
School of Aquatic and Fishery Sciences

Model selection tools are an integral part of biological inference. However, with the advent of recent software packages, researchers may be placing less thought into model selection uncertainty. After reviewing the existing literature on these methods, evidence for three specific biological processes are examined: intraspecific competition, population catastrophes, and Allee effects. Building on previous research, multiple models of west coast harbor seal population dynamics are compared using four model selection criteria (Akaike's Information Criterion, Schwarz Information Criterion, Deviance Information Criterion, Bayes factor). Bayesian model selection tools appear to favor intraspecific competition to a greater extent than frequentist tools. When random effects are included in the population growth rate, approximately equal weight is given to the model with competition and a model without competition. After developing state-space population models that include catastrophes, the catastrophe model is applied to four populations of harbor seals on the west coast. Because complete population time series do not exist, time series of pup production are used as an index of the total population. The catastrophe model allows the probability of a catastrophe occurring to vary by ecosystem (California, Alaska), but assumes that process error variances are shared among all populations. For the California populations, the catastrophes appear to reduce the pup production by and average of 75% percent per year. The Alaskan populations continue to decline, but do not appear to be strongly influenced by catastrophes. As a third question, the rarity of Allee dynamics is addressed using several thousand population time series in the Global Population Dynamics Database. Using a simple extension of the logistic model that allows for depensation, I illustrate that the Bayes factor estimates the frequency of Allee effects to be 5-6 times larger when compared to estimates from likelihood ratio tests. Because all models compared are nested, Bayesian model averaging is used to calculate parameter estimates across models.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Table Number Page

ACKNOWLEDGEMENTS

Funding for my dissertation research was provided by a National Marine Fisheries Service and Washington State Sea Grant Fellowship in Population Dynamics (National Oceanic and Atmospheric Administration grant number NA17RG1353, project E/I-7). In addition to my doctoral committee, I'd like to thank all of my friends (particularly those in SAFS) and family who have supported me along the way.

Introduction

The process of developing and evaluating biological models for inference is one of the few subjects that arcs across all biological research, from population genetics to ecosystem ecology. Regardless of how many parameters they contain, all models are designed to be simplifications of reality – researchers are tasked with separating models that are satisfactory representations of truth from those that are not. Because it is impossible to compare the performance of models directly to truth, the only way to assess model performance is to judge models relative to one another on some common scale (if truth were to be known, there would be no point of modeling). Choosing a criterion or scale to compare models should be done *a priori*, as each criterion has unique implicit assumptions. The majority of criteria for statistical model selection fall into one of two categories – prediction-based criteria and explanation-based criteria. While many useful prediction-based model selection criteria exist (Laud and Ibrahim 1995; Bernardo and Smith 2000), I have chosen to focus my research on explanation-based criteria. These latter criteria are all similar in that their foundations are rooted in the concept of likelihood (Edwards 1972).

For the majority of the 20th century, biologists conducted parameter estimation and model selection in a frequentist or maximum likelihood framework. The likelihood ratio test is the simplest of these approaches, and is designed to test a hypothesis against a nested null model. In addition to the constraint of nested models, there are several downsides to using likelihood ratios for model selection. Output from likelihood ratios are p-values, which have been misinterpreted by many scientists (Schervish 1996; Cherry

1998; Goodman 1999). Unfortunately, a p-value that rejects a null hypothesis cannot be used to assess evidence in favor of that hypothesis. A second problem is that because the testing framework is binary (the null is rejected or it is not), it is difficult to address model uncertainty.

The first alternative to using likelihood ratios for model selection is the Akaike Information Criterion (AIC; Akaike 1973). AIC relies on the principle of parsimony to select models that simultaneously maximize the fit to the data while using the fewest parameters possible. A second alternative to likelihood ratios is the Schwarz Criterion (BIC or SIC, Schwarz 1978). The behavior of BIC is similar to AIC, however there are several important differences. First, each criterion has different estimates of model dimensionality – BIC incorporates sample size into model complexity, while AIC does not. A second important difference is that the BIC criterion only aims to explain observed data – AIC has the same objective, but simultaneously attempts to find the best model that will fit unobserved future data (Myung 2000).

In the last two decades, Bayesian statistics have emerged as an alternative to frequentist methods (e.g. Ellison 1996). Despite the computational hurdles typically associated with these methods, Bayesian model selection criteria offer advantages over AIC and BIC. One of the most popular Bayesian model selection methods is the Deviance Information Criterion (DIC, Spiegelhalter et al. 2002). DIC has been proposed as a Bayesian alternative to AIC, and is one of the easiest Bayesian criteria to calculate. One advantage in using DIC is that it is better able to estimate model complexity for hierarchical models. Like some frequentist methods, DIC has gained popularity because it is accessible in free software programs (Ellison 2004). The main Bayesian alternative to DIC is known as the marginal likelihood or Bayes factor (Gelfand and Dey 1994; Kass and Raftery 1995). Like DIC, the Bayes factor considers parameter uncertainty when addressing model selection uncertainty. The critical advantage that the Bayes factor has over DIC and other model selection tools is that it is most compatible with decision making. Other model selection tools may be used to compute model weights which may be normalized to 1.0 (Burnham and Anderson 2002), however these weights cannot be interpreted as model probabilities. If each model is assigned equal prior probabilities, the marginal likelihoods from the Bayes factor calculation may be normalized to generate valid probabilities for each candidate model.

The primary objective of my research is to compare and contrast the behavior of frequentist and Bayesian model selection criteria, using both simulated data sets and existing time series. Secondary objectives include to (1) to introduce new software tools for biologists to fit single species population models and implement Bayesian model selection methods; (2) to review the existing literature of model selection methods commonly used in biology; (3) to apply these model selection tools to populations of west coast harbor seals to address evidence for intraspecific competition; (4) to introduce new population models that incorporate catastrophes; (5) to assess the evidence for catastrophes in four populations of U.S. fur seals using Bayesian model selection tools; and (6) to use Bayes factors evaluate the evidence for Allee dynamics in several thousand time series across three taxonomic classes. As each of these chapters has been written as a standalone publication, some overlap may occur between them.

Chapter 1: BEAST Software

The first chapter introduces a new software package (BEAST, "Bayesian Ecological Analysis of Statistical Trends") for Bayesian estimation of population trends. The software is free, and requires the user to do no programming. Instead, the focus is on specifying prior distributions on parameters, and the form of the error distribution. A wide range of single-species population dynamics models are supported, including single-stage population models, stage-structured population models, and age-structured population models. Uncertainty may be incorporated into the models either as process error or observation error. Further, the error may be additive or multiplicative, and may be chosen to by symmetric or asymmetric. Models may be made more complex by including time lags and change points. In addition to Bayesian parameter estimation, this software allows users to do population projections. Results from these projections may be used for inference directly, or the projections may be used by the BEAST software in two ways: to create probability tables and to conduct decision analyses to determine the exact nature of the population trend. In addition to providing managers with information in population assessments, the BEAST program may be useful as a teaching tool by agencies and universities.

Although BEAST was initially designed to do Bayesian parameter estimation, the ultimate goal was to make Bayesian model selection methods more accessible. The first model selection criterion outputted by BEAST is the Deviance Information Criterion (Spiegelhalter et al. 2002). A second criterion, involving slightly more computation, is the Bayes factor. The BEAST program does not calculate the Bayes factor directly, because a smoothing function must be specified for each individual problem (Gelfand and Dey 1994). Source code is provided to calculate the Bayes factor in R using a multivariate normal smoothing density (e.g. Cui et al. 2002).

Chapter 2: Model Selection Review and Comparison

In the second chapter, I compare the ability of four commonly used statistical criteria (Akaike's Information Criterion, Schwarz Information Criterion, Deviance Information Criterion, Bayes factors). These model selection tools are applied to thousands of simulated abundance time series to assess the ability of each to (1) evaluate the evidence of density dependence, (2) evaluate the evidence of Allee effects, (3) determine whether multiple life history stages can be detected with total population counts, and (4) determine which criteria are able to distinguish between density dependence occurring in survival parameters and density dependence occurring in fecundity parameters. All model selection criteria failed equally in this latter aim; with a single time series of data, separating the confounded parameters is virtually impossible. Across simulations with single-stage models, the frequentist criteria consistently favored simpler population models when compared to Bayesian criteria. Among the Bayesian criteria, the Bayes factor appeared to favor the simulation model more frequently than the Deviance Information Criterion (DIC). In all comparisons, DIC tended to favor simpler models, while more uncertainty was associated with the Bayes factor.

Chapter 3: Intraspecific Competition in West Coast Harbor Seals

As new model selection criteria are applied to biological problems, it is important to understand the implications of each tool on model inference. Traditionally, population assessments of the Pacific harbor seal (*Phoca vitulina richardsi*) on the west coast have been done using Akaike's Information Criterion (AIC). In addition to the fact that AIC cannot be used as a decision tool, AIC cannot be used to evaluate hierarchical models with intermediate level parameters. Extensions of AIC that consider random effects have been proposed (Vaida and Blanchard 2005), but their performance is still being investigated. Alternatives to AIC include Bayesian approaches, such as the Bayes factor. In this analysis, I extend single-species population models to include the effects of intraspecific competition for 11 west coast harbor seal populations. Both fixed effects and random effects models were considered. However, the models with random effects appear to be strongly favored over the fixed effects models. Regardless of whether random effects are included or not, some populations appeared to have autocorrelated residuals that were unexplained by the model. Comparing the models with and without competition illustrates that the posterior probability is highest for the random effects model that included competition between adjacent geographic regions. One problem with this model, however, is that the strength of density dependence is strongly dependent upon the choice of prior, and is not estimable in the presence of competition.

Chapter 4: Incorporating Catastrophes Into Population Models

Catastrophic events are considered a major contributor to extinction threats, yet are rarely included in population models. In this chapter, I extend the basic state-space population dynamics model to include a mixture distribution for the process error. The mixture distribution consists of a "normal" component representing regular process error variability, and a "catastrophic" component, representing rare events that negatively affect the population. Direct estimation of parameters is rarely possible using a single time series. However, estimation is possible when time series are combined in hierarchical models. I apply the catastrophic state-space model to simulated time series of abundance from simple non-linear population dynamics models. Applications of the model to these simulated time series indicate that population parameters (such as the carrying capacity or growth rate), and observation and process errors are estimated robustly. Both the frequency and magnitude of catastrophes are susceptible to bias. These simulations indicate that the power to detect a catastrophe is also a function of the strength of catastrophes, and the magnitude of observation and process errors.

Chapter 5: Catastrophes in Northern Fur Seal Populations

In this chapter, I illustrate an application of a catastrophic state-space population dynamics model, in which normal year-to-year variability and infrequent catastrophic mortality events are combined in a model of pup production for four northern fur seal populations (St. Paul, St. George, Bogoslof, San Miguel). Pup production is commonly used as a proxy for abundance, and may reflect catastrophic events in adult survival or fecundity. The catastrophic model is flexible in that it allows the frequency and magnitude of catastrophes to vary by ecosystem. One limitation of this model is that for all parameters to be estimated, the process error variances must be assumed to be shared among all populations. Each of the 48 models considered allows each distinct population to have unique population dynamics equations. Different forms of observation uncertainty are considered, including survey-specific CVs, and an additive scaling parameter, to evaluate whether the standard errors of the survey counts are underestimated. Using the deviance information criterion (DIC) the data strongly support a model that includes catastrophic events. The estimated annual probability of a catastrophe for these populations was found to be approximately 1.8% for the Bering Sea ecosystem, and slightly higher for the California Current ecosystem (7.1%). The estimated magnitude of catastrophes on San Miguel Island was a 75% reduction in the number of pups. The best fitting model also included survey-specific observation error CVs, which is expected because of how the data are collected. The models selected in a state-space framework are consistent with the pup production models used in previous fur seal assessments.

Chapter 6: Evidence For Depensation Across Taxa

Depensatory population dynamics (Allee effects) may be defined as a decrease in population growth rate at low density. Understanding why Allee effects occur and developing tools for detecting them is critical for population dynamics modeling. Recent large-scale meta-analyses suggest that depensatory dynamics are extremely rare across multiple taxonomic groups. Several previous approaches have focused on traditional hypothesis testing. In this chapter I present an alternative approach, analyzing the evidence of Allee dynamics in a Bayesian framework, incorporating both observation and process errors. Three single species population dynamics models are considered: a theta-logistic population model, an extension of the logistic model that includes Allee effects, and a geometric growth model. To evaluate the evidence for depensation, I applied Bayesian tools (Bayes factors, model averaging) to 1900 records from the Global Population Dynamics Database. The results applied to the Global Population Dynamics

Database time series suggest that evidence for depensation is 5-6 times more common than previous analyses have suggested, but still rare (~1.2%). Across taxa, the geometric growth model without density dependence received much more weight than models with density dependence.

Chapter 1: New Bayesian Software Tools For Single Species Models

Introduction

Statistical trend analysis in ecology is useful both as a tool for understanding population dynamics (estimating a growth rate, for instance) and as a predictive tool for predicting future population states. While many methods have been proposed for conducting trend analyses, the majority of these methods require some knowledge of statistics, as well as some computer programming skills. The programming necessary to estimate parameters may vary greatly, depending on whether frequentist or Bayesian statistics is applied. In a frequentist (or maximum likelihood) setting, biologists may implement existing algorithms such as the Nelder-Mead simplex routine (Press et al. 1992) to obtain maximum likelihood estimates (MLEs). Similar search routines are built in to a variety of free (e.g. R) and commercial software programs (S-plus, Systat, Excel, Matlab). The main alternative to estimating trend parameters in a frequentist setting is Bayesian statistics. Implementing Bayesian methods poses two main problems: 1) often the exact routine is case-specific (meaning it is more difficult to implement general versions in software), and 2) due to the high-dimensional integration required in Bayesian estimation, more advanced programming is required.

A range of free and commercial software programs have been available for fish and wildlife biologists to analyze ecological data. Packages such as MARK (White and Burnham 1999) or Distance (Thomas et al. 2005) are useful for specific sampling designs (capture-recapture and point or line transects respectively). Other packages, such as Coleraine (Hilborn et al. 2003) or EcoSim (Christensen and Pauly 1992) are specific to unique population models (fisheries catch-at-age and multi-species bioenergetics models respectively). The number of Bayesian packages available for estimating population trends is limited. The BATS package (Pole et al. 1994) implements dynamic linear models (or state space models), and allows a range of trends, but has some limitations (for a full review see Pedregal 1996). Additional packages that may be used to estimate trends, such as WinBugs (Speigelhalter et al. 2003) or PyMC (Fonnesbeck 2006) are much more flexible, but require users to write code. Although both packages may be used to generate posterior distributions of population trends, the downside to both BATS and WinBugs is that neither allow users to conduct formal Bayesian decision analyses.

Decision analyses (sometimes referred to as risk analyses) in ecology may be conducted using Bayesian or frequentist statistics. The rationale for conducting a decision analysis in a Bayesian context is that the wildlife managers are able to fully specify all costs (or penalties) of making wrong decisions (Wade 2000). This is not true when statistical tests of trends are done in a frequentist setting – tests based on likelihood ratios assign fixed costs to Type I and Type II errors, and do not give users full control over the costs of each error. Some researchers have proposed ad hoc methods for changing the definition of a significant biological trend (Hayes and Steidl 1997, Dixon and Pechmann 2005), but these methods still do not give researchers full control over all error costs.

There are three main objectives for the BEAST software project: (1) develop a flexible population modeling software package that will allow biologists not familiar with programming to conduct parameter estimation via Markov chain Monte Carlo (MCMC), (2) allow biologists to conduct population projections based on those parameter values, and (3) incorporate user-specified cost functions to allow users to conduct Bayesian decision analyses on the population trends.

Methods

When designing the BEAST software package (http://students.washington.edu/warde/Beast.html), I chose to not restrict the flexibility or complexity of the models supported. The essential ingredient in Bayesian analyses is incorporating uncertainty or error – in this software, error may be built-in under the class of observation error models or process error models. For each class of models, the likelihoods available are dependent on the chosen form of the error (additive or multiplicative). While error distributions in ecology tend to be symmetric (normal, Student's t, double exponential), several asymmetric distributions (gamma, exponential, half-normal) have been included. The BEAST software package is written to run on the Windows platform, and takes advantage of several programming languages (Java for the user interface, C for the computation) to offer an extremely fast tool for Bayesian parameter estimation.

The models supported by BEAST fall into three categories: single stage-structured models, stage-structured models, and age-structured models. Currently, there are 7 single stage-structured models available: Random Drift, Exponential, Theta-Logistic, Theta-Ricker, Depensation, Allee, and Gompertz models (see Ricker 1954; May 1974; Lewis and Kareiva 1993; Case 2000; Turchin 2003). Stage-structured models supported by BEAST may have 2–100 stages. To make these models as general as possible, users are able to incorporate density dependence into survival or fecundity (or both) in any of the stages. The form of density dependence may be chosen from one of several functions (both linear and non-linear). Age-structured models supported by BEAST are also limited to 100 age classes. The assumptions for all age-structured models are: juvenile survival is constant to the age at which individuals become adults, adult survival is constant for all age classes to senescence, and fecundity is equal for all age classes greater than the age of reproductive maturity.

In addition to specifying various forms of the population model, the BEAST package allows users to change other features of the model. One of the most important features of single-species modeling of density dependence is the choice of time lag (Lande 2003; Turchin 2003). In BEAST, the default time lag is 1, but may be changed by the user to another positive integer. The time lag is not allowed to be a random variable because is the likelihoods associated with two competing lag values were to be compared, they would use different data. As a simple example, assume we have a data set with N abundance estimates and we wish to calculate the likelihood of two models with only observation error present. A model with a time lag of 1 will have N components of the joint likelihood, whereas a model with a time lag of 2 will have N-1 components. A second option that users have is to specify change points in models (also referred to as 'broken stick' or 'open gate' models). Change points represent drastic changes in population parameters, and may be assumed to be a result of internal (e.g.

disease) or external (e.g. environmental change) population processes. Unlike time lags, change points are allowed to be random variables and may be estimated as parameters.

Because the majority of trend analyses are conducted on population sizes, we chose to limit the allowable data to only support abundance estimates and count data. There may be an unlimited number of data points and gaps in data are unlimited in BEAST. Each data point is allowed to have a standard error associated with it, however the standard error may also be estimated (if the standard error is not constant between years, it may be entered up to a proportionality constant, which becomes the estimated parameter). The final piece of data that may be entered is additional mortality or harvest data for each year.

Each parameter of the population model may be assumed known (and assigned a fixed value), may inherit the properties of other parameters, or may be assigned one of 25+ prior distributions. Parameter estimation in BEAST is done in a fully Bayesian framework via Metropolis-Hastings MCMC sampling written in the C programming language (Gelman et al. 1995). Further details of MCMC sampling, including a discussion of burn-in, thinning, and chain length can be found in Gilks et al. (1996). Output from BEAST may be loaded into packages such as CODA (Best et al. 1996) to assess chain convergence and diagnostics. This output may also be loaded into packages such as Excel or R to produce plots of posterior distributions.

While pre-existing software packages may be able to conduct parameter estimation, what sets BEAST apart is the ability to conduct Bayesian trend analyses. In the parameter estimation phase of BEAST, the user has the option of doing stochastic population projections. If the model is an observation error model, stochasticity in the future population state is incorporated via uncertainty in parameters (using each MCMC draw to project the population forward). If the model is a process error model, stochasticity is integrated by sampling from the estimated process error distribution at each time step. One advanced option is that future population sizes may be subjected to user-specified fixed harvest (e.g. 100 animals per year). Results from these projections are used by the BEAST program in two ways: 1) creating population projection tables, and 2) conducting formal decision analyses.

Population projection tables (or decision tables) are one way of conveying information about population size under alternative management actions (Punt and Hilborn 1997). Because the BEAST package does not support the incorporation of different management scenarios, we chose to construct decision tables based on current and future population size. An example table (Table 1.1) illustrates one of the many ways that probability values in the table may be interpreted.

The second use of population projections is to conduct a formal Bayesian decision analysis. Very few examples of ecological decision analysis exist in the literature (Wade 2000; Field et al. 2004). Conducting a decision analysis in BEAST requires identifying penalties (or costs) associated with making various wrong decisions. These costs are case-specific, and may be subject to temporal and spatial effects. Following from Wade (2000), one simple application might include three scenarios: a population is decreasing, the population is stable, or the population is increasing. Each of these costs must be specified, and are case-specific. For example, it may be more appropriate to penalize every wrong decision equally when a population is near carrying capacity, but if a population is at risk of becoming extinct, a cost matrix based on the precautionary approach may be more suitable.

Sibly et al. (2005) estimated the theta parameter in the theta-logistic population model for 1849 populations from the Global Population Dynamics Database. I illustrate one application of BEAST to real data by focusing on a time series of stock pigeon (*Columba oenas*) from Sibly's analysis. The *C. oenas* data are a continuous 12-year time series of breeding pairs, taken from Williamson (1983). Lognormal priors were placed on the initial population size and carrying capacity, uniform priors were placed on the growth rate and strength of density dependence, and a log-uniform prior was placed on the observation error standard deviation. After a burn-in of 10 million, every 100th MCMC draw was stored until 100,000 samples were saved (total run time < 2 minutes). Upon completion, the BEAST output was loaded into CODA, and the chain was found to pass both the Geweke and Heidelberger-Welch diagnostic tests (Geweke 1992; Heidelberger and Welch 1983). The mode of the posterior distributions were calculated after binning each parameter into 200 bins. Sibly et al. (2005) estimated the MLE estimate of θ as 0.2, however the estimate of the posterior mode is slightly larger (0.38, Figure 1.1). This difference indicates that if the Bayes posterior mode was used as the best point estimate of θ, the strength of density dependence would be greater when compared to the MLE. It is unclear whether the difference between the MLE and

Bayesian estimates is a result of using an informative prior for θ, or that estimation is done by integrating (rather than maximizing) over the parameter space.

To illustrate BEAST's ability to do population projections, I projected the number of *C. oenas* breeding pairs 5 years into the future. Two somewhat intuitive results are immediately obvious from these projections (Figure 1.2). First, the uncertainty associated with the number of future breeding pairs increases with time. Second, there is a slight trend (4%) in the mode and median of the projected population sizes, indicating that the population may increase slightly. The use of Bayesian statistics allows other questions may be addressed from these posterior distribution plots. One useful tool might be to generate Bayes predictive intervals, giving the probability that a parameter is between two values (unlike frequentist statistics where the probability a parameter lies in confidence bounds is always 0 or 1). Additional questions that may be important are "What is the probability that in 5 years, the number of pairs will be greater than 50?" The solution may be found by simply cumulating the portion of the histogram for the five-year projection that is greater than 50.

Results and Discussion

Conducting parameter estimation in a Bayesian framework allows several advantages over traditional frequentist (or likelihood) parameter estimation methods. First, a wide range of uncertainty is allowed to be included via the prior distribution. By creating a simple user interface in BEAST, ecologists will be able to conduct simple sensitivity analyses to the choice of prior distributions. A second advantage of Bayesian methods is that inference from parameter estimates is expressed as probabilities –

researchers may address questions such as "What is the probability that the population size is less than 20?" to provide advice to managers. Output from BEAST may be used to address these types of issues, and may be imported into virtually any database or plotting software.

While the BEAST program is relatively flexible, it does have several limitations that need to be addressed in future releases. First, density dependence can only be modeled in response to the total population size, N_t. This assumption is probably valid for very few biological systems (especially those in which one particular stage is known to interact strongly with another, e.g. Benoit et al. 1998). Second, the current version of the software is limited in the sense that it only allows trend analyses to be done on count-type data. Future versions will allow more flexibility, allowing trend analyses on other important factors, such as survival rates and genetic variability. Currently, a single time series is supported, but future releases will allow multiple data sets will also be supported (for example, some surveys might estimate the number of adults and juveniles in a population). A third limitation is that we currently have implemented single-species population dynamics models. Complete ecosystem models (e.g. Ecopath) are beyond the scope of the BEAST project, however it would be possible to incorporate two species predator-prey models into future versions of BEAST. A larger goal will be to include hierarchical modeling, which has an infinite number of applications in ecology. Ideally, fish and wildlife biologists will be able to use this tool to provide advice to natural resource managers, facilitating the decision making process. Because BEAST is provided as free software, we hope to encourage future development and suggestions.

Table 1.1 Example of a decision table, showing the probability of the future state of the population (Z) and the current state of the population (X). An example of how values in the table can be read is: given the population size is less than 100, the probability of the population remaining less than 100 is 0.06 (or 0.01/[0.01 + 0.10 + 0.05]). The denominator in this calculation represents the sum over all future population states.

	P(Z <= 100)	P(100 < Z <= 200)	P(200 < Z <= 500)	P(500 < Z)
P(X <= 100)	0.01	0.1	0.05	0
P(100 < X <= 200)	0	0.1	0.15	0.01
P(200 < X <= 500)	0	0	0.02	0.08
P(500 < X)	0	0	0	0.13

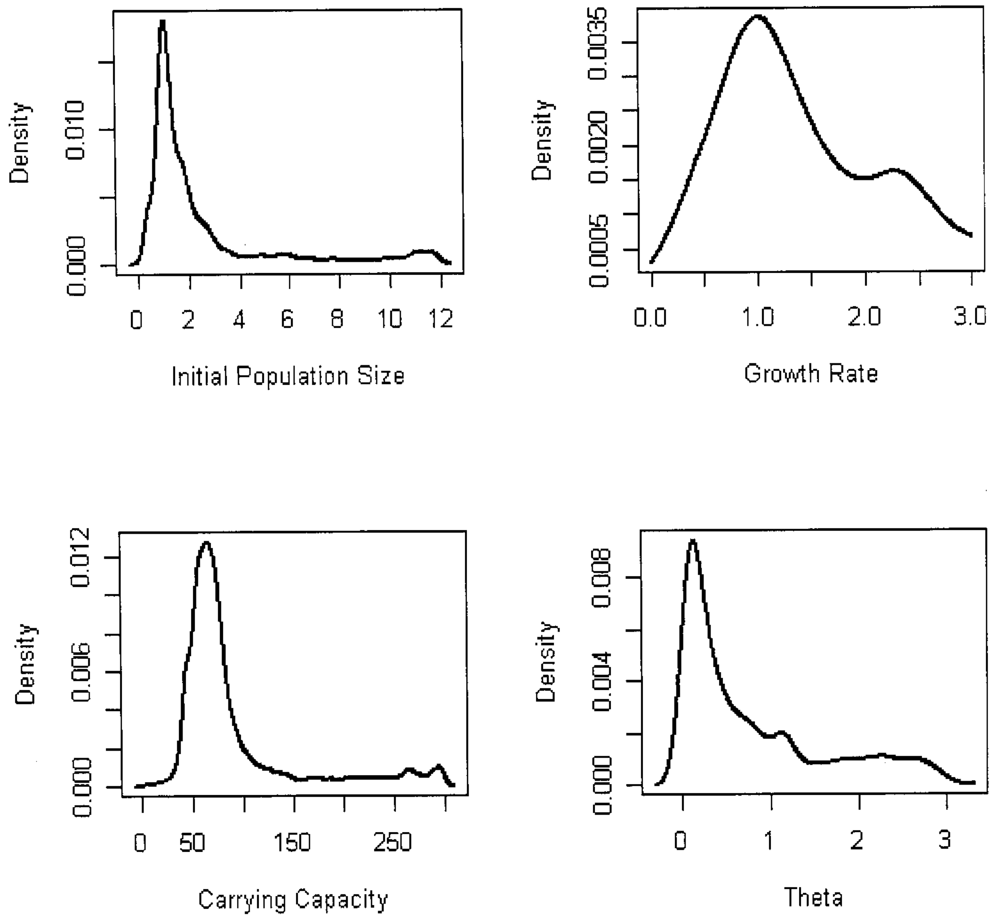

Figure 1.1 Posterior distributions for the initial population size, growth rate, carrying capacity (K), and theta-logistic parameter (q). Results from 100,000 samples, thinning every 100th (total chain length = 10,000,000).

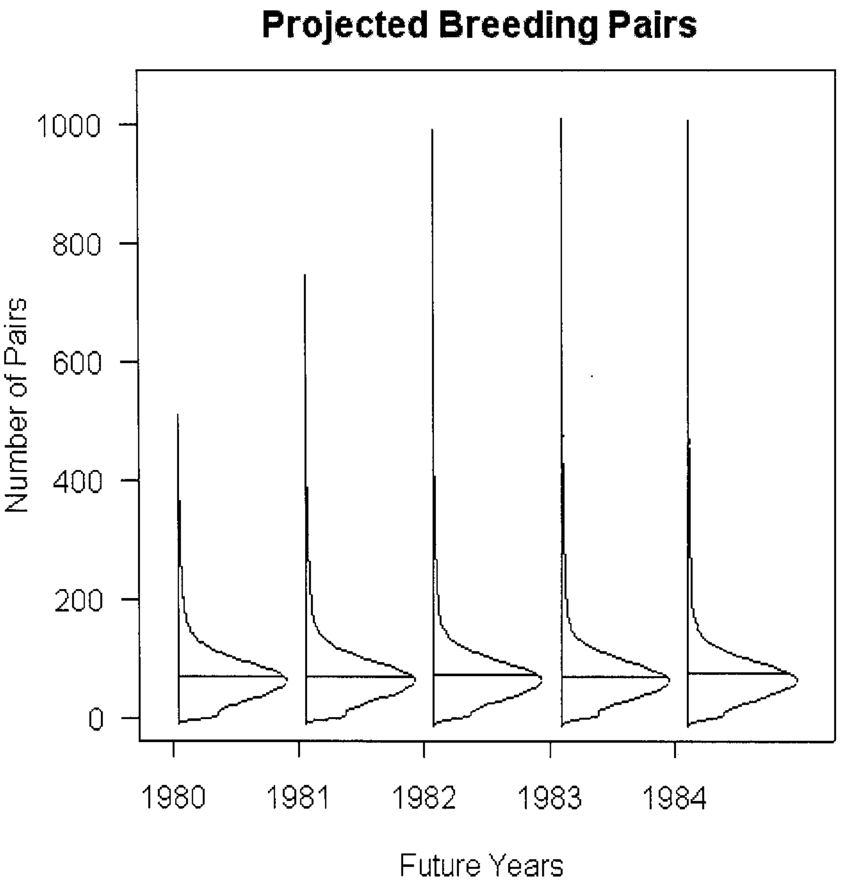

Figure 1.2 Projected breeding pairs for the stock pigeon data set. Each projection is done with 100,000 samples from the MCMC chain.

Chapter 2: A Review and Comparison of Four Popular Model Selection Criteria

Introduction

The goal of developing biological models is to capture the behavior of natural processes and patterns using relatively simple mathematical models. Models may serve as both explanatory and predictive tools; if a model of the abundance of a species appears to realistically track variation in time, it may be used to predict future population sizes or to assess extinction risk. Population dynamics models are often complex, involving age-structured or sex-structured dynamics, the effects of both time and space, and potentially time lag effects between years. In developing statistical models for inference, researchers must decide *a priori* which (if any) of these factors may be important in the life-history of the organism being modeled. A second decision researchers must make is how model performance should be assessed. For the purpose of this analysis, model performance is defined as the ability of a model to meet some specified objective (e.g. how well a model explains observed data, makes predictions, or can be used in the context of decision analysis). Statistical tools allow models to be compared relative to one another, but each criterion differs in what is considered a 'good' model. Differences between model selection criteria are rooted both in the philosophy of science, as well as statistics. Criteria may view good models to be those that make as few assumptions as possible (Occam's razor; Forster 2000), those that minimize the tradeoff between bias and variance (using the principle of parsimony; Burnham and Anderson 2002) or those that minimize the Type I and Type II error rates (Weiss 1997). Because these tools are so widely used in the biological sciences (Johnson and Omland 2004), it is important to understand the differences in the assumptions and performance of each criterion.

Over the last 20 years, the most widely used model selection tool used in the biological sciences has been Akaike's Information Criterion (AIC; Akaike 1973). AIC began to see widespread use by biologists since the 1990s after several papers illustrated its application to capture-recapture analyses (Anderson et al. 1994; Burnham et al. 1995). AIC is calculated as $AIC = D(\underline{\hat{\theta}}_{MLE}) + 2K$, where the function $D(\cdot)$ represents twice the negative log-likelihood, $\underline{\hat{\theta}}_{MLE}$ is a vector of maximum likelihood parameter estimates (MLEs), and K represents the number of model parameters. The second term in the AIC calculation represents a complexity term, and has been suggested by some researchers to be a measure of model complexity (Burnham and Anderson 2002). A popular variant of AIC is the small sample AICc, where the complexity term $2K$ is replaced by $\frac{2 \cdot K \cdot n}{n-K-1}$ (Hurvich and Tsai 1995), where n represents the samples size. When the sample size is small, AICc favors models with fewer parameters compared to AIC, but as the sample size becomes relatively large, the behavior of AICc and AIC becomes similar (Figure 2.1). AIC has inspired several additional variants to account for overdispersion (QAIC), correlations between model parameters (CAICF), and situations where models might not be good approximations to truth (TIC; Burnham and Anderson 2002). Because these criteria are not as commonly used as AIC or AICc, they are beyond the scope of this review.

A second important model selection criterion that is frequently compared to AIC, is the Bayesian Information Criterion (BIC or Schwarz Information Criterion; Schwarz

1978). It is important to note that while the two criteria may appear similar, BIC has a completely independent derivation (Burnham and Anderson 2004). Like AIC, the calculation of BIC involves 2 terms: $BIC = D(\underline{\hat{\theta}}_{MLE}) + K \cdot \ln(n)$. The first term is identical to that used in the calculation of AIC (representing the fit of the model to data), however the complexity term is slightly different, being a function of both the number of parameters (K) and sample size (n). When n is < 7.4, BIC assigns more weight to complex models than does AIC, but as the sample size increases, BIC assigns more weight to simpler models when compared to AIC or to p-values (Figure 2.2; Raftery 1995; Forster 2000; Burnham and Anderson 2002). Generally, BIC is interpreted as a rough approximation to the logarithm of the Bayes factor (Kass and Raftery 1995). BIC exists in somewhat of a grey area – while the computation of the criterion is not itself Bayesian, there are some situations where the BIC model weights may be interpreted as posterior model probabilities (Raftery 1999). Although BIC does not require the explicit specification of a prior distribution on model parameters, the prior implicitly assumed by BIC is a multivariate normal distribution centered on the MLE (with a covariance matrix equal to the inverse of the Fisher information matrix). Another name for this implicit prior is the unit information prior, because it contributes approximately the same information as one additional data point (Raftery 1995). In cases where the implicit BIC prior is similar to the prior used in the Bayes factor calculation, BIC can be used to calculate posterior probabilities for each model being considered (posterior probabilities could be computed using normalized BIC weights; Burnham and Anderson 2002). BIC may be poorly suited for some biological problems, particularly when the multivariate normal distribution is not a reasonable prior and few data points exist (skewed parameters may include variance parameters, non-linear density dependence, habitat quality, sex ratios, etc.).

One downside to the conventional forms of AIC and BIC is that they are not well suited for complicated ecological models that include hidden states with non-Gaussian errors, or hierarchical parameters (Burnham and Anderson 2002; Vaida and Blanchard 2005). Spiegelhalter et al. (2002) proposed using the Deviance Information Criterion (DIC) as a Bayesian equivalent of AIC for complicated hierarchical models. Like BIC and AIC, the calculation of DIC involves a measure of model fit, and a measure of model complexity: DIC $= D(\hat{\theta}) + 2p_D$, where $D(\hat{\theta})$ represents the deviance evaluated at some posterior parameter estimate, and p_D represents the effective number of model parameters. The DIC equation closely resembles the calculation of AIC, and in the absence of informative priors on model parameters, the two criteria are expected to be equal (Ellison 2004). The term p_D can also be expressed as the deviance function (twice the negative log-likelihood) evaluated at the posterior means subtracted from the mean deviance across all possible parameter values, $p_D = \overline{D(\theta)} - D(\hat{\theta})$. The mean deviance and mean parameter values were originally proposed by Spiegelhalter et al. (2002) as plug-in estimates (and are currently implemented in as the default option in WinBugs; Spiegelhalter et al. 2003), however the posterior median or mode may also be appropriate (e.g. Celeux et al. 2006). Proponents of DIC argue that it is a Bayesian equivalent of AIC (Spiegelhalter et al. 2002), however the similarity between the two methods is still being investigated.

A final model selection criterion that has seen increased use by biologists in the last decade is the Bayes factor (Jeffreys 1935; Kass 1993; Kass and Raftery 1995). The Bayes factor for two models (M_1, M_2) can be interpreted as the ratio of posterior odds to prior odds, $BF = \frac{P(x \mid M_1)}{P(x \mid M_2)} = \frac{P(M_1 \mid x)/P(M_2 \mid x)}{P(M_1)/P(M_2)}$ (Good 1958). As the Bayes factor updates the prior distribution upon observing the data, it is essentially a measure of how much a researcher learns by observing data (Bernardo and Smith 2000). If more than two models are being compared, and all models are given equal prior weight, the Bayes factor in favor of model *j* becomes $BF = \frac{P(\underline{x} \mid M_j)}{\sum_i P(\underline{x} \mid M_i)}$. There are several approaches for approximating the marginal likelihood of the data, $P(\underline{x} \mid M_j)$, all of which are computationally intensive. For this analysis, I adopted the approach proposed by Gelfand and Dey (1994): $P(\underline{x} \mid M_j) = \left[\frac{1}{n} \sum_{i=1}^{n} \frac{h(\theta_i)}{P(\theta_i \mid \underline{x})} \right]^{-1}$, where *n* is the number of Markov Chain Monte Carlo (MCMC) samples, $P(\theta_i \mid x)$ is the posterior probability of parameter vector *i* for model *j*, and $h(\theta_i)$ represents an importance function evaluated at parameter vector *i* (a multivariate normal density centered at the posterior mode, with a covariance matrix estimated from the MCMC chain). Aside from computational challenges, the primary problem in using Bayes factors is specifying reasonable prior distributions. Bayes factors are known to be extremely sensitive to the choice of priors (Kadane and Lazar 2004), and only proper priors may be considered (they must be valid probability distributions that integrate to 1.0; Kass and Raftery 1995). If no information about model parameters is known, a default option is that non-informative prior distributions such as Jeffrey's prior may be constructed (Kass and Wasserman 1996). What makes the Bayes factor different from the other criteria discussed here is that it does not explicitly include a term that quantifies model complexity – this term still exists, however, and overly complex models are penalized in the marginal likelihood calculation (see Dawid's discussion in Spiegelhalter et al. 2002).

There are several important differences between AIC, BIC, DIC, and Bayes factors (summarized in Table 2.1). All of the frequentist and Bayesian criteria are alike in that model selection is linked to parameter estimation. In the calculation of AIC and BIC, parameter estimation is done by maximization in an attempt to find the single best point estimate. Parameters for DIC and Bayes factors are estimated using Bayesian methods, which integrate rather than maximize over the parameter space (Hobbs and Hilborn 2006). While both DIC and Bayes factors incorporate parameter uncertainty and correlation in the sampling of the joint posterior distribution , one criticism of AIC and BIC is that neither consider parameter uncertainty in their calculations. This criticism is not a frequentist-Bayesian issue; several frequentist criteria, including the Information Complexity Criterion (ICOMP, Table 2.1; Bozdogan 2000), do include parameter correlation and uncertainty, however these methods have seen little use in biology. In principle, the calculation of ICOMP is similar to AIC, $ICOMP = D(\underline{\hat{\theta}}_{MLE}) + \ln(n) \cdot C(\Sigma)$, where $C(\cdot)$ represents a complexity function based on the covariance between parameters (Σ).

From a philosophical point of view, there are also differences in the objective of each model selection criterion considered here. Myung (2000) divided model selection tools into two groups: generalization-based criteria, and explanation-based criteria. Generalization-based criteria (AIC, ICOMP) seek to find the best model that fits both current data in addition to hypothetical future data that may be observed from the same process that generated the original sample. This predictive quality may be partially responsible for the tendency of AIC favoring models that are too complex (Burnham and Anderson 2004). Explanation-based criteria (BIC, DIC, Bayes factors) are only concerned with correctly identifying the process from which the data arose, and are not influenced by hypothetical future observations. Many biologists, particularly those working with natural resource management, may be more interested in predicting future quantities, rather than explanation (e.g. Fried and Hilborn 1988). Prediction-based Bayesian model selection tools exist for these types of problems (Madigan and Raftery 1994; Gelman et al. 1995; Bernardo and Smith 2000), but will not be discussed here because the current analysis is only concerned with explaining population processes.

A final important difference between the model selection tools is that other than Bayes factors, none of the criteria presented were ever intended to be used as decision tools, yet they are routinely used by biologists to make decisions (or provide decision making advice to managers). An argument could be made that BIC is compatible with decision making because it is an approximation to the Bayes factor, however the approximation only holds when the prior implied by BIC is thought to be reasonable (Raftery 1999). If a researcher wished to incorporate additional information into a model by developing priors based on expert opinion, a Bayesian criterion other than BIC would have to be chosen. Proponents of both AIC and DIC strongly discourage using model weights for decision making (Burnham and Anderson 2002; Spiegelhalter et al. 2002). When weights are used to make decisions, there may be inferential consequences and costs that many researchers are not be aware of. For comparative purposes, assume that two nested modes are to be compared. One traditional approach for comparing these models might be to view likelihood ratio tests (LRTs) as decision rules. For two models (M_1, M_2), the likelihood ratio in favor of model M_1 is computed as $\lambda(x) = L(x \mid M_1) / L(x \mid M_2)$. This statistic rejects model M_1 when $\lambda(x) \leq C$, C representing the critical region of the test (note: this can also be rewritten as the ratio of the cost of a Type I error to the cost of a Type II error; Bernardo and Smith 2000). The significance level of the LRT (α) represents the probability of rejecting M_1 when M_1 is actually true, and is generally set *a priori* at 0.05. The relationship between AIC and LRT in a testing framework has been reviewed extensively in the literature (Forster 2000; Burnham and Anderson 2002; Kuha 2004). In this framework, the significance level of AIC varies as a function of the complexity between the models considered (with 1 degree of freedom, α is automatically fixed at 0.157 rather than 0.05). The relationship between BIC and LRTs is more complex because the significance of the test is a function of both the difference in model complexity, and sample size. With small sample sizes (~ 7.4), BIC gives similar results to AIC, but the significance level decreases rapidly with increasing sample size or an increasing difference in complexity (BIC eventually favoring simpler model than would be chosen by LRTs). Forster (2000) summarized the differences between AIC and LRTs by illustrating the fact that LRTs give more weight to simpler models than AIC. The performance of DIC relative to other criteria is still being investigated (e.g. Ellison 2004), however the performance of Bayes factors relative to other criteria has been well studied. Bernardo and Rueda (2002) showed that Bayes factors converge to AIC when the precision of the prior is close to the precision of the likelihood (the prior variance is a function of the sample size). This scenario is rather unlikely in ecological problems; as more data is collected, researchers would expect the precision of the prior to decrease. Although the relationship between BIC and Bayes factors has already been discussed, further comparisons in a hypothesis testing framework may be found in Kass and Raftery (1995).

The purpose of this analysis is to address the performance of four model selection criteria (AICc, BIC, DIC, Bayes factors) applied to simulated population abundance data. The majority of previous studies evaluating the performance of model selection criteria may not be completely applicable to biological data, either because of the criteria compared, or the sample sizes involved. For example, Myung (2000) conducted numerous simulations for polynomial models, but did not consider the small sample variant of AIC (AICc), which has seen widespread use in ecology (Burnham and Anderson 2002). Other simulations have focused on sample sizes of more than 100 (Kuha 2004) which may be unrealistic for the natural sciences – many biologists work with sample sizes an order of magnitude smaller. Research with biological models utilizing few samples has been explored in a maximum likelihood framework (Shono

2000), however few studies to data have bridged the gap between maximum likelihood and Bayesian model selection criteria (e.g. A'Mar 2004; Ellison 2004).

All four criteria will be evaluated across multiple scenarios, and will be evaluated in their ability to (1) detect non-linear density dependence, (2) detect Allee effects (or inverse density dependence; Courchamp et al. 1999), (3) detect evidence for multiple life stages, and (4) detect whether there is more evidence for density dependence occurring in survival or fecundity parameters. In Monte Carlo comparisons of model selection performance, the true model is almost always considered among the candidate models (McQuarrie and Tsai 1998; Forster 2000; Myung 2000). While the goal of Monte Carlo comparisons is to estimate the frequency of selecting the true model, the goal of model selection applied to real data is to find a model that best approximates truth. If the simulation-based rules were applied to inference in the real world, all criteria would be wrong 100% of the time. Despite these issues, understanding the performance of model selection over many data sets is extremely valuable – not only can the performance of alternative criteria be compared, but it is also possible to examine the influence of different kinds of error on model selection.

Methods

In this analysis, I considered the ability of four model selection tools to select among seven discrete time population models. The first four population models represent single stage models: (1) the geometric model $N_{t+1} = N_t(1+r)$; (2) logistic model $N_{t+1} = N_t + rN_t(1 - N_t / K)$; (3) theta-logistic model $N_{t+1} = N_t + rN_t\left(1 - (N_t / K)^{\theta}\right)$ (Gilpin and Ayala 1973); and (4) the Allee model proposed by Lewis and Kareiva (1993),

$N_{t+1} = N_t + rN_t(N_t - a)(K - N_t)/K^2$. In these models, the parameter r represents the growth rate, K represents the carrying capacity, θ represents the strength of density dependence, and a represents the critical depensation point, below which per capita growth becomes negative. Three additional models were considered, using variations on a 3-stage transition matrix, $\begin{bmatrix} 0 & 0 & F_3 \\ s_1 & 0 & 0 \\ 0 & s_2 & s_3 \end{bmatrix}$. In this model, the parameter s_1 represents the probability of surviving from stage 1 to stage 2 (e.g. juvenile to sub-adult), s_2 represents the probability of surviving from stage 2 to stage 3 (e.g. sub-adult to adult), s_3 represents the probability of mature individuals surviving to the next year, and F_3 represents the fecundity of mature individuals. In order to assist convergence of MLEs, I assumed that $s_3 \geq s_2 \geq s_1$. Three parameterizations of this matrix were considered, with density dependence being incorporated into s_1, s_3 or F_3. Each model was assumed to have the same form of logistic density dependence, $v(t) = v_{\min} + (v_{\max} - v_{\min}) \cdot (1 - \sum_x N_{x,t} / K)$, where $v(t)$ represents the value of the function at time t, $v_{\min}$ and $v_{\max}$ represent the minimum and maximum survival (or fecundity) values, respectively, and $N_{x,t}$ represents the abundance of animals in stage x at time t.

In turn, I considered each of the models as the simulation model and generated 2500 time series. Each time series was assigned a random length (n = 10, 15, 20, 25). To make the data as realistic as possible, I included random lognormal observation error (CV_{obs} = 0.1, 0.3, 0.5), and random lognormal process error (CV_{pro} = 0.0, 0.1, 0.2). Each data set was assumed to have the same carrying capacity (K = 1.0E5), and each was initialized from some random fraction of K (N_0 = 0.1, 0.3, 0.5, 0.7, 0.9). When generating logistic data, I varied the growth rate over the range 0.1 to 1.1 (in steps of 0.2) and the density dependence parameter θ over the range 0.7 to 1.3 (in steps of 0.2). For the Allee model, the Allee threshold (a) was also treated as a random fraction of K (a = 0.1, 0.3, 0.5, 0.7, 0.9), subject to the constraint that $N_0 > a$. For the stage-structured model with density dependence in adult survival, adult survival was varied over the range (0.75,0.99) in increments of 0.04, and the quantities ($s_3 - s_2$) and ($s_3 - s_2$) were varied over the range (0.05, 0.20) in steps of 0.05. When density dependence was modeled in fecundity, I varied the minimum fecundity of stage-3 individuals over the range 0.33 to 1.66 in steps of 0.33 and varied the difference between maximum fecundity and minimum fecundity ($f_{max} - f_{min}$ = 0.1, 0.5, 1.0, 2.5, 5.0).

For each estimation model, I assumed that observation error was the only type of error present, and it was assumed to be lognormally distributed (multiplicative). Prior distributions were assigned based on similar models in the literature (e.g. McAllister and Kirkwood 1998). For the single-stage models, uniform priors were placed on K (500, 1.0E6), r (-0.5, 2.5), and the critical Allee point as a fraction of K (0, 0.9). These priors were chosen because they are known to not contain much information with respect to model parameters. Following Punt and Hilborn (1997), the prior on the initial population size was assumed to be Log-Uniform (500, 1.0E6) and the CV of the observation error was assumed to be Log-Uniform (0, 2). For the theta-logistic model, the prior on θ was chosen to be ~ Lognormal (-0.5, 1). This prior has an expected value of 1.0 (linear density dependence), and is favored over the log-uniform prior because the log-uniform prior assigns more weight to small values of θ. For the stage-structured model, priors on N_0, K, and the observation error CV were identical to those priors used in the single-stage model. Placing priors on the survival rates was difficult because of the model assumptions ($s_3 \geq s_2 \geq s_1$). A Uniform (0.6, 1.0) prior was placed on the adult survival rate (s_3). Similar priors cannot be placed on s_2 and s_1 - instead, Uniform (0, 0.4) priors were placed on the quantities ($s_3 - s_2$) and ($s_2 - s_1$). A small number of MCMC parameter draws ($<$ 0.1%) were discarded because s_1 became negative, however because the population growth rate was positive, this was generally not a problem.

After each of the new data sets was generated, the seven models were in turn treated as the estimation model and fit to the simulated data set. Rather than estimate both process error and observation error, I assumed that the estimation models only contained observation error. The form of the error was assumed to be multiplicative and lognormally distributed, just as in the simulation model. Maximum likelihood parameter estimation was done in AD Model Builder (Fournier 1996), and the output from each estimation procedure was used to calculate AICc and BIC. AD Model Builder was also used to do Metropolis-Hastings MCMC sampling (Gelfand et al. 1995). A small number of data sets were used to determine whether MCMC chains converged (Best et al. 1995). Results from this analysis were used to choose a burn-in length of 5.0E5 samples, followed by a chain length of 1.0E6 samples. Every 200th sample was retained from the MCMC chain, resulting in 5000 posterior vectors for each estimation model. Following

MCMC sampling, the DIC value and marginal likelihood for the Bayes factor were calculated.

For each comparison between estimation models, the normalized model weights for AICc, BIC, and DIC were computed following the general method for computing AIC weights (Burnham and Anderson 2002). After the ΔAIC values are computed for each model relative to the model with the lowest AIC score, the normalized weight for the ith model can be calculated as $w_i = \frac{\exp(-0.5 \cdot \Delta\mathrm{AIC}_i)}{\sum_j \exp(-0.5 \cdot \Delta\mathrm{AIC}_j)}$. Although some studies have presented these model weights as Bayesian posterior probabilities (e.g. Wintle et al. 2003), and others have suggested that AIC may be justified in a Bayesian framework (Burnham and Anderson 2004), model weights are most commonly considered in a frequentist framework, where they are not treated as probabilities. After computing the posterior model probability via the Bayes factor for each candidate model, I normalized the posterior model probabilities to allow comparison to the model weights.

Results

The first comparison I examined was the ability of each model selection criteria to detect density dependence, and the rate with which each was subject to Type II error. In this comparison, Type II error represents instances where more weight is assigned to an estimation model that includes density dependence when the simulation model is geometric. Across all simulated data sets that included density dependence, I computed the frequency with which each criterion gave a model weight or posterior model probability greater than 0.5. Using a cutoff value of 0.5 is somewhat arbitrary, however it was chosen because it represents the point at which one model is favored over all others. The Bayes factor and DIC were able to detect density dependence in almost all data sets (99.7%, 99.3% respectively), while AICc and BIC did slightly worse (80.3%, 85.3% respectively). One potential reason for the difference between AICc and BIC is the assumption that each makes about whether the true model is included among the candidate models. McQuarrie and Tsai (1998) observed that BIC outperformed AICc when the simulation model was among those considered in the list of candidate models. Even though the growth rates in data generated from the geometric model were small (0.01-0.1), both the frequentist and Bayesian criteria had low Type II error rates (all < 4%).

As a second comparison, I examined the ability of each model selection criteria to detect non-linear density dependence. For simplicity, this involved only the theta-logistic and logistic population models. Before comparing the performance of the frequentist and Bayesian criteria, I examined the inferential consequences of choosing a particular summary statistic (mean or mode) in the calculation of DIC and the Bayes factor. For each summary statistic, I calculated both components of the DIC calculation ($\overline{D(\theta)}, D(\hat{\theta})$) as a function of that statistic (Spiegelhalter et al. 2002). Two versions of the Bayes factor were also calculated – one with the multivariate normal importance function centered on the posterior mode, and the other with the importance function centered on the posterior mean. When the mean likelihood and parameter means were used in the calculations of DIC, more weight was given to the complex model (Figure 2.3). If the mode of the deviance is used in place of the mean deviance, and parameter modes are used instead of the parameter means, DIC tends to favor the simpler model. One problem with using the posterior mode in the calculation of $\overline{D(\theta)}$, however, is that more than 80% of the time series resulted in negative estimates of model complexity (p_D). Using the parameter means in the importance function of the Bayes factor tends to favor the more complex model, while there appears to be much more uncertainty in which model is favored when the posterior mode is used.

To understand why DIC may favor complex models, I examined the distribution of bias ($\hat{\theta} - \theta$) for each parameter in the theta-logistic model. Of all parameters, the largest bias was observed in the density dependence parameter θ. The posterior distribution of θ is often skewed, with the posterior mean greater than the mode (Figure 2.4). Although the prior on θ is centered at 1.0 (the point where density dependence is linear), varying the standard deviation of the lognormal distribution largely impacts DIC. As the standard deviation of the prior increases, the expected value of the posterior of θ decreases, influencing both the mean deviance and deviance at the mean (Figure 2.5). The log-likelihood at the mean appears to be unstable as the prior CV increases because of the negative correlation between the parameter θ and the logistic growth rate, *r*. Large values of θ in particular have negative impacts on model stability, and allow only small growth rates to be considered. It is possible that the lognormal prior may be a poor choice for θ, or that the assumption about its location may be wrong (it may be better for the prior to have an expectation that is less than or greater than 1.0).

When the frequentist model selection tools are compared to their Bayesian counterparts in their ability to detect non-linear density dependence, it appears that both

AICc and BIC tend to favor the simpler logistic model, even when data are generated from a theta-logistic process (AICc favors the simpler model slightly more; Figure 2.6). Although there is considerable uncertainty associated with the Bayes factor posterior probabilities, Bayes factors have the highest rate of successfully identifying the true model. The behavior of DIC is similar to that of AIC in that it tends to favor the simpler model, favoring linear density dependence over non-linear density dependence.

A third comparison investigated here was the ability of each criteria to detect Allee population dynamics. For this comparison, I examined the performance of the logistic, theta-logistic, and Allee models on data generated from those three processes. The performance of AIC, BIC, and DIC was similar in that when the logistic and theta-logistic models are considered together against the Allee model, all three criteria tend to give less weight to the Allee model, even when the Allee model generates the data (Figure 2.7). This same result held when the theta-logistic model was compared alone against the Allee model, which is surprising because the Allee and theta-logistic models have the same number of parameters ($n = 5$). The posterior model probabilities estimated by the Bayes factor leads to different results (Figure 2.7). First, there is less uncertainty in whether or not Allee dynamics are present. Second, the Bayes factor gives slightly more weight to the Allee model, which is the opposite of the other criteria.

While stage-structured or age-structured population models are more realistic than single-stage models, estimating population parameters from a single time series is often difficult, if not impossible. Life history parameters may be estimated from other data sources and treated as fixed constants, or assigned informative Bayesian prior distributions based on existing knowledge. In this analysis, I wanted to determine whether it was possible to use model selection criteria to distinguish single-stage dynamics from multi-stage dynamics. A secondary goal was to determine whether it was possible to differentiate between density dependence in survival parameters, and density dependence in fecundity parameters. Unfortunately, the results from these simulations are inconclusive. On the surface, it appears that all model selection criteria strongly favor the simpler single-stage model over the stage-structured model (Figure 2.8). Upon closer inspection, however, both maximum likelihood and Bayesian parameter estimates failed to converge for many of the data sets. Of the 7,500 data sets generated from stage-structured population models, only a small fraction (< 10%) had both types of parameter estimates successfully converge for all models.

Discussion

There are several approaches to model inference – multiple models might be compared with the goal of selecting a single best model, or model averaging might be used (Hoeting et al. 1999; Burnham and Anderson 2002). Although model averaging is desirable when the true model is unknown, the goal of this analysis is to examine the distribution of estimated model weights (or model probabilities) favoring the true simulation model. It is important to note that the results in this analysis are specific to the types of data used, choice of time series length, as well as the priors that have been assigned to all model parameters.

Because each model selection tool has different intrinsic assumptions and behavior, it is crucial to understand the differences between each. Before using any of the model selection criteria discussed in this paper, every biologist should ask themselves the following questions to help determine which criteria are most appropriate: (1) Is the purpose of your analysis to make predictions, or to decide which model best represents reality? (Ghosh 1999); (2) How confident are you that the true model is being considered among the subset of candidate models? If the true model is not among those considered, are you willing to accept a model that is 'close enough'? (Forster 2000); (3) Does your model contain hierarchical parameters, random effects, or highly correlated parameters? (4) Is anything known about biases in model parameters? While this might be impractical in reality, parameter bias has been shown to impact model selection (Forster 2000). Final questions are specific to Bayesian methods: (5) If using DIC, does the posterior mean a reasonable estimator? If not, is there a better alternative? (e.g. Celeux et al. 2006); (6) How sensitive is the chosen criterion to your choice of prior distributions on model parameters?

The first important result from this analysis is that despite different origins and computation of model complexity, AICc and BIC have remarkably similar behavior. AICc tends to favor simpler models to a slightly greater extent than BIC, however both criteria strongly favor simpler models in all comparisons considered here. This result has implications for model inference – if a researcher attempts to use either to detect non-linear density dependence or Allee effects, both criteria will tend to favor models without these processes. To illustrate this, consider a simple comparison between the logistic and theta-logistic models. The logistic model has one fewer parameter (it assumes θ is fixed at 1), but because there is a strong negative correlation between the parameters r and θ, a logistic model can fit theta-logistic data almost as well as the theta-logistic model by simply changing the growth rate. When the difference in log likelihood values between the two models is smaller than the difference in their complexity (AIC = 2, BIC = ln(K)), the simpler logistic model is assigned more weight. Even for a time series with precise abundance estimates (CV ~ 0.01), a population exhibiting non-linear density dependence ($N_0 = 200$, $r = 0.2$, $K = 1000$, $\theta = 0.7$) would have to be continuously observed for 25 years – to more than 90% of K – before the theta-logistic model would receive more AIC or BIC weight than the logistic model.

A second point is that this paper raises more questions than provides answers considering the use of DIC. Kadane and Lazar (2004) illustrated that one concern about the use of Bayes factors is that they are sensitive to the choice of priors on model parameters. Like the Bayes factor, DIC is strongly sensitive to the choice of priors. Highly informative priors for parameters that are difficult to estimate, such as the theta-logistic parameter θ, tend to influence both the posterior and the DIC statistic (Figure 2.5). In the comparison between logistic and theta-logistic models, decreasing the variance of the prior on θ has the effect of making DIC smaller, giving more model weight to the theta-logistic model. A second point about DIC is that when posterior mode is used in the calculation the average model performance ($\overline{D(\theta)}$), most of the estimates of model complexity tend to be negative. As negative values of p_D generally indicate poor model fit and are not interpretable (Spiegelhalter et al. 2002), using summary statistics other than the mean to calculate DIC is not advised. A third potential issue is that while DIC is supposed to be a Bayesian analogue of AIC, it may behave quite differently. Recent studies have indicated that DIC may favor complex models to a greater extent than AIC (*Link et al., In Press Ecology*).

While no one model selection tool is superior to all others under all conditions, the Bayes factor appeared to do slightly better than other model selection criteria in minimizing the total error (both Type I and Type II). For example, examining the distribution of weights greater than 0.5 illustrates that AICc selected the simulation model only 50% of the time, BIC selected the simulation model 49.3% of the time, DIC selected the simulation model 47.6% of the time, and the Bayes factor selected the simulation model 53.1% of the time. This result is true in evaluating non-linear density dependence and Allee effects, however all criteria did poorly in assessing evidence for multi-stage models. This result is promising, because of the four criteria considered here, Bayes factors are most compatible with model averaging and decision making. Although Bayes factors have been studied and used in practice to a greater degree than DIC, more research needs to be done on several components of the Bayes factor, including sensitivity of prior distributions (specifically applied to non-linear population models) and the importance function used in the marginal likelihood calculation. As some biological parameters tend to be skewed, it may be worth examining whether bias is introduced when the multivariate normal distribution is applied to these parameters. A second question of interest may be to study whether it is preferable to estimate the covariance matrix and posterior modes from all MCMC draws, or whether it is better to estimate these quantities based on a subset of the entire sample.

Table 2.1 Summary of various properties of the model selection criteria used in this analysis. In addition to AIC, BIC, DIC, and Bayes factors (BF), the ICOMP criterion has been included to illustrate an example of a maximum likelihood criterion that incorporates relationships between parameters. The property "Complex Models" is meant to include state-space models, or models with hierarchical parameters. The property "Complex Likelihoods" is meant to designate multimodal densities, likelihoods that are composed of mixture distributions, and missing data problems.

	Model Selection Criteria				
Property	AIC	BIC	DIC	BF	ICOMP
Estimation	MLE	MLE	Bayesian	Bayesian	MLE
Model Performance	Best	Best	Average	Integrated	Best
Complex Models	N	N	Y	Y	N
Non-nested Models	Y	Y	Y	Y	Y
Prior	N	Implicit	Explicit	Explicit	N
Parameter Uncertainty	N	N	Y	Y	Y
Parameter Correlation	N	N	Y	Y	Y
Complexity Term	Explicit	Explicit	Explicit	Implicit	Explicit
Explanation-based	N	Y	N	Y	N
Asymptotic Consistency	N	Y	N	Y	Y
Complex Likelihoods	N	N	N	Y	N

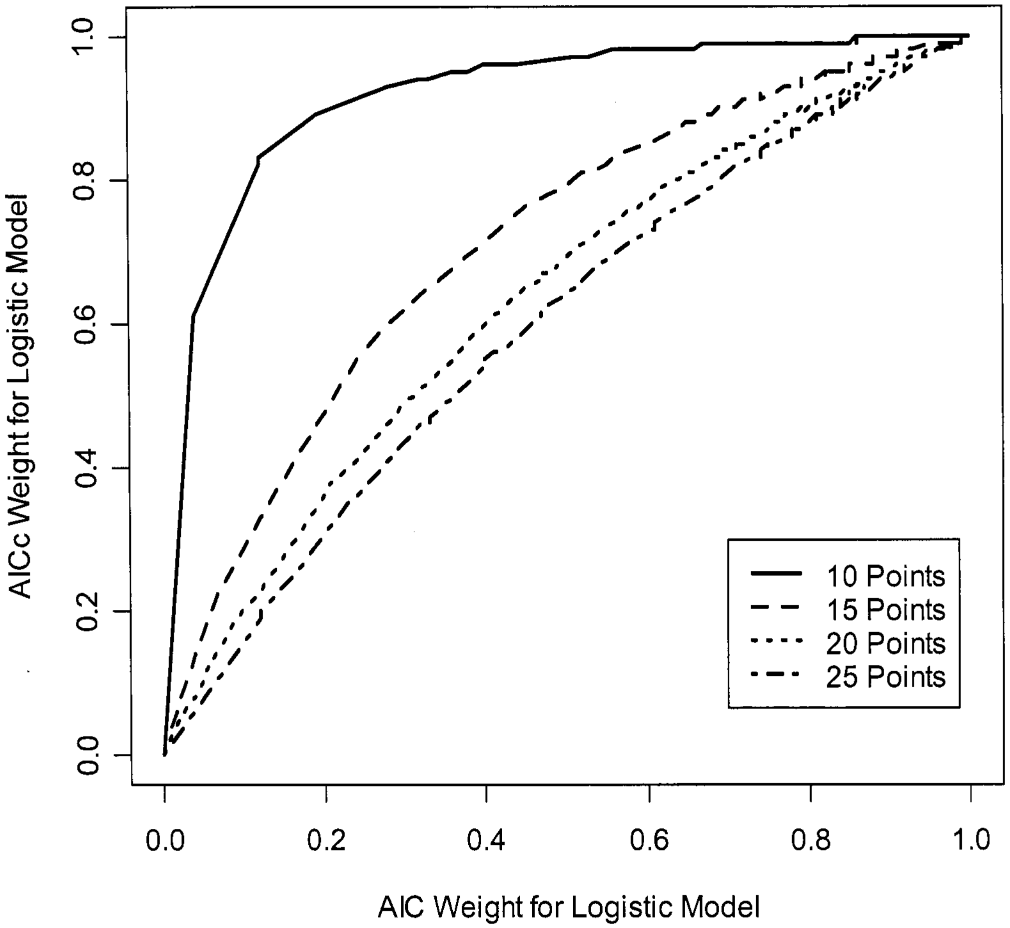

Figure 2.1 The asymptotic relationship between AIC weights and the small sample AICc weights as a function of the number of data points. Data were simulated from a theta-logistic process, and two models were considered: logistic and theta-logistic. When the sample size is large (~ 30) the methods perform similarly, but when less data are available, AIC tends to favor more complex models.

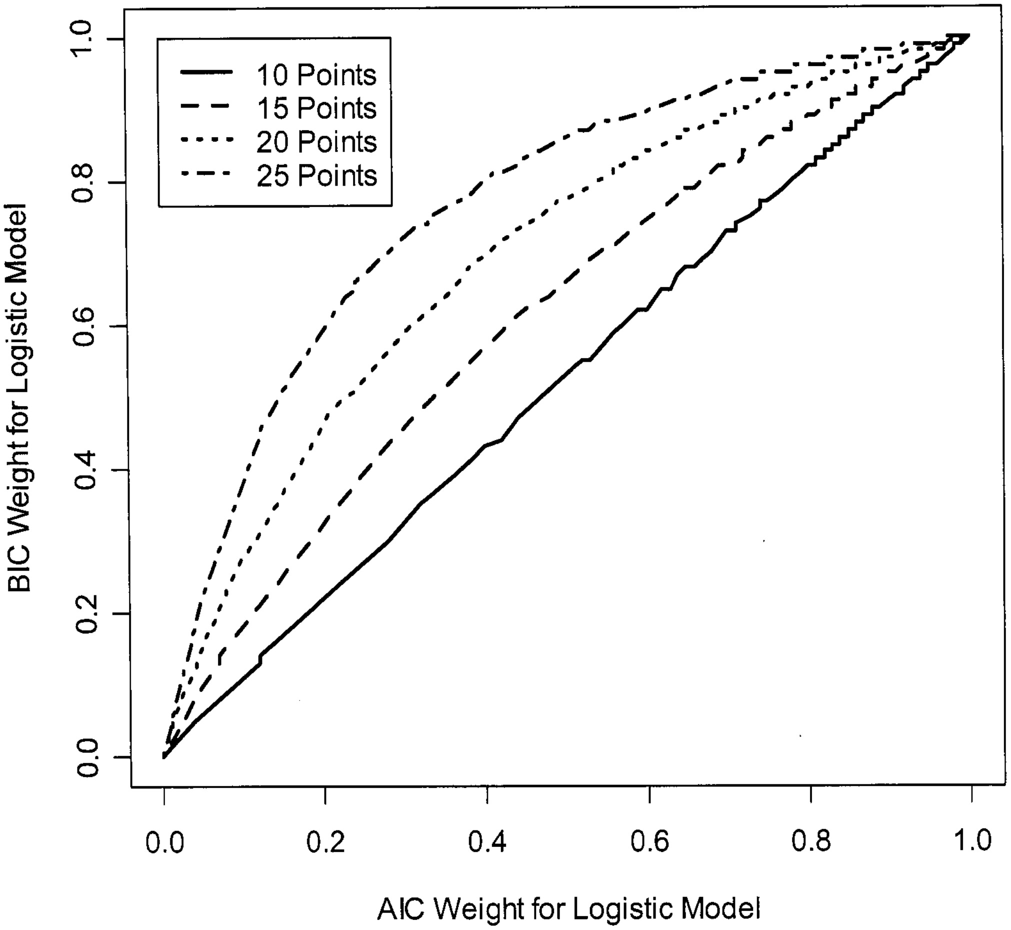

Figure 2.2 The relationship between AIC weights and BIC weights as a function of the number of data points. Data was simulated from a theta-logistic model, and two competing models were considered as estimation models: logistic and theta-logistic. When the sample size is small the methods perform similarly, but as more data is collected, BIC favors the model with fewer parameters.

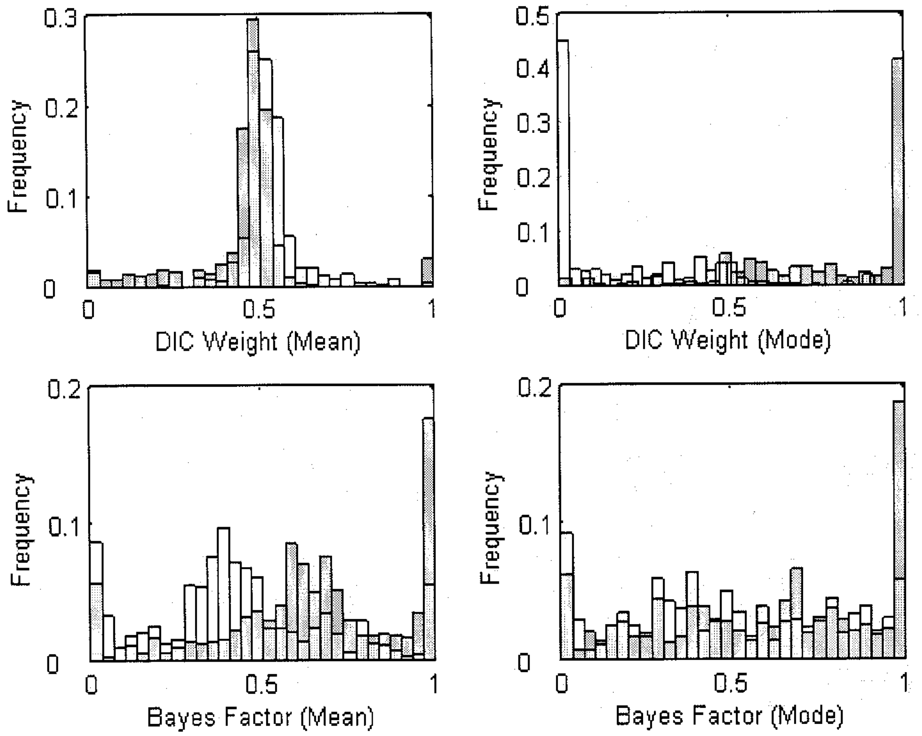

Figure 2.3 Distribution of DIC weights and posterior model probabilities (Bayes factor) evaluated at both the mean and mode for data generated from a logistic model (dark grey bars) and theta-logistic model (white bars). As the dark grey histogram is transparent, the light grey region represents the overlapping area. Each summary statistic (mean, mode) is used in both terms of the DIC calculation ($\overline{D(\theta)}, D(\hat{\theta})$), and each is used to center the importance function in the calculation of the multivariate normal importance function ($h(\theta)$). For both histograms, values close to 1.0 indicate that the true model is favored, while values close to 0.0 indicate that the wrong model is favored. A model selection criterion with low Type I and Type II error rates will have both densities concentrated on the right side of the plot.

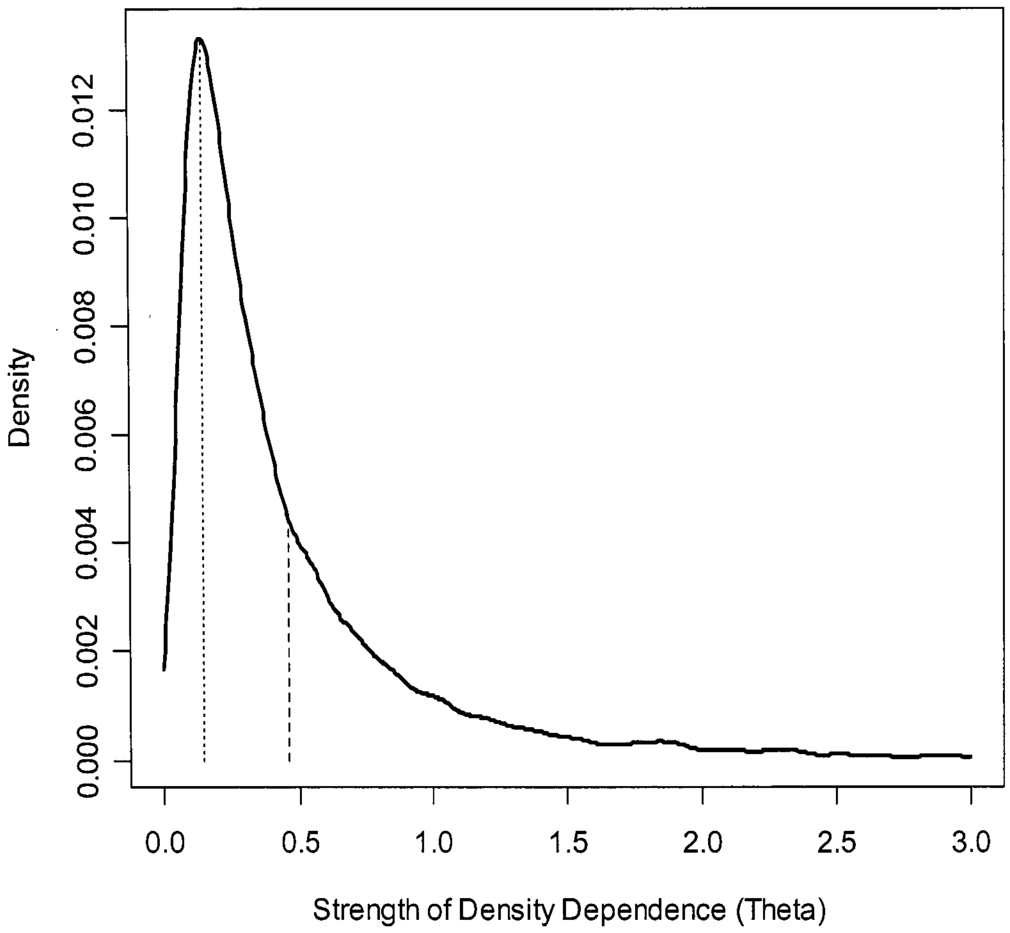

Figure 2.4 Posterior distribution of the density dependence parameter θ, estimated for one of the 2,500 theta-logistic data sets in this analysis (the true value of θ is 1.3) . In this example, the mean (0.463, dashed line) is much greater than the median (0.295) and the mode (0.147, dotted line).

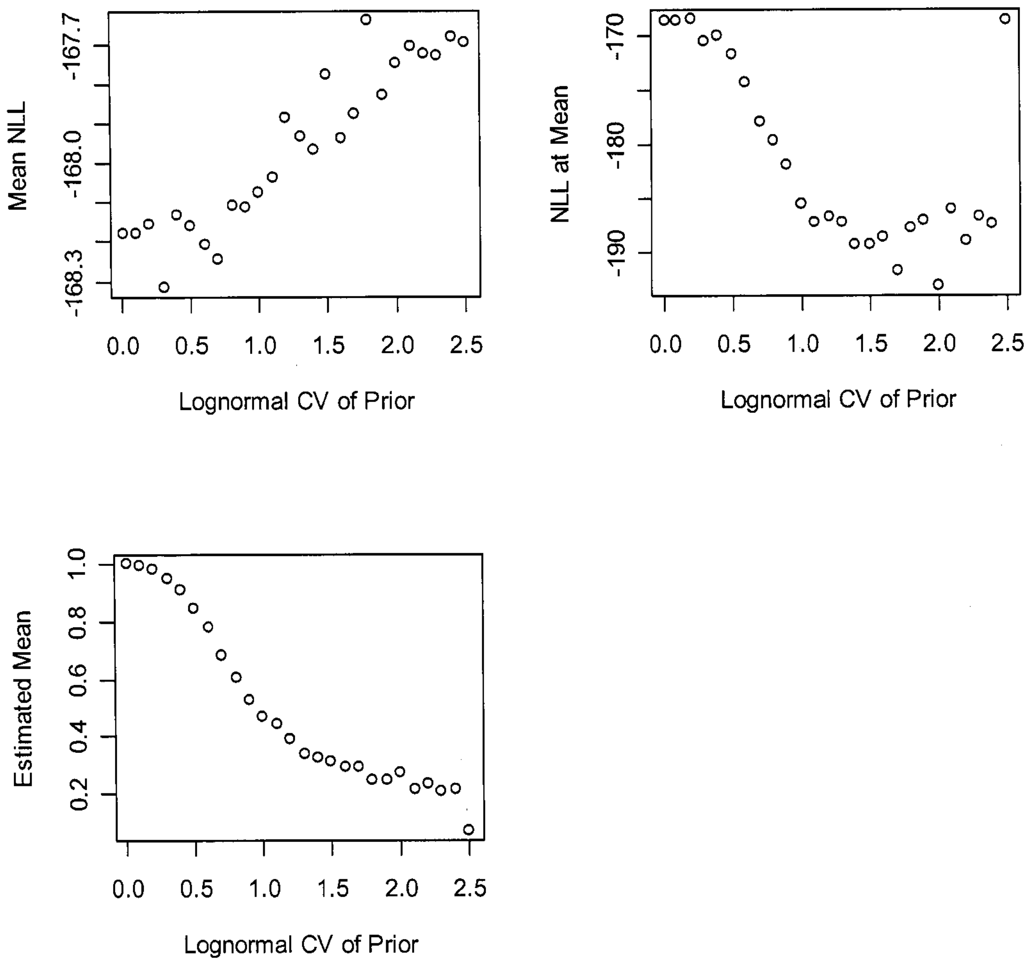

Figure 2.5 Sensitivity of the prior for the density dependence parameter θ on the components of the DIC criterion. The prior on θ is assumed to be lognormally distributed, with an expected value of 1.0. As the prior standard deviation (CV) increases, the expected value of the posterior of θ decreases, resulting in an increase in the parameter bias (the true parameter value is 1.3) and a large difference in the negative log-likelihood (NLL) at the posterior means.

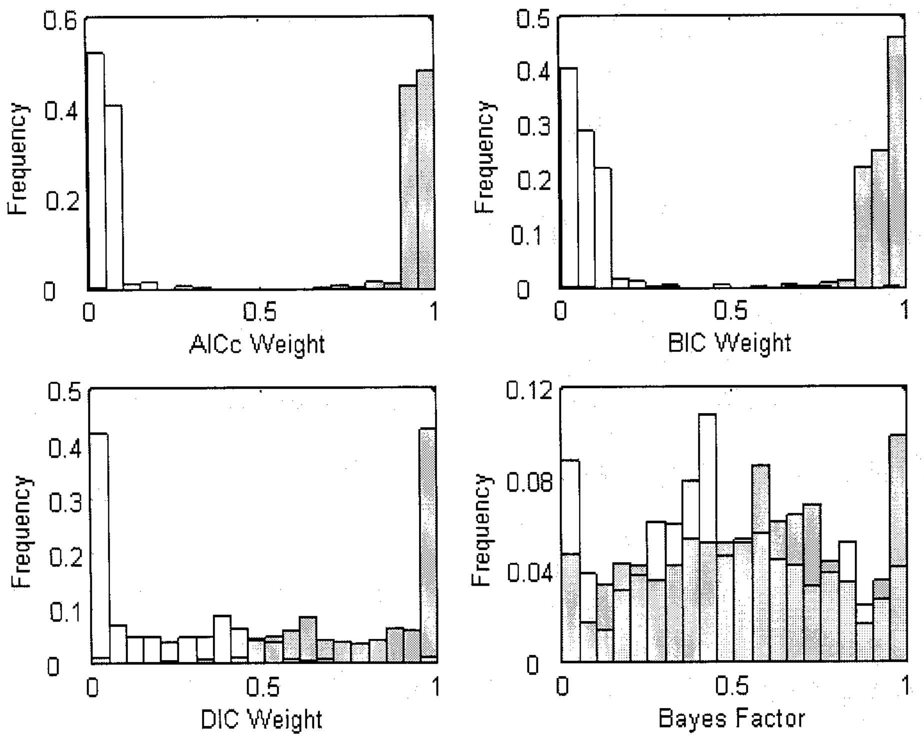

Figure 2.6 Model weights for evaluating the evidence of non-linear density dependence. In each case the model weight (or in the case of the Bayes factor, the posterior probability) represents the frequency of selecting the correct model that generated the data. 2500 data sets have been generated from a logistic model (grey bars) and 2500 data sets have been generated from a theta-logistic model (white bars). For both histograms, values close to 1.0 indicate that the true model is favored, while values close to 0.0 indicate that the wrong model is favored. A model selection criterion with low Type I and Type II error rates will have both densities concentrated on the right side of the plot.

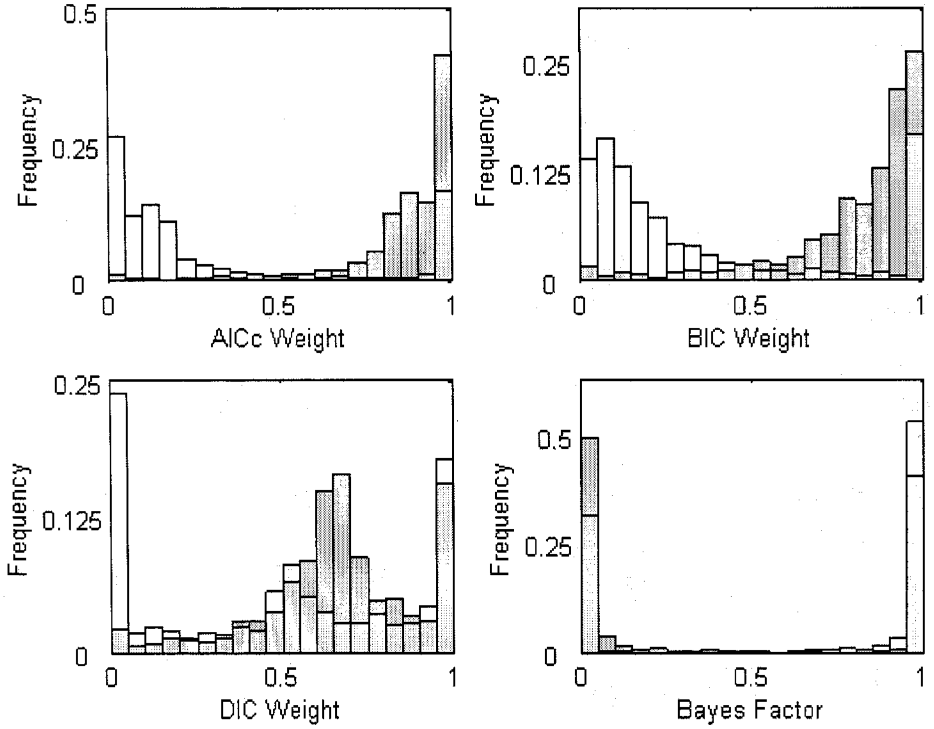

Figure 2.7 Model weights for evaluating the evidence of Allee dynamics (depensation). In each case the model weight (or in the case of the Bayes factor, the posterior probability) represents the frequency of selecting the correct model that generated the data. 5000 data sets have been generated from logistic and theta-logistic models (grey bars) and 2500 data sets have been generated from an Allee model (white bars). For both histograms, values close to 1.0 indicate that the true model is favored, while values close to 0.0 indicate that the wrong model is favored. A model selection criterion with low Type I and Type II error rates will have both densities concentrated on the right side of the plot.

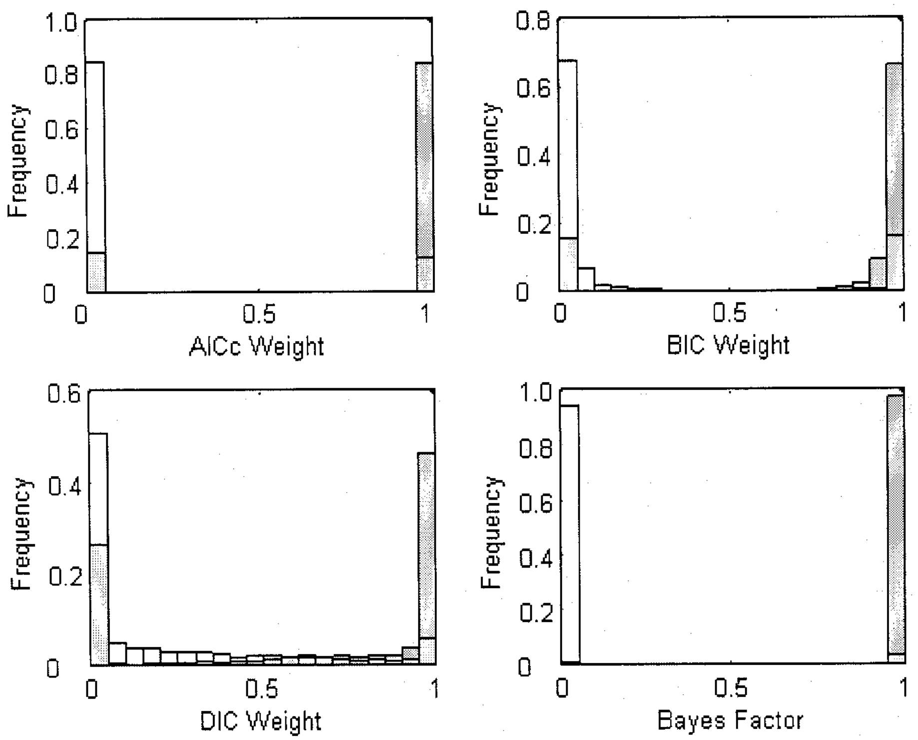

Figure 2.8 Model weights for evaluating the evidence of single stage versus 3-stage population models. In each case the model weight (or in the case of the Bayes factor, the posterior probability) represents the frequency of selecting the correct model that generated the data. 5000 data sets have been generated from single stage models (Allee, logistic, theta-logistic) (grey bars) and 5000 data sets have been generated from a 3-stage models with density dependence occurring in juvenile survival, adult survival, or fecundity (white bars). For both histograms, values close to 1.0 indicate that the true model is favored, while values close to 0.0 indicate that the wrong model is favored. A model selection criterion with low Type I and Type II error rates will have both densities concentrated on the right side of the plot.

Chapter 3: Bayesian Approaches to Harbor Seal Population Dynamics: New Support for Intraspecific Competition

Introduction

The harbor seal (*Phoca vitulina*) is one of the most widely distributed pinneped species in the world, with two of the four recognized subspecies currently inhabiting coastlines of the United States. The subspecies *P. vitulina richardsi* inhabits the eastern Pacific Ocean, ranging from Baja California to Alaska, while the subspecies *P. vitulina concolor* inhabits waters of the western Atlantic Ocean. Since the late 1970s, both subspecies have shown dramatic increases in abundance, and appear to be reaching equilibrium (Brown et al. 2005; Gilbert et al. 2005). While most population sizes prior to the 1970s are unknown, harbor seal populations were probably at low density, in part because of a 1947-1960 bounty that removed an estimated 17,000 individuals in Washington state alone (Newby 1973). Similar bounties were in place on the East coast through the late 1960s (Katona et al. 1993), and are thought to be responsible for the decline of *P. vitulina concolor*. One benefit of these large-scale harvests, in addition to the development of systematic monitoring programs, is that growth rates of *P. vitulina* populations were observed over a wide range of population sizes (Jeffries et al. 1993).

In the eastern Pacific, the management area for *P. vitulina* is divided into six stocks: California, coastal Oregon – Washington, inland Washington (Puget Sound), and three stocks in southeastern Alaska. The Alaskan stocks will not be considered in this analysis, as they have not exhibited similar trends observed in the California, Oregon, and Washington populations (e.g. ver Hoef 2003). Each of the three remaining west coast stocks can be further subdivided into regions: the California stock consists of the

Mainland and Channel Island regions, the coastal Oregon – Washington stock consists of two Oregon regions (Southern, Northern) and two Washington regions (Coastal Estuaries, Olympic Peninsula), and the inland Washington stock can be divided into five regions (Strait of Juan de Fuca, San Juan Islands, Eastern Bays, Puget Sound, Hood Canal; Figure 3.1).

Count data have been collected for many of the harbor seal regions over the last 40 years, making these populations an excellent case study for comparing the performance of traditional likelihood-based model selection methods with Bayesian approaches. These data are generally collected with aerial surveys, and the estimated number of animals observed at any one time only represents a fraction of the total individuals ashore. The ratio of animals ashore to the total population has been assumed to be relatively constant (e.g. Brown et al. 2005), because if it were to exhibit a trend, modeling the population would be extremely difficult. All recent studies investigating the trends of west coast harbor seal populations have used the Akaike Information Criterion (AIC, Burnham and Anderson 2002). Potential alternatives to using AIC are the Schwarz Criterion (BIC; Schwarz 1978), Deviance Information Criterion (DIC, Spiegelhalter et al. 2002) and Bayes factors (Kass 1993). Each of these tools has tradeoffs and differences – most importantly, the Bayes factor is the only tool compatible with decision making and model averaging (Hoeting et al. 1999). Applying each of these tools in turn to *P. vitulina* populations may illustrate potential tradeoffs of choosing any one criterion over alternatives.

The primary objectives of this analysis are to determine whether existing time series provide evidence for intraspecific competition between harbor seal populations, and to quantify the support for or against including competition. Competition likely affects harbor seals via multiple processes on several scales. At the individual level, competition is likely to occur for shared resources (e.g. prey items). Because the breeding system for this species consists of harems, competition also occurs between breeding males for both mates and territories. On an intermediate scale, intraspecific competition may occur between subpopulations or between haulout sites. On a global scale, competition may occur between collections of haul-out sites (regions) or even between stocks.

Previous studies have considered the dynamics of each harbor seal population to be independent and unaffected by competition – while regions may share growth parameters, it has been assumed that growth rate in any particular region is unaffected by the abundance of other regions. Recent observational and genetic studies have reflected uncertainty in the degree to which competition between regions may affect the population dynamics. Despite the fact that recent stock assessments have assumed that movement between regions occurs (NMFS 2000), a 1995 tagging study suggested that movement between Washington regions is rare (Huber 1995). From a behavioral standpoint, harbor seals are known to travel large distances in search of food, and are not limited by geographic barriers, but simultaneously are known for fidelity to haulout sites (Pitcher and McAllister 1981). Recent genetic studies have suggested that harbor seals may be spatially structured in the Pacific on a scale of 600-800km (Westlake and O'Corry-Crowe

2002), however some populations – particularly those in Puget Sound – appear to be genetically isolated from other U.S. populations (Lamont et al. 1996). Additional movement may include individuals immigrating from Canadian waters (Jeffries et al. 1997), but the effect of those individuals on the dynamics of Washington populations remains unknown. For the purposes of this analysis, I will assume that in increase in density within a particular region has a negative effect for surrounding regions (e.g. competition for prey items).

In addition to evaluating evidence for competition between regions, a second objective of this analysis is to evaluate whether there is support for a hierarchical model with random effects in the growth rate. Rather than allow each region to have unique growth rates, the random effects model considers a 'population' of growth rates (Pinhiero and Bates 2000). These growth rates are assumed to have some underlying distribution (e.g. Normal), which may be described with fewer parameters (u_r, σ_r) compared to the number of parameters required to estimate region-specific growth rates (n = 11). For both the individual growth rate model and the random effects model, I will assume non-informative priors with respect to the logistic growth rates. Growth rate data exist for several *P. vitulina* subspecies not found on the west coast (Gilbert and Guldager 1998; Harding et al. 2002; Harding et al. 2003), but these populations experience different environmental conditions and frequency of disease outbreaks than *P. vitulina richardsi.*

A third goal of this analysis is to develop a Bayesian estimate of the density dependence parameter (θ) for the theta-logistic model. The discrete rate of growth in this model can be expressed as $\Delta N = rN_t[1-(N_t/K)^{\theta}]$ (Gilpin and Ayala 1973), where r

represents the maximum growth rate, *K* represents carrying capacity, and θ represents a parameter controlling both the shape of density dependence and return time when the population experiences small perturbations (Figure 3.2; Lande et al. 2003). Understanding the role of θ in population dynamics is also important for management, because it alone controls the point at which maximum sustainable yield (MSY) occurs relative to *K*, $N_{MSY} / K = \exp[-\ln(1.0+\theta)/\theta]$. As the strength of density dependence (θ) increases, the value of N_{MSY} as a fraction of *K* asymptotically approaches 1. While θ is difficult to estimate with a single time series, multiple time series may be combined if the strength of density dependence is assumed to be similar between them. Published estimates of θ suggest a wide range of uncertainty for *P. vitulina*. Jeffries et al. (2003) suggested a maximum likelihood estimate of ~ 2.4 for the Washington stocks (95% CI = 1.07, 8.57), while Brown et al. (2005) estimated a value of 10.1 (95% CI = 1.06, 11.49). If all stocks and regions are combined in a single hierarchical Bayesian model, the posterior variance of θ may be reduced.

Methods

The population dynamics of each of the 11 regions in this analysis was initially assumed to be governed by a theta-logistic population model, $N_{t+1,j} = N_{t,j} + r \cdot N_{t,j} \cdot \left[1 - \left(N_{t,j} / K_j\right)^{\theta}\right]$, where $N_{t,j}$ is the abundance in region *j* in year *t*, and K_j is a region-specific carrying capacity. The parameters *r* and θ represent the logistic growth rate and strength of density dependence, and are not subscripted because they are assumed to be shared between regions. Several extensions of this model were also considered, with varying numbers of growth rate parameters. In the first case, each region was allowed to have a specific growth rate (11 parameters), and in the second case each stock was allowed to have its own growth rate (3 parameters). Models with subscripts on θ were not considered because the parameter θ is generally unstable (whether estimation is done via maximum likelihood or Bayesian methods), and it may be biologically reasonable to assume that the degree of density dependence is relatively constant between subpopulations of the same species that have experienced relatively similar conditions (management, predation, environmental stochasticity, etc.).

Although dispersal of harbor seals between regions may be rare (Huber 1995), I explored three models that included the density dependent effects of intraspecific competition. For each model, I extended the basic Lotka-Volterra model of interspecific competition (Volterra 1926; Lotka 1932) to allow for intraspecific effects. The general form of the interspecific competition model describing the abundance of species *j* in the presence of species *k* is $N_{t+1,j} = N_{t,j} + rN_{t,j}\left(1-\left(\frac{N_{t,j}+a_{jk}N_k}{K_j}\right)^{\theta}\right)$, where a_{jk} represents the conversion factor that translates species *k* individuals into species *j* equivalents (Begon et al. 1996). This model can also be used to describe intraspecific effects by replacing ‘species’ with ‘regions’, and assuming that every region competes against others as strongly as it competes against itself ($a_{jk} = 1$). In the first intraspecific model (‘All’), each region was assumed to be equally influenced by the abundance of seals in all other regions: $N_{t+1,j} = N_{t,j} + r_j \cdot N_{t,j} \cdot \left[1-\left(\left(\sum_{j=1}^{11} N_{t,j}\right)/K_j\right)^{\theta}\right]$. In the second case

('Stock'), I assumed that competition only occurred between regions within a stock (but not between stocks). In the third case ('Neighbor'), I allowed the dynamics of each region to only be affected by the dynamics in adjacent regions (Table 3.1).

For each model without random effects, I assigned Log-uniform priors to the initial population sizes (log(N_0) ~ Uniform (exp(2), exp(9))), and uniform priors to the region-specific carrying capacities (K ~ Uniform (500, 10e04)). Growth rates were assigned Uniform (0,1) priors. Rather than attempt to assign a non-informative prior to the parameter θ, I assumed that the prior for $\theta \sim \text{Lognormal}(-0.125, \sigma = 0.5)$. For the random effects model, I included random effects by assuming that the growth rate of each region was normally distributed, $r_j \sim \text{Normal}(u_r, \sigma_r)$. The global mean growth rate (u_r) was assigned a Uniform (0,1) prior, and the variance of the mean growth rate (σ_r^2) was assigned an Inverse Gamma prior, $1/\sigma_r^2 \sim \text{Gamma (shape} = 4, \text{rate} = 0.1)$. Previous modeling efforts have assumed that only observation error was present (Jeffries et al. 2003), and that the additive error was normally distributed with a constant coefficient of variation (CV). Instead of using a normal likelihood, I assumed that the observation error for all models was multiplicative and lognormally distributed. Inverse Gamma priors were placed on the squared observation error CVs so that
$1/CV^2 \sim \text{Gamma (shape} = 4, \text{rate} = 0.3)$.

Maximum likelihood parameter estimates were calculated for each model using AD Model Builder (ADMB, Fournier 1996). The maximum likelihood estimates were used to calculate AICc and BIC,for each model, in addition to the relative model weights. Bayesian parameter estimation was done via Markov Chain Monte Carlo (MCMC), also in ADMB. After initializing each MCMC chain at a random parameter vector, I ran the chain for 2 million iterations (discarding the first 50% of the chain, and storing every 20th sample). The random effects model was implemented in WinBugs, due to the large number of integrals involved in integrating out the random effects. For each Bayesian model, I calculated DIC based on the vector of estimated posterior means. The marginal likelihood was calculated for each model (Gelfand and Dey 1994), and then normalized to calculate the posterior probabilities for each model.

Results

Among the models that did not include random effects, AICc and BIC differed from the Bayesian methods in assessing both the number of growth rates (constant, region-specific, stock-specific) and the evidence in favor of competition. AICc split the model weight between the model with a constant growth rate and the model with region-specific growth rates (0.51, 0.49 respectively). This result is expected, because Jeffries et al. (2003) observed similar results when examining just Washington populations. Using BIC assigned more weight to the model with a single growth rate (0.67) than the model with stock-specific parameters (0.33). DIC and Bayes factors differed from the maximum likelihood approaches because both strongly supported the most complicated model with region-specific growth rates. When competition was included in the model, both Bayesian methods gave zero weight to the model without competition, while AIC and BIC strongly supported the simpler model without competition.

Comparing the performance of the hierarchical random effects models to the models without random effects was only done using DIC and Bayes factors. The DIC statistic was found to be very unstable across models (in some cases negative complexity terms were estimated). One reason for the poor performance of DIC is that some likelihood surfaces were multimodal (e.g. Spiegelhalter et al. 2002). The marginal likelihood was computed over a range of thinning intervals (10, 20, 100, 200) and MCMC chain lengths (1000, 5000, 1.0E6), and found to be relatively stable for each model. Regardless of the competition scenario included, the posterior probability of each random effects model was much greater than its counterpart without random effects (Table 3.2). Among the models with random effects, there appears to be nearly equal weight assigned between the model without competition ('None') and the model that includes competition between adjacent regions ('Neighbor'). These two models differ in their interpretation of density dependence – in the 'Neighbor' model density dependence is allowed to vary over space and time, while in the 'None' model density dependence is constant through space and time.

As part of the model selection process, a range of priors for the density dependence parameter θ were explored. When random effects are incorporated into the logistic growth rate, and competition between regions is not included, the posterior estimate of θ is greater than 1.0 and similar to previously published estimates (Figure 3.3; Jeffries et al. 2003). Estimating θ becomes more difficult when competition is included in the model because the posterior becomes strongly influenced by the prior. When a lognormal prior with an expectation of 1.0 is compared to a uniform prior with an expectation of 10.0, the posterior for θ becomes centered near the prior expectation

(Figure 3.3). Across models and priors, there is a clear tradeoff between small growth rates and large values of θ.

Because estimates of θ in the 'Neighbor' scenario are suspect, I did not generate model averaged parameter estimates between the two best models. Using the MCMC output from the random effects scenario without competition, I calculated the 95% posterior probability intervals for each population in each time step (Figures 3.5-3.6). With the exception of the Hood Canal and Mainland California regions which appear to be at carrying capacity, all of the west coast harbor seal populations appear to be currently increasing and approaching equilibrium. One potential problem with the fit to the observed count data is that several regions appear to have high autocorrelation in the residuals (Southern Oregon = 0.63, Northern Oregon = 0.47, Washington Coastal Estuaries = 0.42). Similar patterns are found in the competition model between adjacent regions (Southern Oregon = 0.57, Northern Oregon = 0.52, Washington Coastal Estuaries = 0.44), indicating that including competition does little to reduce the autocorrelation. The correlated residuals are likely an artifact of correlation in external covariates (environmental variables, prey availability), suggesting that it may be beneficial to include such variables in future extensions of this model.

Discussion

Comparing traditional model selection criteria (AIC, BIC) to more recently used Bayesian tools (DIC, Bayes factors) illustrates that model inference depends on the model selection tool used. Neither AIC nor BIC appear to support including competition in models of *P. vitulina* populations, while DIC and Bayes factors appear to support models with competition. These model selection criteria also appear to have different views of stock structure – Bayes factors and DIC have given weight to the most complicated model (region-specific growth rates) over time, while AIC and BIC tend to favor models that assume growth rates are more constant across regions. When Bayes factors are used to compare models with hierarchical random effects, there is overwhelming support for modeling growth rates with random effects over modeling growth rates as region-specific parameters. The posterior distributions of the random effects model used in this analysis (Figure 3.4) may be of use for future assessments of west coast populations, or assessments of other *P. vitulina* subspecies.

When density dependence is assumed to have a linear relationship ($\theta = 1$), the growth rate r is alone responsible for the stability of the logistic model (May 1976). One feature of the theta-logistic model is that both θ and r control the stability of the model. When density dependence is assumed to be linear ($\theta = 1$), values of r greater than 2 are necessary to create 2-point oscillations, and the oscillations become chaotic when $r > 2.6$. As θ takes on larger values, the range of growth rates that are stable becomes smaller. To illustrate this, consider the differences between the situation where $\theta = 10$ and $\theta = 20$. From the point of MSY, there is little difference between these values (MSY occurs at 79% and 86% of K, respectively). The range of stable growth rates is strongly affected by θ; when $\theta = 10$, stable growth rates lie in the interval (0, 0.28), but when $\theta = 20$, this range is cut in half (0, 0.14). This result might not have a large impact on Bayesian estimates of harbor seal growth rates because the majority of the posterior density is less than 0.14 (Figure 3.3), however constraints on θ may be necessary for populations that are faster growing.

Regarding parameter estimation, results from the random effects modeling indicate that the parameter θ appears to be more precisely estimated when compared to previously published estimates and confidence intervals (Brown et al. 2005). When competition scenarios are considered, it becomes nearly impossible to estimate θ. This result may be caused by the fact that in the model without competition, the shape of density dependence is controlled solely by the parameter θ. When competition is included, however, the shape of density dependence becomes a function of the model parameters from both competing populations (r, K). If two species or regions are assumed to be competing, the per-capita growth rate of the weaker competitor will experience the effects of density dependence to a greater degree than the stronger competitor (per-capita growth for the weaker competitor will behave as if θ were increased; Figure 3.2) Jeffries et al. (2003) suggested that observing populations at low density is important when estimating θ. Because 1965 population size for each time series was at a different fraction of K, it may be possible in future analyses to quantify the contribution of each time series to the posterior estimate of θ.

Regardless of the model chosen, some of the harbor seal time series used in this analysis appeared to fit the logistic model better than others. Jeffries et al. (2003) presented standard errors for parameter estimates, illustrating that the Hood Canal region in particular had the highest associated observation error CV. When Bayesian methods were applied to the same models in this analysis, a similar result was found. Parameter estimates from the Hood Canal region had the largest variance relative to other regions. The two forces that may be responsible for this result are a data quality (or quantity) issue, as the Hood Canal region was surveyed less frequently than other regions over the period 1978-1999. A second potential impact may be that the Hood Canal experiences different dynamics than other regions. Since 1999, two significant mortality events in Hood Canal have occurred as a result of predation by killer whales (London 2006). In terms of the intraspecific competition scenarios evaluated in this analysis, large removals may result in Hood Canal acting as a population sink (Pulliam 1988), receiving immigration from other regions.

Unfortunately, collecting more count data might not provide insight into the magnitude of intraspecific competition affecting harbor seals in Washington. Competition in these populations likely has indirect effects (competition for prey) and direct effects (competition for haul-out sites and mates). Both types of competition, particularly movement, may be also be density dependent. The best opportunity to quantify density dependent dispersal is likely to occur when the population experiences future depletions. If the population were to experience catastrophic events similar to European populations (e.g. disease, Harding et al. 2002), or chance reductions similar to other west coast otariid species, movement studies could be conducted over a range of population densities to observe potential source-sink dynamics, and how populations respond to those events. In addition to estimating movement and competition between regions between west coast populations, it might be possible to estimate migration to and from both Baja California and British Columbia.

Table 3.1 Matrix of competition between nearest neighbors. ('SMI' = San Miguel Island, 'MC' = Mainland California, 'SO' = Southern Oregon, 'NO' = Northern Oregon, 'CE' = Coastal Estuaries, 'OP' = Olympic Peninsula, 'SJF' = Strait of Juan de Fuca, 'SJI' = San Juan Islands, 'EB' = Eastern Bays, 'PS' = Puget Sound, 'HC' = Hood Canal. A '1' indicates that competition between regions was included in the model, a '0' indicates that competition was not included.

	SMI	MC	SO	NO	CE	OP	SJF	SJI	EB	PS	HC
SMI	1	-	-	-	-	-	-	-	-	-	-
MC	1	1	-	-	-	-	-	-	-	-	-
NO	0	1	1	-	-	-	-	-	-	-	-
SO	0	0	1	1	-	-	-	-	-	-	-
CE	0	0	0	1	1	-	-	-	-	-	-
OP	0	0	0	0	1	1	-	-	-	-	-
SJF	0	0	0	0	0	1	1	-	-	-	-
SJI	0	0	0	0	0	0	1	1	-	-	-
EB	0	0	0	0	0	0	0	1	1	-	-
PS	0	0	0	0	0	0	0	0	1	1	-
HC	0	0	0	0	0	0	0	0	0	1	1

Table 3.2 Table of marginal likelihoods for alternative competition models. The probability of models without random effects (RE) represents the sum of the posterior probability for all three growth rate scenarios (constant growth rate, stock-specific growth rate, region-specific growth rate). All models are assumed to have region-specific initial population sizes, observation error CVs, and carrying capacities.

Competition Scenario	Probability (no RE)	Probability (with RE)
All	<0.001	0.013
Neighbor	<0.001	0.477
None	<0.001	0.502
Stock	<0.001	0.008

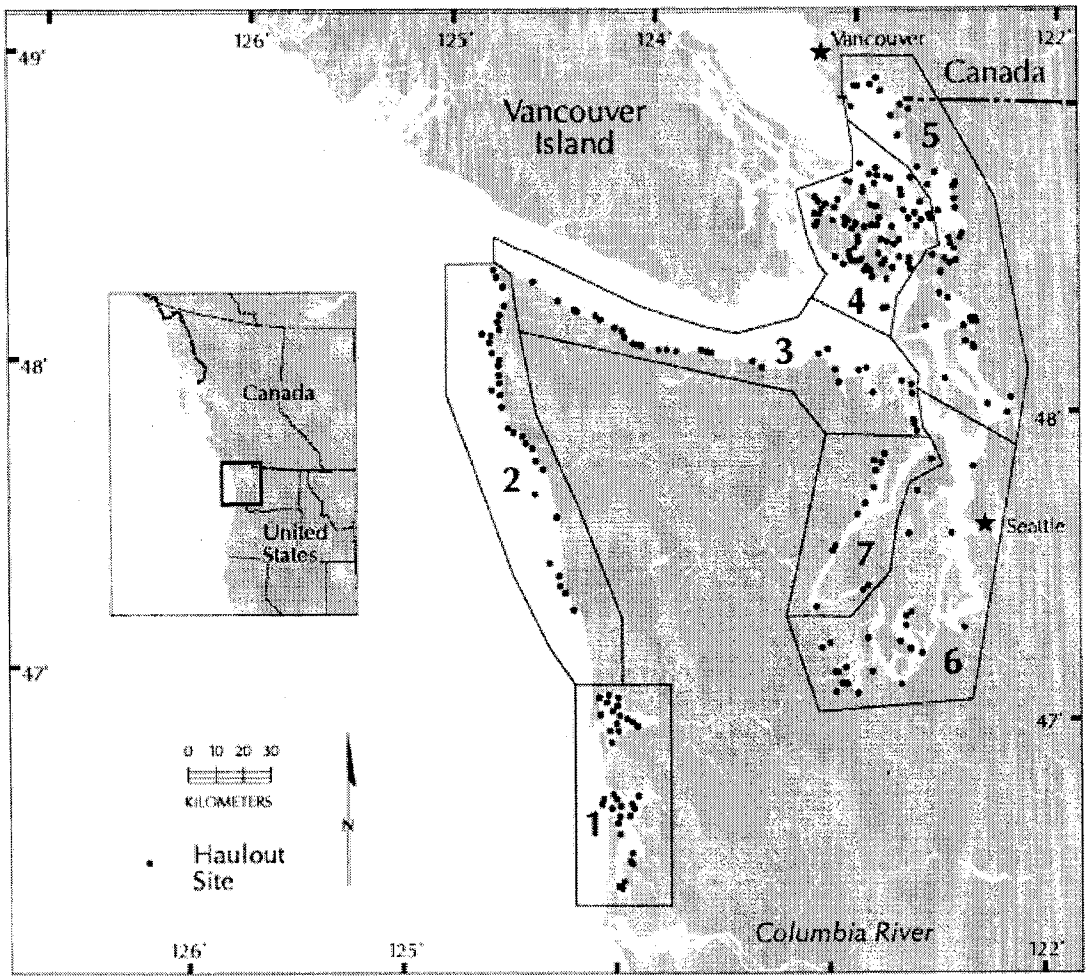

Figure 3.1 Harbor seal regions in Washington State, taken from Jeffries et al. (2003). Regions 1-7 The seven regions are the Coastal Estuaries (1), Olympic Peninsula (2), Strait of Juan de Fuca (3), San Juan Islands (4), Eastern Bays (5), Puget Sound (6), and Hood Canal (7).

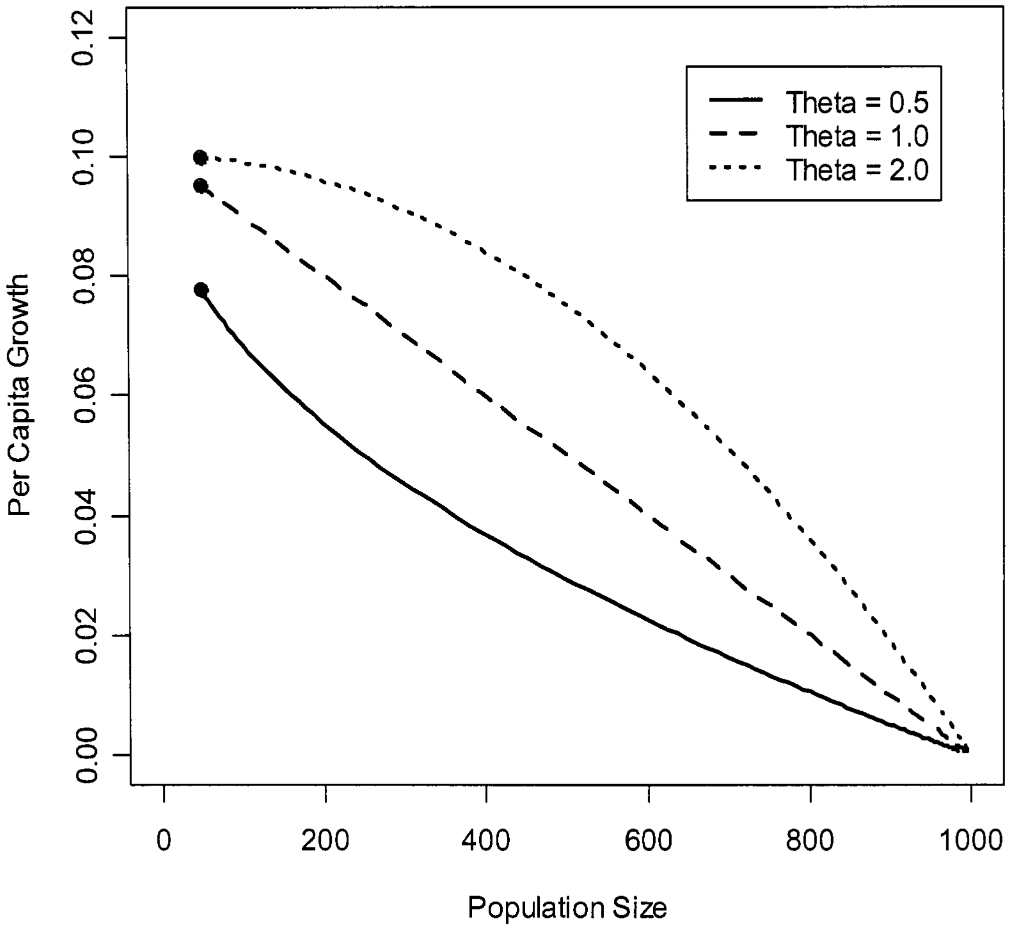

Figure 3.2 Simulated data showing the effect of different values of the density dependence parameter, theta. For all three data sets, the growth rate was assumed to be 0.1, initial population size (black circle) was assumed to be 50, and the carrying capacity was assumed to be 1000.

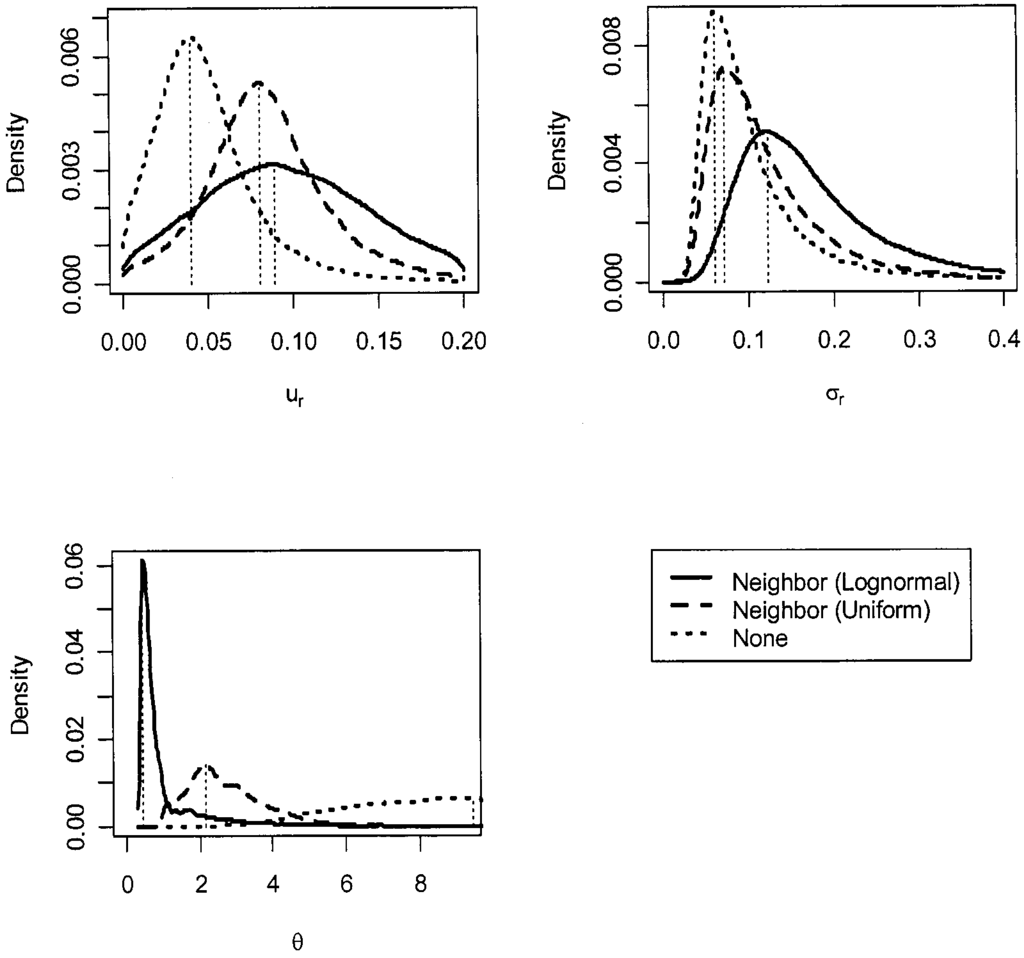

Figure 3.3 Bayesian posterior distributions for the three shared parameters in the random effects model. The global mean and standard deviation of the growth rate (u_r, σ_r) are parameters controlling the random effects; the parameter θ controls the shape of density dependence. Results from three models are shown: a model with no competition ('None'), and two models with competition occurring between regions that are geographically adjacent to one another ('Neighbor'). In the first 'Neighbor' model, θ is assigned a Lognormal(−0.125, 0.5) prior (identical to the prior used in the 'None' scenario) and in the second model, $\theta \sim$ Uniform (0.01, 20).

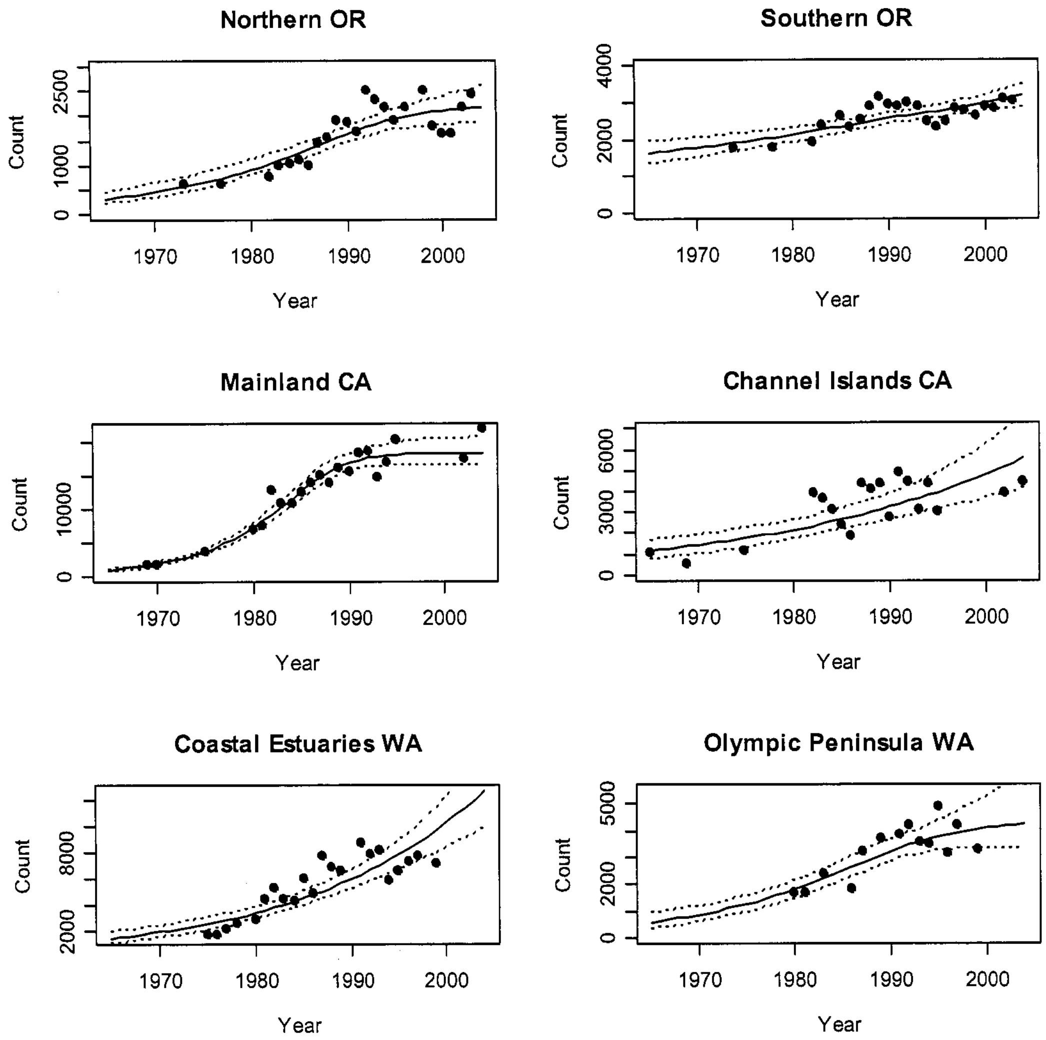

Figure 3.4 Observed and model predicted population sizes for the harbor seal regions located in the California and coastal Oregon – Washington stocks. The results depicted are from the random effects model without competition. The predicted median population size is shown for each year (solid line) in addition to the 95% probability intervals (dashed lines).

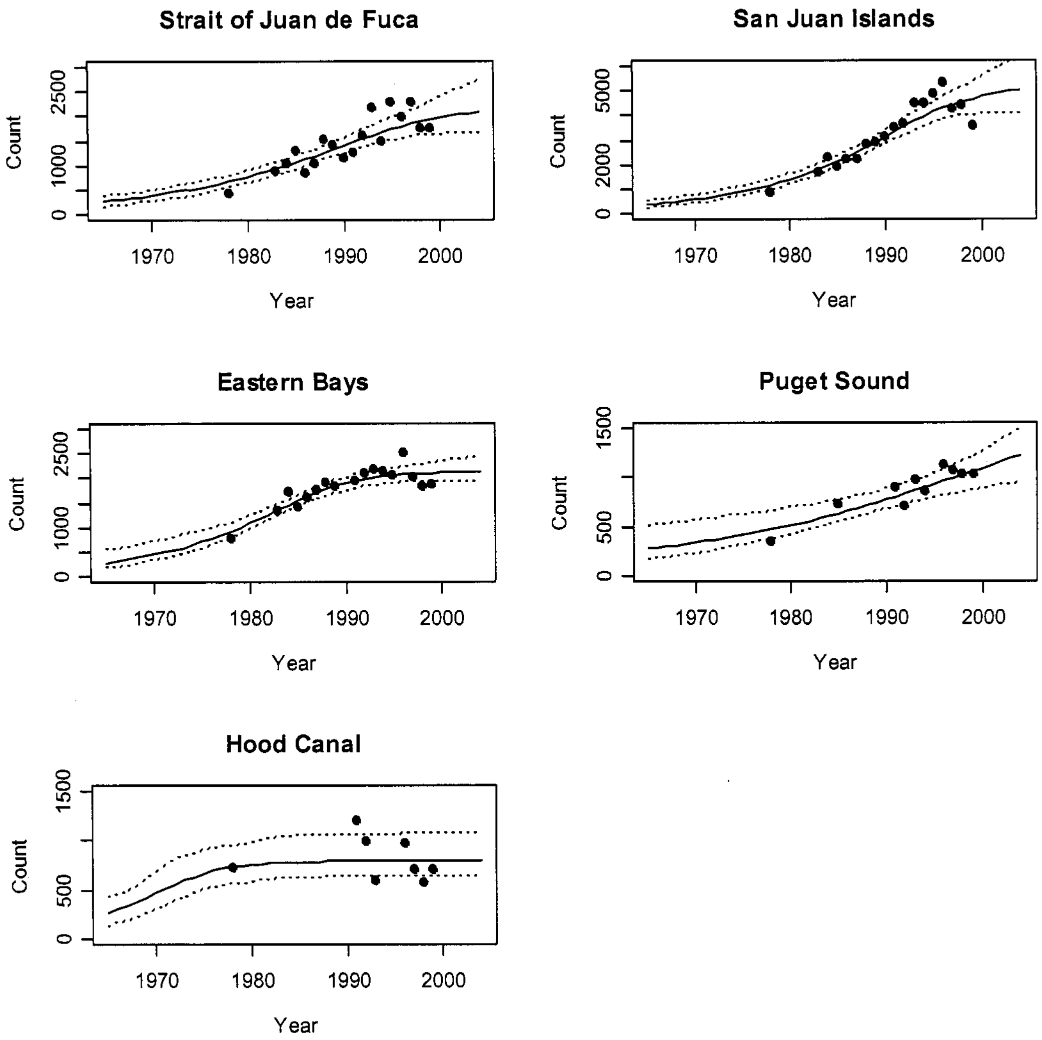

Figure 3.5 Observed and model predicted population sizes for the regions within the Inland Washington stock. The results depicted are from the random effects model without competition. The predicted median population size is shown for each year (solid line) in addition to the 95% probability intervals (dashed lines).

Chapter 4: A State-Space Mixture Approach for Estimating Catastrophic Events in Time Series Data

Introduction

State-space population dynamics models have become increasingly popular in ecology with the advancement of new statistical methods and software capable of computing high-dimensional integrals. The main advantage of using state-space models is that both observation error and process error can be modeled simultaneously. Before the advent of these models, ecologists were forced to model only one of the two processes, assuming the other to be negligible. While this assumption may be valid for some populations, it is likely to introduce bias (Walters and Ludwig 1981). The most common application of state-space models has been to time-series of population abundance data, because understanding natural fluctuations in population size is critical for management (Hilborn and Mangel 1997; de Valpine and Hastings 2002; Calder et al. 2003; Clark and Bjørnstad 2004). While a variety of processes may be responsible for observed fluctuations in abundance, the most important factors determining a species' persistence may be the frequency and magnitude of catastrophic events (Mangel and Tier 1994). Not including these events in stock assessments and population viability analyses (PVA) may create an overly optimistic view of a population's status (Young 1994).

All natural populations are subject to a variety of stochastic processes, including both demographic and environmental variability (Lande et al. 2003). The key difference between these two types of stochasticity is that demographic variability is simply the random process of survival or reproduction, with each individual having the same probability of surviving or reproducing. Environmental variability changes the probability of survival or reproduction for every member of a population, but all individuals still experience the stochastic process of survival and reproduction. In this paper, we assume environmental stochasticity to be the dominant process. Most PVAs and stock assessments assume that the extreme events may be represented by the tail of the process error distribution (Hilborn and Mangel 1997; Figure 4.1). The tradeoff of using single distribution is that the frequency of extreme negative events can only be increased by increasing the variance, thereby also increasing the frequency of extreme positive events. For many K-selected species that experience slow growth, incorporating extreme positive growth events into population models is not biologically reasonable (e.g. even in the presence of unlimited resources, very few large mammal populations would be able to experience an annual doubling in population size). As a more flexible approach, we adopt a mixture distribution to model environmental variation. This mixture consists of a distribution representing the normal process error variability of the population, and a distribution representing the variability that the population experiences during catastrophes. Catastrophes have been given multiple definitions in the literature (Erb and Boyce 1999; Mangel and Tier 1994; Young 1994). We choose to define catastrophes as being rare negative events that affect the probability of survival or reproduction in a single time step, and reduce a population by more than 50%. Some populations that have experienced steady declines over multiple years, such as the Western Aleutian stock of Steller's sea lion (Angliss and Lodge 2002), would not be considered catastrophic under our definition.

A variety of approaches have been developed to address parameter estimation in a state-space framework. Numerical search algorithms such as the Kalman Filter (Kalman 1960) are limited to linear problems with normal error distributions. More recently, restricted maximum-likelihood (REML) methods have been suggested (Staples et al. 2004), but these approaches also have some limitations, including a constraint that time series do not contain missing observations. An extension of the Kalman approach, the numerically integrated state-space (NISS) method, has been developed as a more general solution to both linear and non-linear systems (de Valpine 2002; de Valpine and Hastings 2002). A third approach to state-space models are Bayesian statistics, which have seen increased use in fisheries science (Meyer and Millar 1999b; Rivot et al. 2004), and will be the focus of this paper. One advantage of dealing with state-space models in a Bayesian or NISS framework is that non-normal error distributions may be applied to non-linear population models (Punt 2003). In a hierarchical modeling framework, Bayesian methods allow for multiple datasets to be combined in an analysis, and prior distributions to be placed on shared population parameters. While Bayesian methods may give similar results to traditional maximum likelihood approaches, the key difference is that Bayesian sampling integrates over the parameter space, rather than maximizing (Hobbs and Hilborn 2006).

The objectives of this paper are to (1) develop a model for incorporating catastrophic events time series of abundance data, (2) evaluate the ability of Bayesian methods to estimate the underlying processes, (3) evaluate the power of this method to detect catastrophic events, and (4) examine what factors influence the precision of Bayesian posteriors in a hierarchical framework.

Methods

State-space models require the formulation of two models, an observation model, and a process model describing the transition between unobserved population states. For time series of population abundance, the observation model relates the set of population abundance estimates based on observed data to the true unobserved states of the population. Typically, the observation model takes the form $\hat{N}_t = S_t \cdot \varepsilon_t$, where $\hat{N}_t$ is the estimated abundance (e.g. from a survey), S_t represents the true unobserved abundance in year *t*, and ε_t is a multiplicative error term. The set of observation error terms, $\underline{\varepsilon}$ are assumed independent between years and for the purpose of this paper are assumed to be lognormally distributed,

$$(1) \qquad \mathrm{P}(\varepsilon_t | S_t, \sigma_o) \sim \text{Lognormal}\left(-\frac{\sigma_o^2}{2}, \sigma_o\right)$$

where S_t is the true population size. Regardless of the form of the error, it is possible to model the error terms directly or model the ratio of the observed population size to the unobserved state of nature in normal space, $\varepsilon_t = \frac{\hat{N}_t}{S_t}$.

The general form of the state equation for the process component of the model is: $S_{t+1} = g(S_t) \cdot \delta_t$, where the function $g(\cdot)$ represents a transitional relationship or growth function between states S_t and S_{t+1}, and δ_t is a multiplicative process error term. If a discrete-logistic growth model is used for example, the process model can be written as $S_{t+1} = S_t \cdot \left[1 + r \cdot \left(1 - \frac{S_t}{K}\right)\right] \cdot \delta_t$, where *r* is the logistic growth parameter, *K* is the logistic carrying capacity, and the error term may be distributed $\delta_t \sim \text{Lognormal}\left(-\frac{\sigma_p^2}{2}, \sigma_p\right)$. As with the observation error model, we can model the process error term as the ratio of the true state of nature to the predicted state, $\delta_t = \frac{S_t}{g(S_{t-1})}$.

Our catastrophic state-space model extends the basic state-space model by also assuming that each year of a particular time series is an unknown categorical state: it may be a normal year in which the population experiences regular process error variability, or it may be a catastrophic year in which the population experiences the variability associated with catastrophic events. There is no constraint placed on the magnitude of process errors in catastrophic years; the variability in those years may be less than or greater than the expected variability in normal years. Assuming independence between years, the chance of a catastrophe occurring in any particular year can be modeled as a Bernoulli trial with parameter ϕ, where ϕ represents the probability of a catastrophe (catastrophes representing successful trials). Using the laws of conditional probability, we can write this process as $P(S_{t+1} \mid Y_t, \underline{\theta}) = P(S_{t+1} \mid Y_t, \underline{\theta}) \cdot P(Y_t \mid \phi) \cdot P(\phi)$, where $\underline{\theta}$ represents a vector of all model parameters (population parameters, variances, etc.), and Y_t represents a categorical variable (normal or catastrophic year). The vector of categorical states may be treated as nuisance parameters, because they are integrated (or summed) out of the model. The magnitude of a catastrophic event (λ) is modeled as a multiplicative term, so that when the process errors are lognormally distributed,

$$\delta_t \sim \begin{cases} \text{Lognormal}\left(-\frac{\tau_{norm}^2}{2}, \tau_{norm}\right) \text{with P}(1-\phi) \\ \text{Lognormal}\left(\lambda - \frac{\tau_{cat}^2}{2}, \tau_{cat}\right) \text{with P}(\phi) \end{cases}$$

The parameter controlling the magnitude of a catastrophe (λ) represents the average percent reduction in the population size – values close to 1.0 indicate that the population is nearly driven extinct. The probability density function (p.d.f.) of this mixture is described by the equation:

$$P(\delta_t | S_{t-1}, \phi, \lambda, \tau_{cat}, \tau_{norm}) = \frac{\phi}{\sqrt{2\pi} \cdot \tau_{cat} \cdot \delta_t} \cdot \exp\left[-\frac{\left(\ln(\delta_t) - \lambda + \frac{\tau_{cat}^2}{2}\right)^2}{2\tau_{cat}^2}\right] + \tag{2}$$

$$\frac{1.0 - \phi}{\sqrt{2\pi} \cdot \tau_{norm} \cdot \delta_t} \cdot \exp\left[-\frac{\left(\ln(\delta_t) + \frac{\tau_{norm}^2}{2}\right)^2}{2\tau_{norm}^2}\right]$$

A joint prior distribution must be assigned to all unobservable parameters and states in the model. There are two approaches to dealing with the prior distribution on the first state of the system: a prior may be placed on the state S_0, one time period before the first observation (yielding N conditional elements in the prior), or a prior may be placed on S_1, the state in the same time period as the first observation (yielding N - 1 conditional elements in the prior). In the catastrophic state-space model we choose the latter approach, placing a prior on the first state, S_1. Assuming independence between parameters and conditional states, the joint prior for a simple logistic model can be expressed as:

$$\begin{aligned} &P(r,K,\phi,\lambda,\tau_{cat},\tau_{norm},S_1,S_2,\ldots,S_N) = P(r)\cdot P(K)\cdot P(\phi)\cdot P(\sigma_o)\cdot P(\tau_{cat})\cdot P(\tau_{norm})\cdot \\ &P(S_1)\times\prod_{t=2}^{N} P(\delta_t | S_{t-1},r,K,\phi,\lambda,\sigma_o,\tau_{cat},\tau_{norm}) \end{aligned} \tag{3}$$

If data on a particular species consisted of 10 observations over a 10 year period, there would be a total of 17 parameters (10 population states, 2 logistic parameters, 3 variance parameters, 2 catastrophe parameters). Choosing appropriate prior distributions must be done with caution because arbitrary priors that are not informative with respect to one parameter may be informative with respect to another (Walters and Ludwig 1994; Punt and Hilborn 1997; McAllister and Kirkwood 1998). For the catastrophic parameters, we assigned a Uniform (0, 0.5) prior to the probability of a catastrophe (ϕ) so that catastrophes would be constrained to be rare events. A Uniform (0, 1) prior was placed on the magnitude of a catastrophe (λ) to express our vague knowledge about how much catastrophes reduce populations. This prior was chosen over a (0.5, 1) prior to prevent all of the posterior density from being centered near the lower bound (0.5). For the purpose of this analysis, we assigned an uninformative Uniform (0, 1) prior to the growth rate (r), and an informative lognormal prior to carrying capacity (K). Because we are assuming errors in the model to be lognormally distributed, the standard deviation of the error in log-space can be interpreted as the coefficient of variation (CV) in normal-space. We assumed that all three CVs were assigned inverse gamma prior distributions because these priors are a common choice in Bayesian linear models (Gelman et al. 1995). The likelihood function of the data given the parameters is derived from (1):

$$L(r,K,\phi,\lambda,\tau_{cat},\tau_{norm},S_1,S_2,\ldots \mid \hat{N}_2,\hat{N}_1,\ldots) = \prod_{t=1}^{N} P(\varepsilon_t \mid S_t) = \left(\frac{1}{\sqrt{2\pi}\sigma_{obs}\varepsilon_t}\right)^N \exp\left[-\frac{1}{2\sigma_{obs}^2}\sum_{t=1}^{N}(\varepsilon_t - 1.0)^2\right] \tag{4}$$

Using Bayes' theorem, the joint posterior distribution can be calculated as the product of the joint prior distribution (3) and the likelihood (4).

Several methods are available to generate samples from the posterior. Thomas et al. (2005) implemented a sequential-importance-sampling (SIS) approach, which appears to give similar results to Markov Chain Monte Carlo (MCMC) based approaches, but requires some degree of adaptive tuning. Metropolis sampling methods have also been applied to state-space models, however Gibbs sampling has been the most popular in the form of WinBugs (Meyer and Millar 1999a; Spiegelhalter et al. 2003). We chose to implement the catastrophic state-space model in WinBugs because the batch-mode feature of WinBugs 1.4 enabled us to do large-scale simulations on multiple machines.

Relative to population models than only consider observation error or process error, state-space models are generally data-hungry. Adding a mixture distribution to the basic state-space model introduces a further layer of complexity. When the catastrophic state-space model is fit to a single time-series of data, one of the two process error variances generally converges to zero, because the model is unable to separate out the components of the mixture. Several hierarchical models may be developed to estimate the components of the mixture distribution. One approach might be to consider multiple observations of a single population that overlap in time. For example, relative indices of abundance (e.g. catch-per-unit-effort) might be used alongside absolute survey indices as separate time series. In addition to assuming shared process errors between surveys and CPUE series, this type of model would also make the assumption that the catastrophe parameters were shared between the two indices. A second approach is to combine multiple time series from multiple populations. In a hierarchical setting, this type of model would allow each population to have unique population parameters (e.g. growth rates, carrying capacities), with the only assumption being that process error variances are shared between time series. This model could be made more complicated by combining time series from multiple species or multiple environments, each with different probabilities of catastrophes. We focused our simulations on the latter of the two hierarchical approaches because of the flexibility, and the potential application to many species (rather than just those for which CPUE exist).

We generated 10,000 random datasets to test the catastrophic state-space model, where each dataset consisted of a collection of 2-5 time series, representing data from similar species. We assumed that these populations experienced the same observation and process errors, and were affected by catastrophes similarly (e.g. the probability and magnitude of a catastrophe is assumed constant across time series in a dataset). Each of the time series within a dataset consisted of 5-25 simulated observations. We chose to initially consider only the exponential growth and theta-logistic models as simulation models, but each time series was allowed to have unique population parameters (initial population sizes, growth rates, carrying capacities). Observation error and process error was allowed to be either lognormal or gamma, and was assumed to be constant within a dataset (the 2-5 time series all assigned the same type of error). The probability and magnitude of a catastrophes was independent between years and time series (ϕ ranged from 0.01 to 0.2, λ ranged from 0.6 to 0.9). Instead of drawing random values of the observation and process error standard deviations for each dataset, we considered each of the standard deviations to be a random factors (each ranging from 0.0 to 0.6 in steps of 0.05). For some of the time series, we generated large negative process error events in the absence of catastrophes to evaluate the ability of the model to distinguish between these two types of decline. Although they are unlikely to occur, these large negative events represent the tail area of a single process error distribution (Figure 4.1). After all parameters for each dataset were estimated, data points were randomly deleted and time series removed to examine the effect of missing data on the precision of those estimates.

To monitor MCMC convergence, we implemented the CODA package in R. In addition to monitoring the trace plots, autocorrelation plots, and densities of each parameter, the Geweke and Heidelberger-Welch statistic were used to determine whether or not chains had converged (Gelman et al. 1995; Gilks et al. 1996; Carlin and Louis 2000). Following a burn-in of 20,000 elements, we conducted 30,000 sampling iterations, saving every 10th draw to create a posterior sample of 3000 elements of each parameter per dataset. After parameters for each simulated dataset were estimated, the bias was calculated for each parameter as $\hat{\theta} - \theta$ (the true value of the parameter subtracted from the parameter estimate from the posterior distribution).

Results

Of all parameters in the catastrophic state-space model, we found the population parameters to be estimated robustly, but parameter estimates were sensitive to the initial state in the simulation model. This result is expected, as the same result holds under simpler error models (e.g. models that consider observation or process error only). For the case where the operating model is logistic, a time series with low relative population size provides information about the initial state and population growth rate, but little or no information about carrying capacity. The opposite is also true – when the initial population state is close to carrying capacity, estimates of the initial state and growth rate are biased, however estimates of K are relatively unbiased. Bias in estimating the probability of a catastrophe (ϕ) was slightly negative, indicating that the parameter was underestimated more frequently than it was overestimated (Figure 4.2). The opposite result was true for estimating the magnitude of a catastrophe (λ) – the bias tended to be positive, indicating that the parameter tended to be overestimated. Bias in the observation error is centered close to zero, indicating that this parameter and process error in normal years is centered close to zero, however bias in estimating the process error in catastrophic years appears to be positive (Figure 4.2).

One necessary feature of the catastrophic state-space model is that it must be able to distinguish catastrophes from anomalies caused by negative process errors. Because we know each catastrophe in the simulated time series, it is possible to quantify the both the ability of the model to detect catastrophes, and the error rate with which the model classifies normal variation as catastrophes. Statistical power has been applied to many situations in biology (Taylor and Gerrodette 1993), and for the purpose of this analysis we define power as the ability to detect a catastrophe when it actually occurs. The parameters that affect the power to detect catastrophes in our state-space model are the magnitude of catastrophes, and the magnitude of the observation and process errors. As catastrophes become more severe, the probability of detecting them increases (Figure 4.3). Observation error appears to have a larger effect on power than process error (Figure 4.4), however both influence the rate of falsely detecting a catastrophe (Figure 4.5). At intermediate values of the observation error (CV > 0.35), the probability of detecting a catastrophe that reduces the population by 60% becomes less than would be expected by chance alone. As expected, the power to detect catastrophes is highest under low levels of observation and process error, and when catastrophes are large in magnitude.

A final sensitivity of the model that was examined was the ability to detect catastrophes and estimate population parameters in the presence of error misspecification. Ideally, the form of the error would be known (e.g. lognormal, gamma), however this assumption is often unrealistic. After applying a model with lognormal errors to data generated from a gamma distribution, and vice versa, we calculated the median bias for each parameter. In all cases, the gamma model applied to lognormal data resulted in less biased parameter estimates, when compared to the alternative scenario. For example, error misspecification results in a 50% increase in the bias of estimated probability of a catastrophe (Figure 4.6).

Discussion

The catastrophic state-space model could be applied to a wide range of species, including both terrestrial and marine organisms. One potential source for combining continuous time series in a meta-analysis is the Global Population Dynamics Database (NERC 1999), which has collected 5000+ time series from the ecological and fisheries literature. Applications of the catastrophic state-space model should not be limited to parameter estimation or hypothesis testing – incorporating a more realistic form of catastrophic events will have dramatic implications for population projections and management decisions. The most useful case of the model is the scenario where data from similar populations (or even meta-populations) are combined in a hierarchical model.

The datasets in our simulations included a range of realistic observation and process errors, with CVs ranging from 0.01 to 0.5. Across all simulations, the population parameters (e.g. logistic parameters), observation error and process error in normal years were estimated robustly. The ability of the model to estimate the catastrophic parameters (ϕ, λ, τ_{cat}) was highly dependent on the frequency and magnitude of catastrophes involved. The relationship between the magnitude of the catastrophic parameters and the bias in estimating them is somewhat intuitive – as catastrophes become more frequent and stronger, the model is better able to discriminate them from normal years with extreme process error, thereby decreasing the bias. Regarding missing data, we found that removing 50% of observations has less influence on bias than when the frequency and magnitude of catastrophes was changed. A state-space catastrophic mixture model with gamma process errors and gamma observation error was found to be less sensitive to error misspecification, indicating that it would be a more robust choice as the default model, compared to a model with lognormal errors. This result is in agreement with previous studies, indicating that the gamma distribution is less sensitive to outliers than the lognormal distribution (Myers and Pepin 1990).

One potential criticism of this model is that for some datasets (particularly those in which no real catastrophes exist), the magnitude of a catastrophe may be estimated as being approximately 1.0 and the probability of a catastrophe may be non-zero. Our mixture model could still be applied to such data, but results would have to be interpreted differently. Instead of interpreting the process error mixture as the error for normal and catastrophic years, the mixture might be interpreted as a measure of data quality. In this setting, some small fraction of the data would be treated by the model as outliers, and those data points would be assigned to the component of the mixture associated with higher larger variance (Chen and Fournier 1999). For each dataset, researchers should consider the relationship between catastrophic parameters, magnitudes of observation and process errors on power. A second potential criticism of the model is that the exponential or logistic population dynamics models in this analysis are too simple to represent reality. For many species, data that would allow the use of age-structured models has only been collected recently (Holmes and Fagan 2002). A potential future application of the catastrophic state-space model would be to apply it to age-structured data, with catastrophes only affecting one particular age class (e.g. juvenile survival).

The catastrophic state-space model is not limited by the logistic function – if population datasets were combined in a meta-analysis, it would be valid to allow each population to have a unique growth function. For example, it might be reasonable to assume that a smaller population would exhibit exponential growth, while a larger (and possibly more stable) population could be modeled using a density-dependent function. To validate the choice of growth models, it would also be possible to apply Bayesian model selection criteria, such as Bayes factors or DIC, to evaluate which model is most appropriate for each time series. Future analyses will also demonstrate the model selection procedure for distinguishing between catastrophic state-space model proposed here, and the simpler alternative with a single process error.

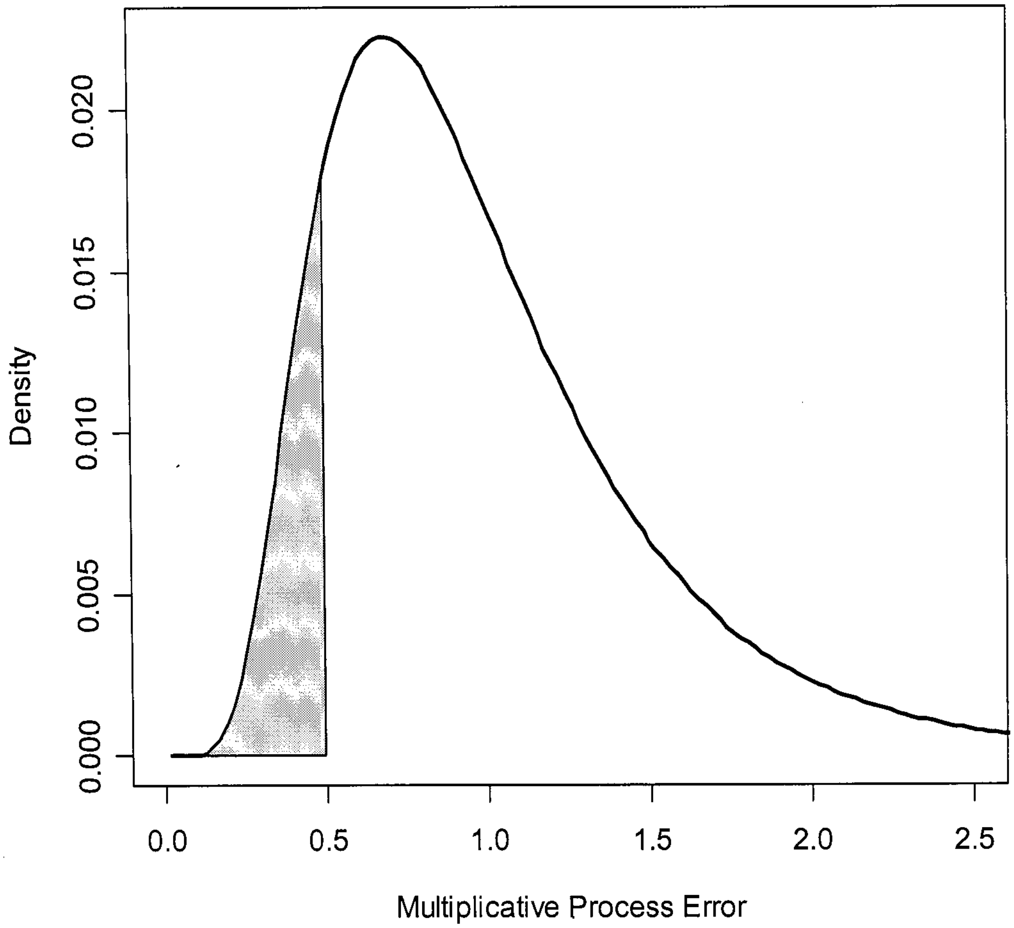

Figure 4.1 Lognormal density with an expected value of 1.0 and CV = 0.5. The region of the curve less than 0.5 has been shaded to illustrate those process errors that might be considered catastrophes (in this region the population is reduced by more than 50%).

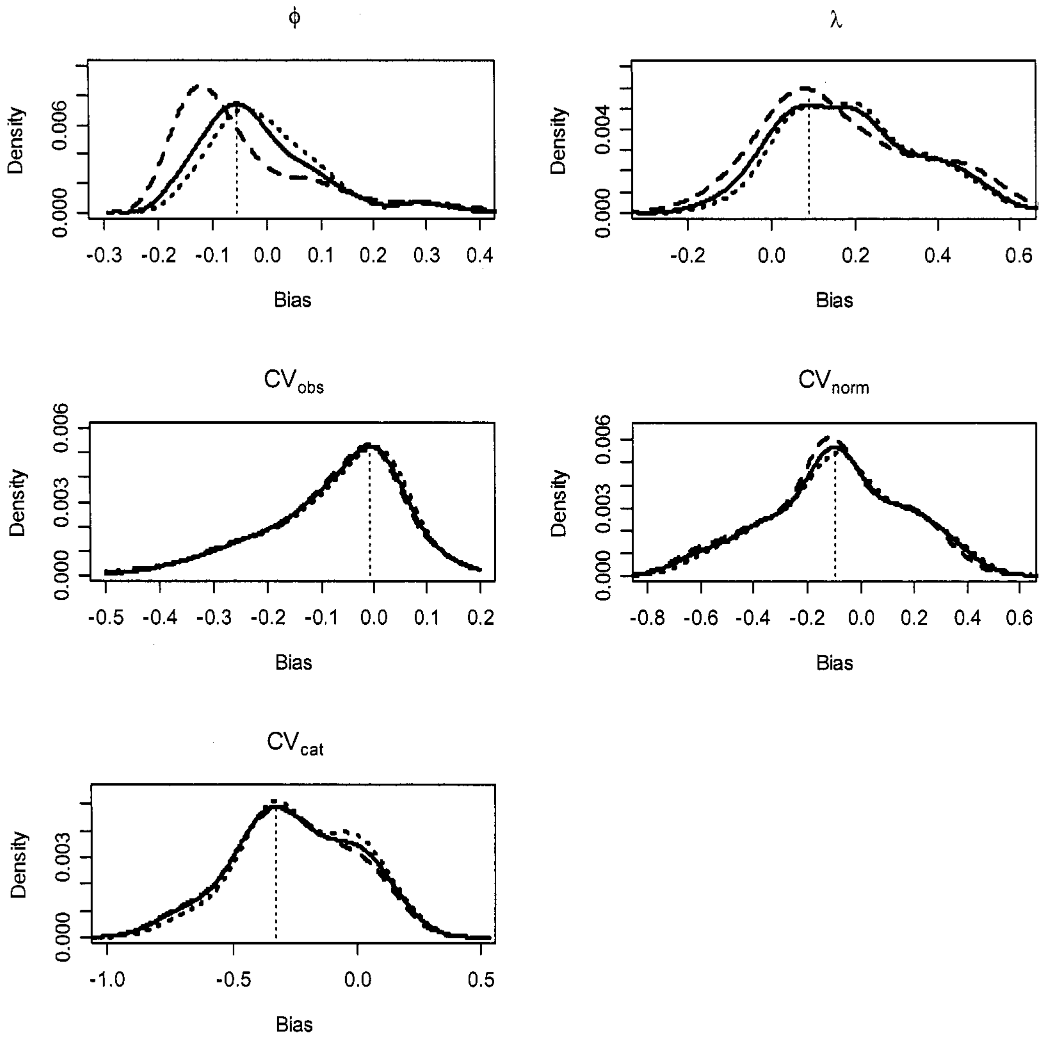

Figure 4.2 Density plots of the bias in posterior estimates of the probability of a catastrophe (ϕ), the magnitude of a catastrophe (λ), the observation error CV (σ_o), process error CV in normal years (τ_{norm}), and process error CV in catastrophic years (τ_{cat}). The solid line represents the distribution of the median bias (posterior median minus true value), the dotted line represents the bias of the posterior mode, and the dashed line represents the bias of the posterior mean.

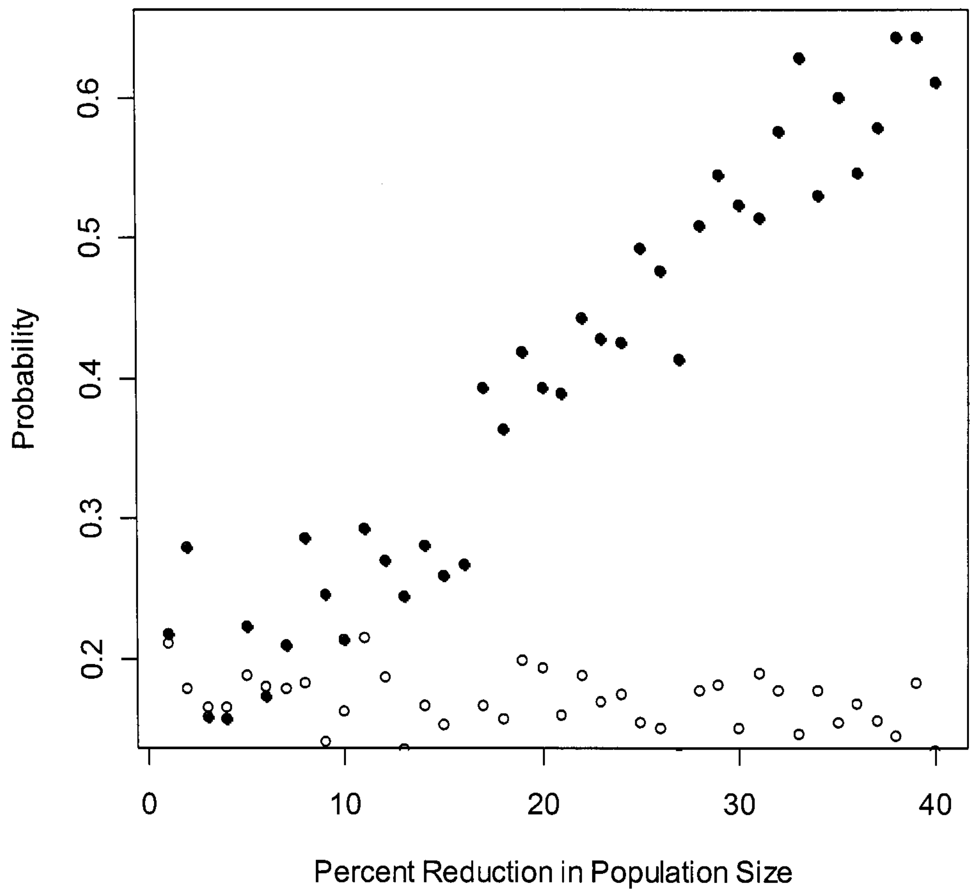

Figure 4.3 The relationship between the power to detect a catastrophe (solid circles) and the magnitude of the catastrophe. Also included is the probability of falsely detecting a catastrophe (open circles). Power is calculated across all levels of observation error and process errors.

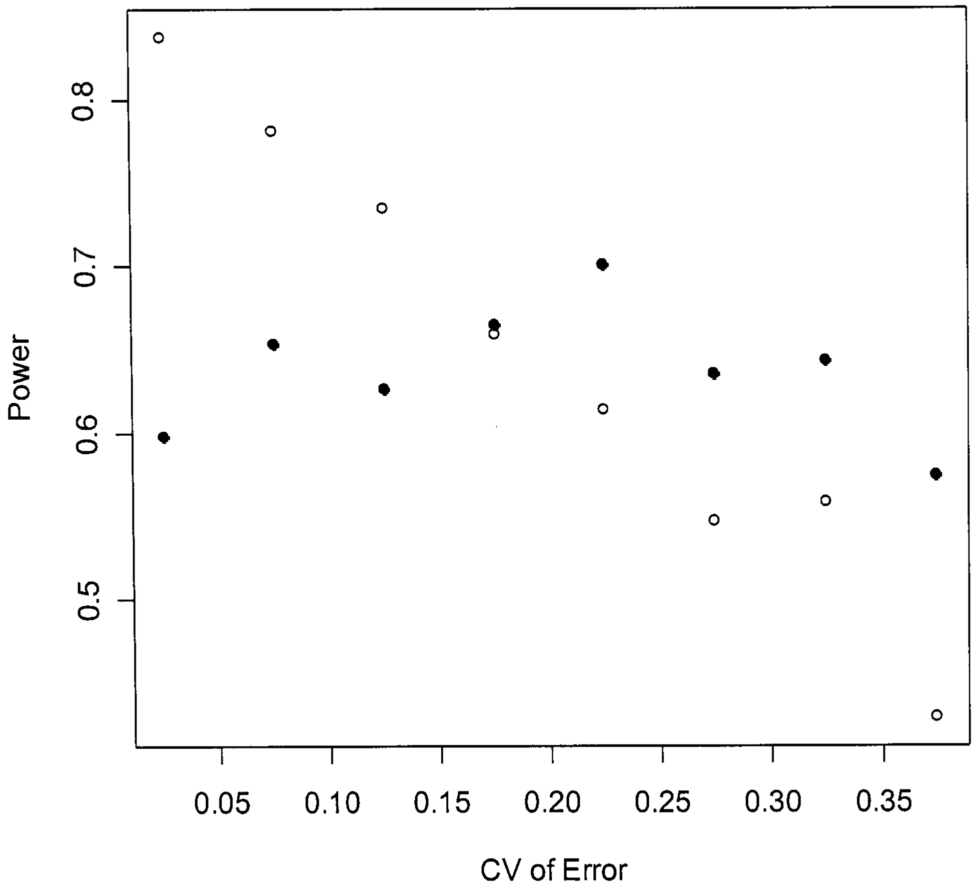

Figure 4.4 The relationship between the CV of the errors (when the errors are lognormally distributed) and the power to detect a catastrophe for 1000 datasets. The magnitude of a catastrophe is held constant at 0.6 across all datasets. Open circles represent varying levels of observation error (across all levels of process error), and solid circles represent varying levels of process error (across all levels of observation error).

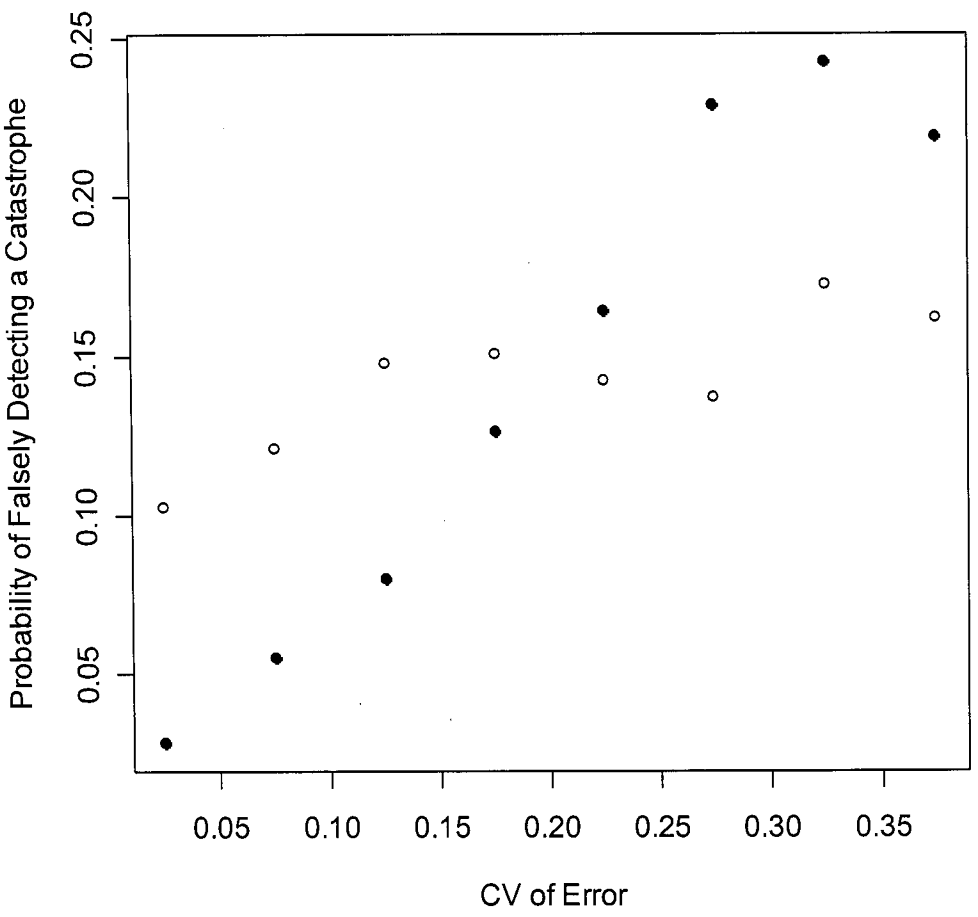

Figure 4.5 Relationship between the CV of the errors (when the errors are lognormally distributed) and the probability of falsely detecting a catastrophe for 1000 datasets. The magnitude of a catastrophe is held constant at 0.6 across all datasets. Open circles represent varying levels of observation error (across all levels of process error), and solid circles represent varying levels of process error (across all levels of observation error).

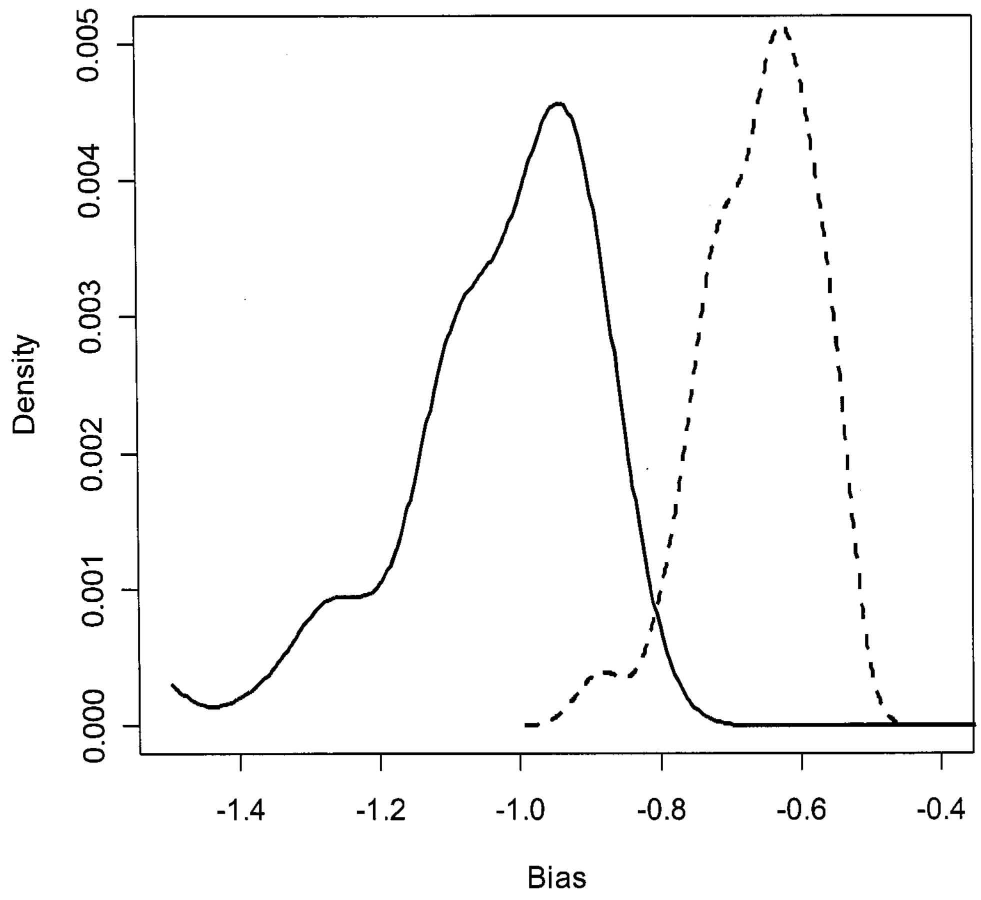

Figure 4.6 The effect of model misspecification on the bias in estimating the probability of a catastrophe, ϕ. Bias is defined as the posterior median minus the true probability of a catastrophe. The solid line represents an estimating model with lognormal error applied to an simulation model with gamma errors, and the dashed line represents an estimating model with gamma error applied to an simulation model with lognormal error. A value of zero indicates no bias.

Chapter 5: Do Catastrophes Affect Northern Fur Seal Pup Production?

Introduction

Catastrophic events are one of the most important features of population dynamics (Lande 1993; Mangel and Tier 1994), yet they have been incorporated into few population viability analyses or stock assessments. Catastrophes may be included in population projections to provide realistic scenarios for managers, but for the frequency and magnitude of catastrophic events to be estimated reliably, a reasonable number of catastrophes must be observed in historical data. There are few individual data sets that are relatively continuous and enable estimation of both the frequency and magnitude of catastrophic events. An alternative approach, as suggested by Ward and Hilborn (2006), is to combine multiple time series from similar species or similar environments in a hierarchical modeling framework. In this setting, parameters may be shared between populations, including the process error parameters, and possibly the frequency and magnitude of catastrophes. There are several advantages of the Bayesian approach over restricted maximum likelihood methods (e.g. Staples et al. 2004). Most importantly, Bayesian model output and parameter uncertainty may be expressed in terms of probabilities, which is compatible with decision making (Drechsler and Burgman 2004).

In a review of otariid population dynamics, Gerber and Hilborn (2001) examined the catastrophic events in populations dispersed around the globe and estimated the total probability of catastrophes in otariid species to be 0.25-2.1% per year. For many of the otariid species in decline, the only type of data that exists is records of pup counts. The number of pups in a given year may be estimated with precision, because all pups usually remain on land, concentrated in small spatial locations. Historically, pup counts have been used as indices of abundance for some otariid species, including the northern fur seal *Callorhinus ursinus* (Loughlin et al. 1994). Northern fur seal pup abundance data has been collected by the National Marine Fisheries Service (NMFS) in four locations: the Pribilof Islands (St. George, St. Paul), Bogoslof Island, and San Miguel Island. Surveys on the Pribilof Islands have been conducted since the 1950s, while counts on Bogoslof and San Miguel have been initiated more recently (1980 and 1968 respectively, when the rookeries were discovered). The *C. ursinus* populations that breed on the Pribilof Islands constitute approximately 90% of the U.S. population and 60% of the global population (York et al. 2000). The remaining 40% of the global population is found in the western Pacific (Kuril Islands, Tyuleniy Island, Commander Islands), however annual surveys do not exist for these populations. Data from NMFS surveys since 1960 (annual counts 1960-1990, biennial counts after 1990) indicate that pup production on the St. Paul and St. George rookeries has been declining (Towell et al. 2006; Figure 5.1), and the cause for the decline remains unknown. Many of the Pribilof pup estimates have associated standard error estimates, which are computed based on the traditional Lincoln-Peterson mark-recapture estimator (Seber 1982). The coefficients of variation (CVs) based on these standard errors are generally quite small because of large sample sizes (CVs < 0.10), and may underestimate the true measurement error of the NMFS surveys. Surveys of *C. ursinus* pups on Bogoslof Island were initiated by NMFS when the rookery was established in 1980. The Bogoslof surveys have not been conducted annually, but show an increasing exponential trend in production (Figure 5.1).

Production for the San Miguel stock also has shown an increasing trend, but has experienced several catastrophic events. These negative events may be partially caused by El Niño (York 1991), but the effects are not completely understood. While the two largest declines in San Miguel pup production occurred during strong El Niño years, the *C. ursinus* population on San Miguel has experienced several other El Niño years (such as the ones that occurred over the period 1990-1994), with no apparent decline in pup numbers. While inter-annual fluctuations in the number of births is expected, pup production of both stocks may also be impacted by occasional catastrophic events.

Incorporating catastrophes into population dynamics models is essential for species such as *C. ursinus* that are at population sizes less than carrying capacity (this population is currently considered depleted under the Marine Mammal Protection Act). While several definitions have been proposed for the term 'catastrophe', we choose to adopt the definition that has been generally accepted in the ecological literature (Gerber and Hilborn 2001; Ward and Hilborn 2006): a catastrophe represents a single unpredictable, infrequent negative event that affects a component of the population (e.g. a particular age class), or the entire population as a whole. A catastrophic event should be beyond the range of normal year-to-year variability. Catastrophic reductions in the number of *C. ursinus* pups may be an indicator that pup survival is impacted by catastrophes, but may also be an indicator of changes in the population dynamics of mature adults. If a catastrophe such as a disease outbreak were to reduce adult survival (fewer adults results in less production) or adult fecundity for instance, the catastrophe would be reflected by a drop in the number of births. Previous analyses dealing with catastrophes have placed reasonable limits on how catastrophes are defined, constraining them to reduce a population by more 50% (Erb and Boyce 1999; Gerber and Hilborn 2001). We adopt a similar convention in this manuscript to avoid assigning the term 'catastrophe' to small population declines.

The primary objective of this paper is to determine whether the dynamics of northern fur seals are best characterized by a model that includes infrequent catastrophes, or whether a simpler population model is more appropriate. Catastrophic events for all four of the U.S. populations may be related to changes in sea surface temperature, changes in food web structure (Trites et al. 1999), or indirect fishing pressure (for a Bering Sea Steller sea lion example see Fritz and Brown 2005). While it is impossible to determine which of these factors may be contributing to the population decline, our approach is to determine the potential for these collective factors to pose a risk to northern fur seals. In particular, we seek to determine which population models are best supported by available data, and whether models that include additional observation uncertainty are appropriate.

Methods

The catastrophic state-space approach is an extension of the basic state-space model, allowing process error to be modeled as a mixture distribution. The two components of the mixture are normal inter-annual variability and infrequent negative events. Observation error is assumed to be a multiplicative error term, $\hat{N}_t = S_t \cdot \varepsilon_t$, where $\hat{N}_t$ represents the estimated abundance (or number of pups) at time *t*, S_t represents the true number of pups at time *t*, and ε_t represents an error term with an expected value of 1.0. The basic process error equation is $S_t = g(S_{t-1}) \cdot \omega_{t-1}$, where the function g(·) represents the transition between states, and ω_{t-1} represents the process error. To incorporate catastrophes into the model, we assume that each population in each year experiences catastrophes independently, such that $Y_{i,j}$~Bernoulli(ϕ), where the parameter ϕ represents the probability of a catastrophe. The average magnitude of catastrophe is represented by θ, so that a catastrophe that reduces a population to 70% of its original size would have $\theta = 0.3$. If the process error distribution is assumed to follow some density f(·), the mixture of the process errors in catastrophic and normal years can be written as:

$$\omega_t \sim \begin{cases} \mathrm{f}(a_1, b_1) \text{ with } \mathrm{P}(1-\phi), \mathrm{E}[\omega_t] = 1.0 \\ \\ \mathrm{f}(a_2, b_2) \text{ with } \mathrm{P}(\phi), \mathrm{E}[\omega_t] = 1-\theta \end{cases}$$

where *a* and *b* represent some parameters controlling the shape of the density function. In our application of the catastrophic model to northern fur seal pup production, we chose to use the gamma distribution for both the observation error and process error mixture because it has been shown to be more robust to model misspecification than the lognormal distribution in models with one type of error present (Myers and Pepin 1990; Cadigan and Myers 2001).

In developing a group of candidate models for this analysis, we had to consider population-specific parameters and ecosystem-specific parameters. Each of the four populations was allowed to have an island-specific growth equation. Pup counts at Bogoslof and San Miguel appeared to be increasing exponentially, and were represented by the 2-parameter model: $S_{t,j} = S_{0,j} \cdot \exp(t \cdot r_j)$, where r_j is the rate of change for population j. Counts at St. George and St. Paul appear to show the opposite trend (Figure 5.1). Two alternative models were considered for the St. George and St. Paul populations: a 2-parameter exponential model (similar to the one used for Bogoslof and San Miguel), and a 3-parameter exponential decay model that was used in previous population assessments (York et al. 2002). This latter equation can be represented as $S_{t,j} = a_j + b_j \cdot 2^{(1960-t)/c_j}$, where a_j is a parameter representing equilibrium pup level (asymptote) for population j, b_j is a parameter representing the total decrease in annual pup production between 1960 and the equilibrium level a_j, and c_j is a decay term representing the half-life for population j. Growth rates in the models (r, b) were assigned uniform prior distributions, while the equilibrium (a) and decay term (c) were assigned gamma prior distributions. The equilibrium parameter represents the asymptotic production for each population, and the decay parameter represents the time necessary for the number of pups to decrease to half the difference between the number of pups at time zero and the asymptotic number of pups.

We assumed that the two catastrophe parameters in the model were ecosystem-specific, so that ϕ_j and θ_j represented the probability and magnitude of catastrophes affecting the j^th^ population. The primary justification in not allowing the catastrophe parameters to be shared across all populations is that the San Miguel Island population experiences El Niño events, while the Bering Sea populations do not. We were forced to assume that the process error variances (not the process errors themselves) were shared across all four populations, because the San Miguel time series alone would not be able to separate the components of the process error mixture for a model that allows ecosystem-specific variances. For each ecosystem, we assigned Uniform (0, 1) priors to the probability of a catastrophe, and a Uniform (0, 1) prior to the magnitude of catastrophic events. This latter prior was chosen to include all negative events, however only those in the range (0.5-1.0) will be interpreted as catastrophes.

Process error terms in all candidate models were chosen to be multiplicative gamma distributions. There are several possible parameterizations of the gamma distribution, however we chose to adopt a two-parameter version dependent on a shape parameter and CV parameter. To evaluate whether there was evidence for the catastrophic state-space model over the standard state-space model, we considered some models with a single process error distribution. All process error CVs were assigned gamma prior distributions with an expected value of 0.2.

To assess whether a model that incorporated additional observation uncertainty was appropriate, we developed several models that included additional observation error parameters. Because not all years had standard error estimates present, we were forced to assume that the CV for years with missing estimates was constant (CV_c). In the first set of models ('1 CV', Table 1), we assumed that the CV for the ith year and jth population could be modeled as $CV_{i,j} = \sqrt{CV_{est(i,j)}^2 + CV_{extra}^2}$ if the standard error estimate did exist ($CV_{est(i,j)}$ being the CV calculated from the abundance and standard error estimates), and $CV_{i,j} = \sqrt{CV_c^2 + CV_{extra}^2}$ otherwise. One problem with assuming that the CV is constant when standard errors don't exist is that surveys conducted on the Pribilof Islands may have a different CV than surveys conducted on Bogoslof Island or San Miguel Island.

The justification for this difference is that pup estimates from the Pribilofs represent mark-recapture estimates, while Bogoslof and San Miguel counts are direct counts based on single observations. To account for this additional variation, we developed a second set of models ('2 CV', Table 1) that allowed the observation error CV for Bogoslof and San Miguel to differ from the Pribilof surveys (but which was assumed constant between years and the two islands). Finally, we developed a third set of models ('3 CV', Table 1) that allowed the Bogoslof and San Miguel populations to have unique, island-specific observation error CVs. All observation error CVs were assigned gamma prior distributions with a mean of 0.2, with the exception of the additive parameter (CV_{extra}), which was assigned a uniform prior distribution.

All combinations of population models, process errors, and observation error scenarios yielded a total of 48 candidate models. Parameter estimation for each model was done via Markov Chain Monte Carlo (MCMC) in WinBugs 1.4. After a burn-in of 2.0e05 iterations, we thinned the next 5.0e05 elements of the MCMC chain, retaining every 100th sample. After the sampling was completed, we used the CODA package in R to determine whether the chain had converged and was stationary (Best et al. 1995). In addition to visually examining trace and autocorrelation plots, the Geweke and Heidelberger-Welch statistics were computed for each model to determine chain convergence (Geweke 1992; Heidelberger and Welch 1983). The Deviance Information Criterion (DIC) was also computed as a model selection tool (Speigelhalter et al. 2003). DIC is conceptually similar to the widely used Akaike's Information Criterion (AIC, Akaike 1973), with the major difference being that DIC integrates over the parameter space (instead of maximizing). The DIC computation can be expressed as DIC = $D(\hat{\theta}) + 2p_D$, where $D(\hat{\theta})$ represents the deviance evaluated at the posterior mean, and p_D represents the effective number of model parameters. The term p_D can also be expressed as the deviance function evaluated at the expected posterior values of the parameters subtracted from the mean deviance across all parameter vectors, $p_D = \overline{D(\theta)} - D(\hat{\theta})$.

Results

With all combinations of production models (2-parameter or 3-parameter exponential models for the Pribilofs) and observation error structures, our initial candidate list included 48 models. We omitted 12 models that allowed St. Paul pup production to follow a 3-parameter exponential curve and St. George to follow a 2-parameter exponential curve because the models resulted in poor fits to the data. Of the remaining 36 models, the model with the lowest DIC score appeared to be a model that included catastrophes, allowed both Pribilof populations to follow 3-parameter exponential curves, and allowed both Bogoslof Island and San Miguel Island to have unique island-specific observation errors (Table 1). This model was strongly favored over the alternative models, receiving 97.5% of the normalized DIC weight.

One of the most important results from the model selection procedure was that the state-space model that incorporated catastrophes was always favored over the simpler model with a single process error distribution. This result is important because it indicates that catastrophes should be included in our analysis, regardless of which population models are chosen. A second important result is that for nearly all scenarios, there appears to be no evidence for the inclusion of an additional observation error parameter, suggesting that the estimated CVs for the Pribilof surveys are appropriate. A third result is that the exponential decay models appear to be more appropriate for both the St. Paul and St. George populations. This result is consistent with the previous stock assessment of northern fur seals (York et al. 2002).

For the model with the lowest DIC score, the posterior mode of the probability of a catastrophe (ϕ) was found to be 1.9% for Bering Sea populations (2.5% = 0.001, 97.5% = 0.462) and 7.2% for the San Miguel population (2.5% = 0.021, 97.5% = 0.228). The posterior mode of the magnitude (θ) of catastrophic events for San Miguel population was estimated to be 0.75 (2.5% = 0.591, 97.5% = 0.801). The magnitude of catastrophic events in the Bering Sea was estimated to be close to zero, indicating that while the Bering Sea populations might experience chance reductions in the number of pups, the populations do not experience true catastrophes. There several important differences between the estimated process error CVs in normal and catastrophic years (Figure 5.2). First, the posterior mode of the CV in catastrophic years (0.129) is larger than the mode of the CV in normal years (0.109), implying that there is slightly more process uncertainty in catastrophic years. A second result is that the variance of the estimated process error CV in catastrophic years is greater than the variance in normal years. For example, approximately 17% of the estimated process error CV in catastrophic years is greater than the upper bound of the process error CV in normal years (~ 0.16). This result is somewhat intuitive – because the number of catastrophic years is relatively small compared to the number of normal years, the process error variation in catastrophic years is expected to be estimated with less precision.

Discussion

In this paper we demonstrate the application of a novel technique to estimate the probability of catastrophes. Estimation of catastrophes is important in both basic analyses of population dynamics and in applied analyses of extinction risk. Our analyses demonstrate the potential application of our approach in a meta-analysis, and similar studies should be conducted for taxa for which data are available for multiple populations. Our catastrophe model is able to incorporate catastrophes that affect total population size, or specific components of the population (as in the fur seal example presented here). Future applications of our model should also explicitly consider identifying data requirements to detect catastrophes of a given magnitude.

One potential criticism of the approach used here to incorporate catastrophes into time series is that the state-space framework is overly complicated, and the frequency of catastrophes could be estimated with a simple back of the envelope calculation (dividing the number of years with obvious catastrophes by the total number of years). If such an approach were used for the four populations in this analysis, the estimated probability of a catastrophe for northern fur seal pup production would be estimated as $\hat{\phi} = 0.015$ (or 2 obvious catastrophes in the San Miguel time series divided by 135 total years across all four time series). The estimated probability of a catastrophe using the Bayesian approach is higher for both the San Miguel and Bering Sea fur seal populations (1.9% and 7.2%, respectively; Figure 5.3). One potential reason for this is that the Bayesian approach used in this analysis considers uncertainty in whether each year of each time series experienced a catastrophe or not. A similar back of the envelope calculation could be done to estimate the magnitude of catastrophes on San Miguel, by taking the average change in pup production for the two years with noticeable catastrophes ($\hat{\theta} = 0.705$). This estimate would indicate that on average, catastrophes reduce pup production to 30% of the expected size, but again this estimate is an underestimate of the posterior mode ($\theta = 75\%$).

Although we cannot identify the cause for catastrophic events in the three time series number of northern fur seal pups, it appears that there is strong support for the catastrophic state-space model based on the DIC weights. The catastrophic state-space model that allowed Bogoslof Island surveys and San Miguel surveys to have a separate observation error CVs appeared to be more credible than models that assumed the CV to be constant. There are also large differences between the estimated CVs (Figure 5.2), suggesting that mark-recapture estimates are more precise than direct counts (essentially a minimum estimate). For the catastrophic model, the modes of the posterior distribution indicated that catastrophes occur on San Miguel with a 7.2% probability, and the average magnitude of those catastrophes is a 75% reduction in the number of pups. This estimated probability of a catastrophe is greater than the estimates previously suggested by Gerber and Hilborn (2001) for all otariid species.

A second objective of this analysis was to identify the most reasonable population models for analyzing pup counts, which are a common type of data available for pinnipeds worldwide. We evaluated the performance of the 2-parameter simple exponential and 3-parameter exponential decay model for the Pribilof Island populations. One general result was that the 3-parameter exponential data appeared to fit the data for both the St. Paul and St. George population, when compared to the 2-parameter model (this result is in agreement with the 2000-2001 stock assessment, York et al. 2002). Our parameter estimates for these models are not directly comparable to the last assessment, however, because the previous analysis only considered data since 1973, whereas in the current analysis we considered data collected since 1960.

In applying model selection tools such as DIC, it is important to stress the fact that even though the catastrophic model received the majority of the model weights, none of the models in this analysis should be considered the "true" model. The catastrophic model provides a more conservative estimate for future population projections, however applications of the catastrophic model must also take into account the power to detect catastrophes. As Ward and Hilborn (2006) illustrated, the power to detect catastrophes is strongly affected by observation error, and both the frequency and magnitude of the catastrophic events. Process error also appears to have an effect on power, but increasing the magnitude of the process error also increases the chance of falsely detecting catastrophes. Under the range of errors and catastrophic parameters that affect *C. ursinus* populations and surveys, the power to detect catastrophes is greater than chance alone, but probably in the range of 80-90%. These results underscore the importance of quantitative analyses to determine data needed to detect catastrophes of a given magnitude and to include estimates of probability of catastrophes in viability assessments.

Table 5.1 Delta DIC values for a set of 36 candidate models describing trends in northern fur seal pup production. Models may be represented as basic state-state models, or as state-space mixture models (including catastrophes), and may contain additional observation error in the form of an additive CV parameter. For each model, the state transition equations on Bogoslof Island and San Miguel Island are assumed to be 2-parameter exponential models. Pribilof populations (St. Paul = SP, St. George = SG) may be modeled as 2-parameter or 3-parameter exponential decay models. Scenarios with 1 CV constrain the observation error CVs for all surveys in years without standard error estimates to be constant across the four populations, scenarios with 2 CVs allow the Pribilof surveys to have separate observation error CVs from the Bogoslof and San Miguel surveys, and scenarios with 3 CVs allow Bogoslof and San Miguel to have island-specific observation error CVs (the Pribilof populations are assumed to share the same CV). The best fit model with the lowest DIC score is highlighted in bold.

SP Parameters	SG Parameters	Catastrophes	Extra CV	1 CV	2 CVs	3 CVs
3	3	Y	Y	114.9	67.5	50.1
3	3	N	Y	164.6	102.6	107.0
3	**3**	**Y**	**N**	**48.8**	**15.6**	**0.0**
3	3	N	N	104.6	52.4	52.8
2	2	Y	Y	102.7	93.7	84.8
2	2	N	Y	171.3	134.4	135.0
2	2	Y	N	50.4	27.7	12.4
2	2	N	N	109.5	66.4	66.9
2	3	Y	Y	94.4	68.2	51.1
2	3	N	Y	143.2	108.4	111.4
2	3	Y	N	48.9	20.2	5.0
2	3	N	N	108.1	57.7	70.5

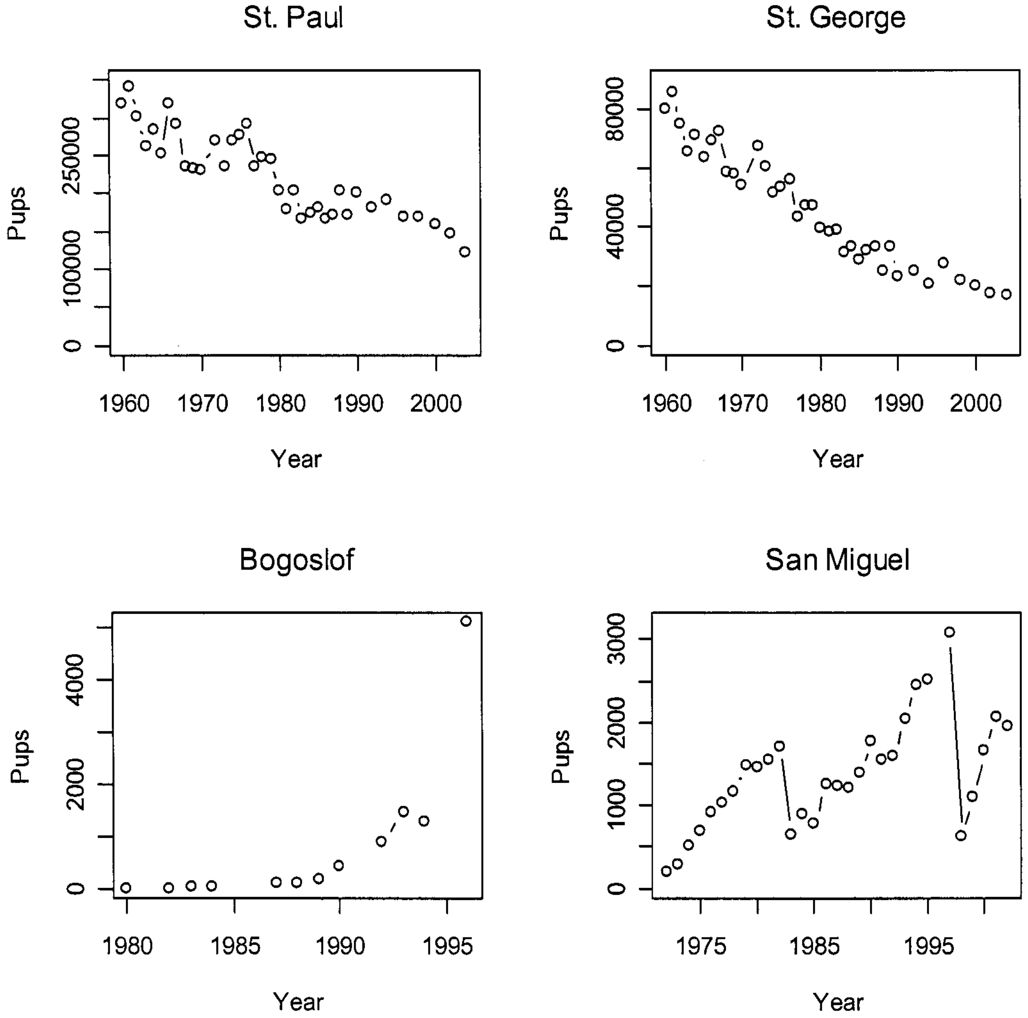

Figure 5.1 Estimates of northern fur seal pup abundance for four U.S. populations. The Pribilof Islands (St. Paul, St. George) and Bogoslof Island are located in the Bering Sea, while San Miguel Island is located off the coast of California.

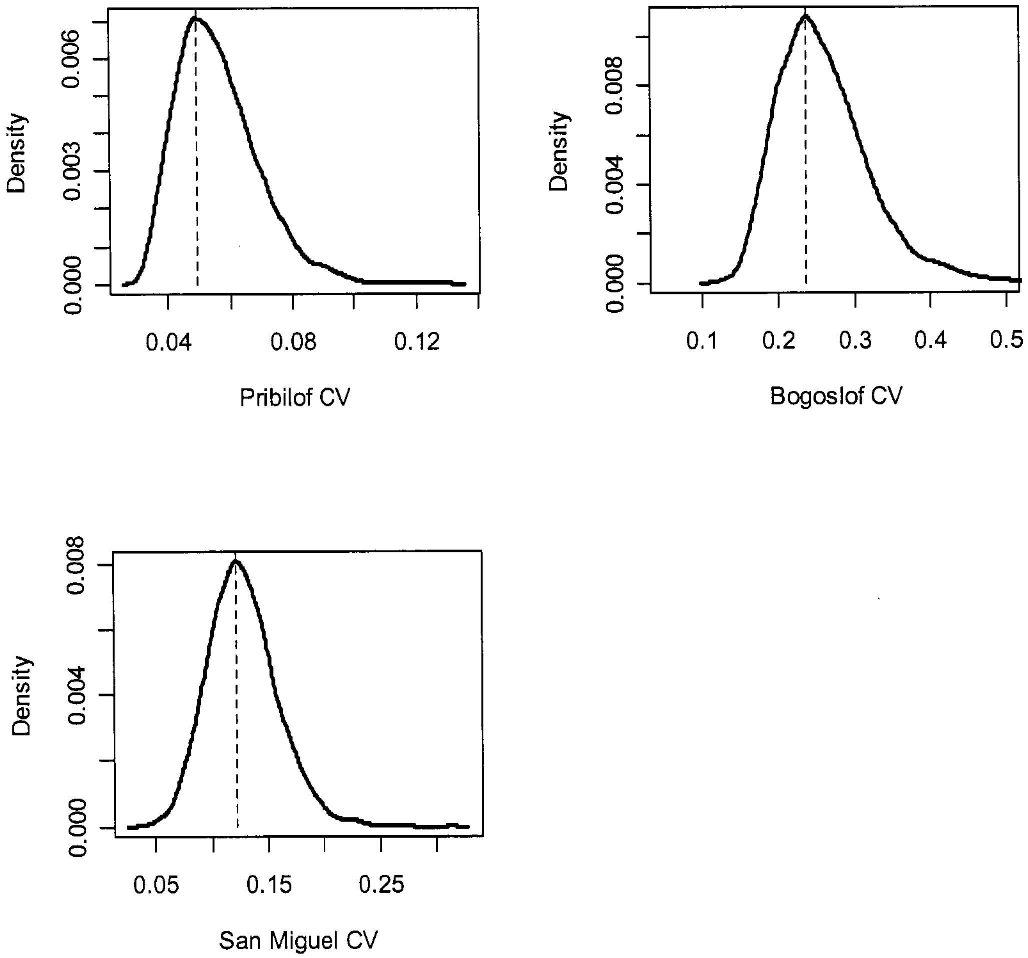

Figure 5.2 Posterior probability distributions of estimated observation error CV parameters for the model with the lowest DIC score, which allows Bogoslof Island and San Miguel Island to each have unique observation error CVs (Table 5.1). The Pribilof Islands have the lowest observation error, while Bogoslof Island appears to have the largest observation error. The dashed vertical line represents the posterior mode of each distribution.

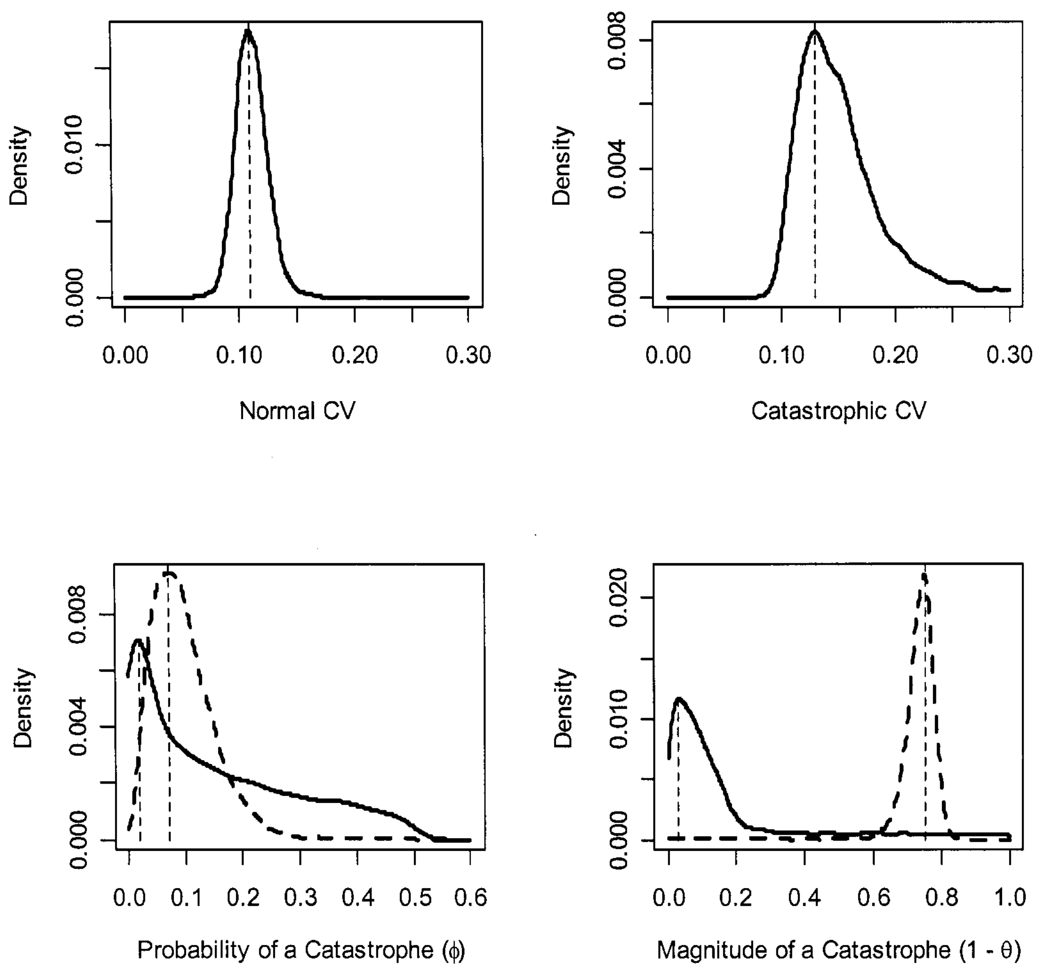

Figure 5.3 Posterior probability distributions for the process error parameters and catastrophe parameters for the model with the lowest DIC score (Table 1). Posteriors of the process error CVs indicate that the variation in catastrophic years (Catastrophic CV) is more skewed than the normal variability (Normal CV). Because we allow the probability of a catastrophe and magnitude of a catastrophe to vary by ecosystem, two posterior distributions are presented (solid line = Bering Sea, dashed line = San Miguel Island). The dashed vertical line represents the posterior mode of each distribution.

Chapter 6: Allee Effects: Really Rare or Just Difficult to Detect?

Introduction

For populations that naturally exist at low density or have been reduced to some fraction of their original numbers, some critical population size may exist below which population growth will become negative, and the population will not persist. These thresholds, described by Allee (1938), are known as Allee effects by ecologists and referred to as depensation by fishery scientists (Liermann and Hilborn 2001). In the last decade, the published literature on Allee effects and depensation has grown rapidly, underscoring the growing importance of this phenomenon in population dynamics modeling. In addition to the obvious risk of extinction for species that naturally occur at low population density, Allee effects may be responsible for the lack of recovery of species that have been harvested to depleted levels (Gascoigne and Lipcius 2004). In addition to understanding the biology behind Allee effects, it is critical to develop techniques to detect possible Allee thresholds before populations are below them.

From a biological perspective, the causes of Allee effects may be divided into four categories, following Liermann and Hilborn (2001): (1) at low density there might be a decreased chance of reproduction (either because of spatial isolation or reproductive asynchrony; Calabrese and Fagan 2004), (2) low density may negatively affect group dynamics (e.g. schooling behavior of fish; Parkinson 1990), (3) organisms may have a reduced ability to modify their environment at low density (e.g. decreased feeding efficiency; Andrewartha and Birch 1954), and (4) predator saturation may act as a depensatory mechanism (Schreiber 2003; Gascoigne and Lipcius 2004a).

Evidence supporting Allee or depensatory behavior is limited (reviewed by Liermann and Hilborn 2001). Many of the recent studies have addressed the evidence of Allee effects in a meta-analysis framework, focusing on vertebrate populations. For example, Fowler and Baker (1991) considered 12 data sets of large mammals, Saether et al. (2002) analyzed 13 species of solitary birds, and Myers et al. (1995) conducted a meta-analysis on 128 fish stocks. Most recently, Sibly et al. (2005) conducted a large scale meta-analysis of 3200 time series divided into four taxonomic groups (Aves, Insecta, Mammalia, Osteichthyes). Depensation has also been implicated in the population dynamics of invertebrate species, including invasive plants (Davis et al. 2004) and marine broadcast spawners (Gascoigne and Lipcius 2004b)

Whether or not Allee dynamics are detected may be a function of the types of data collected, the statistical methods used, and the population models considered. Myers et al. (1995) used likelihood ratio tests to examine whether there was support for depensation in 26 fish stocks, and concluded that there was strong evidence in only 3. Liermann and Hilborn (1997) analyzed the same data in a hierarchical Bayesian meta-analysis, and highlighted the advantages of this technique over traditional hypothesis testing. Most importantly, Bayesian methods were better able to express uncertainty in the strength of depensation, which translated directly into providing better advice to decision makers (see Punt and Hilborn 1997). Sibly et al. (2005) used an approach similar to Myers et al. (1995), and conducted linear and quadratic regressions on the observed per-capita growth rates in 3200 data sets in the Global Population Dynamics Database (GPDD; NERC Center for Population Biology 1999). This method provided little evidence of Allee effects in the GPDD database (0.2% of all records). Several problems with this analysis were that the estimation routine may have been biased (Ross 2006), stochasticity was not included (Doncaster 2006), and that p-values may not be reliable for assessing evidence (Reckhow 1990; Schervish 1996). Recent applications of state-space models to ecological data have demonstrated the benefits of considering both observation and process error (de Valpine and Hastings 2002; Calder et al. 2003; Clark and Bjørnstad 2004). Dennis (2002) found the behavior of a stochastic version of an Allee model to be different than that of a deterministic Allee model, suggesting that including process error in the model will have implications for management.

A variety of model selection tools have emerged to address model uncertainty and model selection in a Bayesian setting. Some model selection methods are designed to explain observed data (Bayes factors; Kass and Raftery 1995), while others have been designed as predictive tools (cross validation; Bernardo and Smith 2002). Some tools such as the Deviance Information Criterion (DIC; Spiegelhalter et al. 2002) may be used to assess the relative weight of one model over competing models, while other approaches have focused on Bayesian model averaging (BMA; Raftery et al. 1997; Hoeting et al. 1999). The DIC criterion has shown promise in several recent ecological analyses (Barry et al. 2003; Ellison 2004; Farnsworth et al. 2006), but has several drawbacks (see discussion in Spiegelhalter et al. 2002). Most importantly, DIC is incompatible with decision making and model averaging. Model weights computed from the DIC criterion may be normalized to sum to 1.0 across models, but should not be confused with posterior model probabilities. A preferred alternative to DIC is to compute the Bayes factor or posterior model probabilities via marginal likelihood computations (Gelfand and Dey 1994; Chib 1995). Although these calculations are often difficult, the posterior model probabilities may be used to conduct formal decision analyses (Wade 2000) and may be used to generate BMA parameter estimates across models.

The purpose of this paper is to address the evidence of Allee dynamics in a Bayesian framework. The performance of three single-species population models will be compared, including two models with density dependence (a logistic model, and an extension of the logistic model that allows for a decline in population growth at low density). To address whether Allee effects are rare in natural populations, I will conduct an analysis of 1900 time series in the GPDD (including a subset of those used by Sibly et al. 2005). Each data set will be examined independently in a Bayesian analysis, and the evidence for or against Allee effects will be measured in terms of Bayesian posterior model probabilities. These posterior model probabilities will be used to generate model averaged Bayesian parameter estimates. Differences in the evidence for Allee effects and the strength of density dependence will be examined between three taxa.

Methods

I chose to use two density dependent population models in this analysis. To model populations not affected by Allee dynamics, I used the familiar theta-logistic population model (Gilpin and Ayala 1973), $N_{t+1} = N_t + r \cdot N_t\left(1 - \left(\frac{N_t}{K}\right)^{\theta}\right)$, where N_t represents the population size at time t, K represents the carrying capacity, r represents the maximum per capita growth rate, and the parameter θ controls the shape of density dependence. A variety of mathematical population models have been developed to describe Allee dynamics (e.g. Lewis and Kareiva 1993). It is difficult to compare many of the existing Allee models to the basic theta-logistic model, however, because their behaviors are different (Turchin 2003). As an alternative, I incorporated Allee behavior into the theta-logistic model with the addition of the parameter a, representing a point below which population growth becomes negative:

$N_{t+1} = N_t + r \cdot (N_t - a)\left(1 - \left(\frac{N_t}{K}\right)^{\theta}\right)$. The parameters in the Allee model retain their meaning in the theta-logistic model, the only difference being that maximum growth occurs at different population sizes. Holding all other parameters constant, increasing the Allee threshold has the effect of increasing the point at which maximum population growth occurs (N^*). The evidence for or against Allee dynamics is strongest at low population sizes – Allee events may be difficult to detect between N^* and K because over this region, the decline in per-capita growth rate is nearly identical for each model.

While the theta-logistic and Allee models are flexible and able to describe a range of time series, they may not be appropriate for records in the GPDD that appear to be stable. If population abundance estimates appear relatively constant, two hypotheses at opposite ends of the spectrum will be supported: (1) the population is at very low density (and possibly has suffered growth rates in recent years that are less than average), or (2) the population is at carrying capacity. Obviously, neither the theta-logistic model or Allee model should be favored under such uncertainty in the population state relative to carrying capacity. To account for cases where the population abundance estimates may be relatively stable, I included the geometric growth model as a third candidate (Case 2000). The geometric growth model can be expressed as $N_{t+1} = N_t \cdot (1+r)$, where r represents the discrete growth rate. In addition to cases where the population states appear stable, the geometric model is able to capture populations that appear to be declining or increasing exponentially.

Data for this analysis was collected from the over 5000 records in the GPDD. Only a subset of the 3200 GPDD data sets used in Sibly et al.'s (2005) analysis were used here. Unlike Sibly et al. (2005), I removed all Insecta records and focused on the remaining three classes: Aves, Mammalia, and Osteichthyes. The primary justification in removing the Insecta records is that the population dynamics of insects is known to be very different from vertebrates - many of the insect time records in the GPDD contain stochastic fluctuations and cyclic behavior, which cannot be captured with single-stage models. Before conducting parameter estimation, I rescaled the abundances each GPDD data sets to be on a 0.0-1.0 scale. Previous studies have shown that this rescaling increases the speed of the Gibbs sampling algorithm (Millar and Meyer 1999).

I assumed that the prior distributions were the same between models and across data sets (Table 6.1). For each model with density dependence, I used an upper bound of 2.0 on the prior for r to prevent chaotic oscillations (May 1976; Case 2000). The uniform prior assigned to K (0,5) implicitly assumes that the largest observation in any one data set is no smaller than 20% of K. While this assumption may appear to be somewhat arbitrary, if a population is only observed at densities < 20% of K, it is improbable that density dependence will not be detected (in this situation, an geometric model would fit the data better). For all three models, I assumed that observation error and process error were both gamma distributed, and parameterized in terms of the coefficient of variation (CV): if X ~ Gamma (v, w) with E[X] = v/w, CV(X) = $1/\sqrt{v}$. I chose to use the gamma distribution over a lognormal distribution because the gamma distribution has been shown to be less sensitive to outliers (Myers and Pepin 1990) and error misspecification (Firth 1988) in observation error models. Further, empirical studies have shown that the gamma distribution may be more appropriate than the lognormal distribution for some taxa in the GPDD (Halley and Inchausti 2002). For both the observation and process error CV, I placed an Inverse Gamma prior on the term $(CV)^2$: $1/CV^2 \sim$ Gamma (0.9, 0.05) , with 10% and 90% quantiles at 0.15 and 0.8.

In addition to requiring priors on population parameters, the Bayesian state-space model requires priors to be placed on the states of nature (true population sizes). For each model, I placed a gamma prior on the initial state, S_1 ~ Gamma (CV_{pro}, $2CV_{pro} / K$), where CV_{pro} represents the process error CV. This prior may appear confusing, but simply assumes that the prior density is centered at 50% of K, and has a coefficient of variation equal to CV_{pro}. For the remaining states of nature in each time series, I assumed that S_t ~ Gamma (CV_{pro}, CV_{pro} / P_t), where P_t represents the model predicted population state, calculated via the model (theta-logistic or Allee), population parameters, and the state of nature one time step earlier.

Parameter estimation for GPDD data sets was done in WinBugs 1.4 (Spiegelhalter et al. 2003). After a burn-in of 5000 iterations, the next 50,000 iterations were thinned to reduce correlation between parameter vectors, and every 10th sample was stored. Using a random sample of 50 data sets, I calculated the Geweke and Hiedelberger-Welch diagnostics in the CODA package to evaluate chain convergence (Best et al. 1995). After diagnostics indicated that the burn-in was sufficient (Gelman et al. 1995), the remaining data sets were analyzed in WinBugs. Of all GPDD records analyzed, a small proportion (< 3%) resulted in WinBugs crashing. Upon further inspection, these records were found to contain negative population sizes (possibly because a standardized index was recorded rather than actual counts). After deleting these records, the remaining 1979 records appeared to be representative of the three taxa (Aves = 498, Mammalia = 935, Osteichthyes = 546).

After the MCMC samples were stored, I calculated the joint prior distribution of all model parameters (including the population states) and the joint likelihood (e.g. Calder et al. 2003). Using a multivariate normal importance function, I computed the marginal likelihood of the data for each of the three models (also referred to as the posterior model probability). Using the BMA procedure to calculate the model-averaged posterior mean is relatively straightforward (Hoeting et al. 1999), however this same procedure may introduce bias when applied to other quantities of interest and asymmetric posterior distributions (e.g. median, mode, quantiles). The reason for this latter result is that the median (or mode) of the weighted posterior probability density function is not necessarily the same as the weighted average of the individual medians. Iterating through data sets, I generated three histograms for each parameter of interest (one per model). Each histogram used the same 500 bin intervals over the same support region of the parameter space. Using the previously calculated posterior model probabilities, I re-weighted the densities of each histogram and summed bins across histograms to generate a model-averaged histogram. This final histogram was then used to calculate the model-averaged posterior median and mode for each parameter.

Results

The posterior model probability favoring the Allee model was greater than 0.5 in only 1.2% of all records (Aves = 2, Mammalia = 19, Osteichthyes = 3). It is important to note that for all three taxa, the geometric growth model received the majority of the support. When the posterior probabilities of the Allee model and theta-logistic model were combined, the total probability favoring density dependence was greater than 0.5 in only 2.6% of all records (Aves = 6, Mammalia = 33, Osteichthyes = 13). This result suggests that while the GPDD time series may not strongly favor Allee dynamics, many of the records also do not support density dependence.

Examining histograms of the geometric growth rate for each of the three taxa shows that the Mammalia records had the highest estimated growth rate, with none of the time series supporting a posterior mode less than zero (Figure 6.1). This result is unexpected because many mammalian populations are thought to be slower growing relative to other vertebrates. For the geometric growth model in particular, the logarithm of the estimated growth rate tended to increase as either of the estimated error CVs increased; this relationship is particularly strong for mammal time series (Figure 6.2).

After computing the posterior model probabilities for each of the three population models, I estimated the Bayesian model averaged posterior modes for the error parameters (Figures 6.3-6.4). For all three taxa, the estimated observation error CVs tended to be smaller than process error CVs. This difference appeared to be greatest in Mammalia records, because the model averaged estimates of the observation error were smaller than those estimated for Aves or Osteichthyes. Another interesting result is that there did not appear to be a trade-off between observation and process errors across time series – when the CVs are examined in log-space, there is a strong positive correlation between the posterior modes for the two error types indicating that the split between the two types of error ($\rho \sim 0.66$, Figure 6.2). For any one time series, the observation and process error CVs tended to be negatively correlated ($\rho < -0.4$).

One final difference between the three taxa is in the estimated Allee threshold (as a proportion of carrying capacity). Examining the histograms of the posterior medians illustrates that the quantity a / K may be smaller in Mammalia records when compared to other taxa (Figure 6.5). Because the weight supporting the Allee model is relatively small, it is unclear whether there actually is a difference between taxa.

Discussion

Previous studies have highlighted the advantages of Bayesian methods in assessing evidence for depensation (Liermann and Hilborn 1997). This analysis provides similar support for Bayesian techniques, and shows that Bayesian model averaging is useful even in high-dimensional problems. Despite the fact that only a small fraction of the GPDD records provide support for Allee dynamics (1.2%), this estimated proportion is 5-6 times larger than the estimate found with maximum likelihood methods (Sibley et al. 2005). The advantage of the Bayesian approach to model selection and hypothesis testing is that it treats uncertainty in a framework that is compatible with decision making (Punt and Hilborn 1997). Rather than viewing evidence for depensation in a binary framework (it either exists or it doesn't), the Bayesian approach used in this analysis assigns a probability to the observed data being generated from a process with Allee dynamics.

For all three taxa in this study, the geometric model generally received much more weight than the Allee and theta-logistic models combined. When alternative parameter combinations are considered for the geometric model, mammal time series only support high growth rates when either observation error or process error is large. One potential explanation for this result is that mammalian populations are expected to grow more slowly than other vertebrate taxa – under some conditions (large measurement uncertainty or process error), the estimated population growth rate may be greater than expected. For other taxa, the relationship between growth rate and errors is slightly more flexible – higher growth rates are supported at low and high errors for the Aves and Osteichthyes time series.

This analysis was unable to reliably estimate the difference in Allee thresholds between taxa because of small sample sizes. Ideally, multiple data sets that support density dependence could be evaluated for evidence of Allee dynamics. The GPDD proved to be too small a sample size, with < 40 time series supporting density dependence. There are a number of reasons why time series in the GPDD might not support Allee effects – some records represent harvested populations (with unknown removals), while other records contain many missing observations. For many of the

Mammalia records, the observed estimates may be better seen as indices of abundance than absolute abundance indices. Most importantly, many GPDD records do not appear to represent populations over a wide range of densities (both at low density, and near carrying capacity). Observing the growth rates at the extremes of population sizes is critical for detecting Allee dynamics - both frequentist and Bayesian approaches will also fail if no portion of the time series occurs at low population density.

After considering the evidence of Allee effects in a Bayesian framework, it appears that Allee dynamics are not widely supported in the GPDD, but more common than the 0.2% suggested by Sibly et al. (2005). It is also important to note that the results described in this analysis should be extended to other populations with caution, as some observed patterns may be sensitive to the models chosen for comparison or the time series analyzed. The Allee model used in this analysis was chosen to provide a nested comparison to the theta-logistic model – in reality a more flexible model might be more appropriate (e.g. Lewis and Kareiva 1993; Keitt et al. 2001).

Table 6.1 Bayesian prior distributions for the geometric, theta-logistic, and Allee population models. The parameters CV_{obs} and CV_{pro} represent the observation and process error CVs, respectively.

Model	Parameter	Prior Distribution	Expected Value
Geometric	r	Uniform (-1, 2)	0.5
Theta-logistic / Allee	r	Uniform (0, 2)	1
Theta-logistic / Allee	K	Uniform (0, 5)	2.5
Theta-logistic / Allee	θ	Lognormal (u = -0.02, σ = 0.2)	1
Allee	a	Uniform (0, K)	$K/2$
All	S_1	Gamma ($CV_{pro}, 2CV_{pro}/K$)	$K/2$
All	$1/CV_{obs}^2$	Gamma (0.9, 0.05)	18
All	$1/CV_{pro}^2$	Gamma (0.9, 0.05)	18

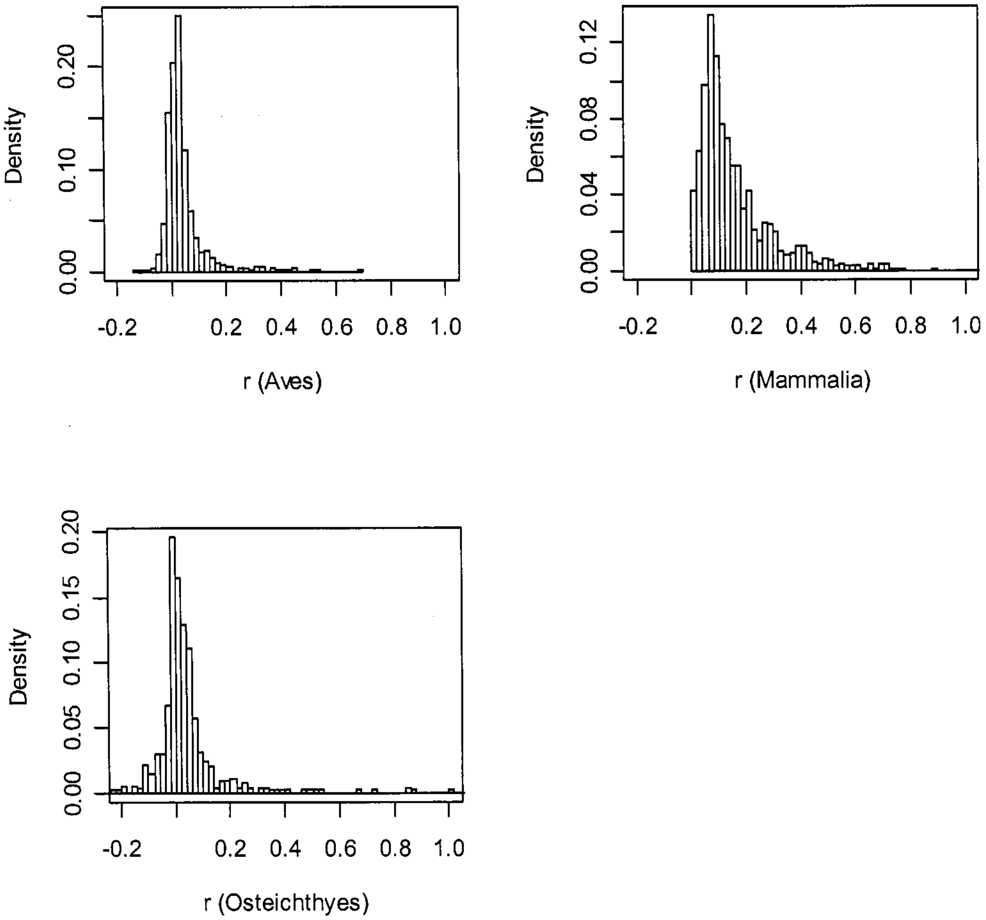

Figure 6.1. Estimated posterior modes of the growth rate (*r*) for the geometric model. Only records where the geometric growth rate received a posterior model probability of > 0.9 have been included.

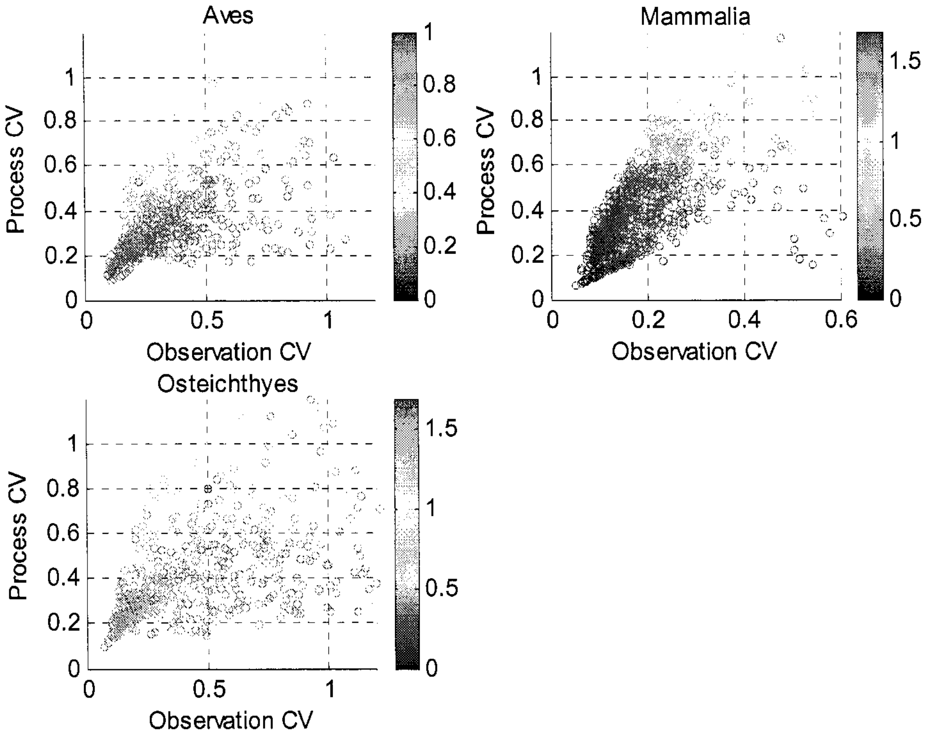

Figure 6.2. Scatterplot of observation and process error CVs for three taxa in the GPDD. Points are color coded by the value of the posterior mode of the geometric growth rate. For all three taxa, larger combinations of error parameters tend to favor larger growth rates.

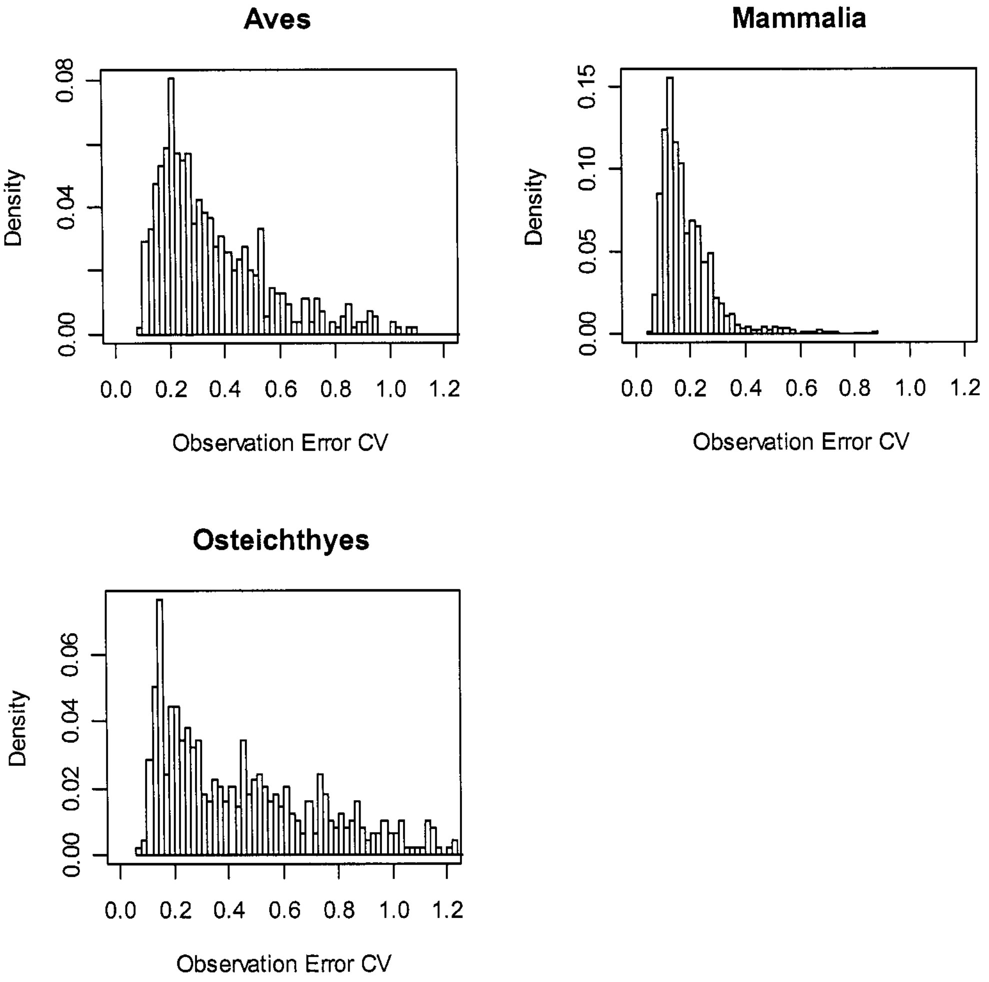

Figure 6.3. Bayesian model averaged estimates of the posterior modes for the observation error coefficient of variation (CV). The model averaged parameter estimate is computed for each time series across the three population models (geometric, theta-logistic, Allee).

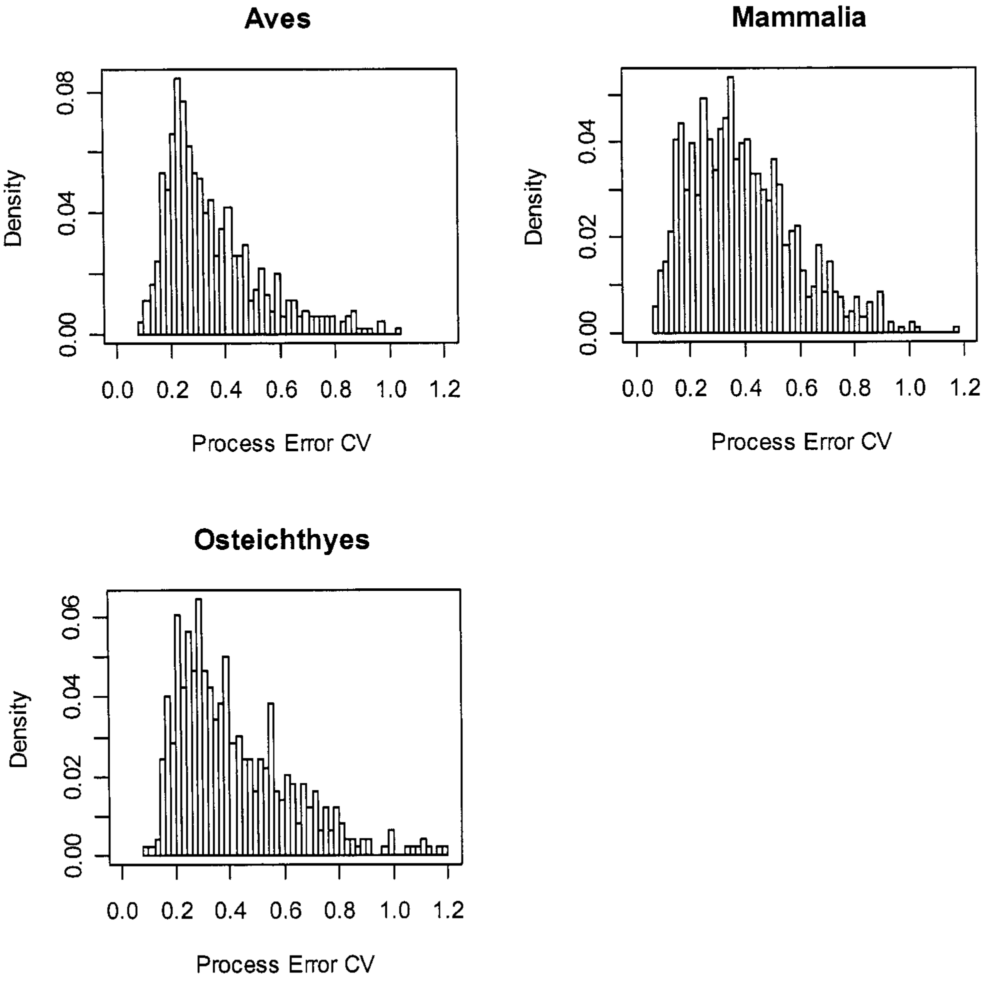

Figure 6.4. Bayesian model averaged estimates of the posterior modes for process error coefficient of variation (CV). The model averaged parameter estimate is computed for each time series across the three population models (geometric, theta-logistic, Allee).

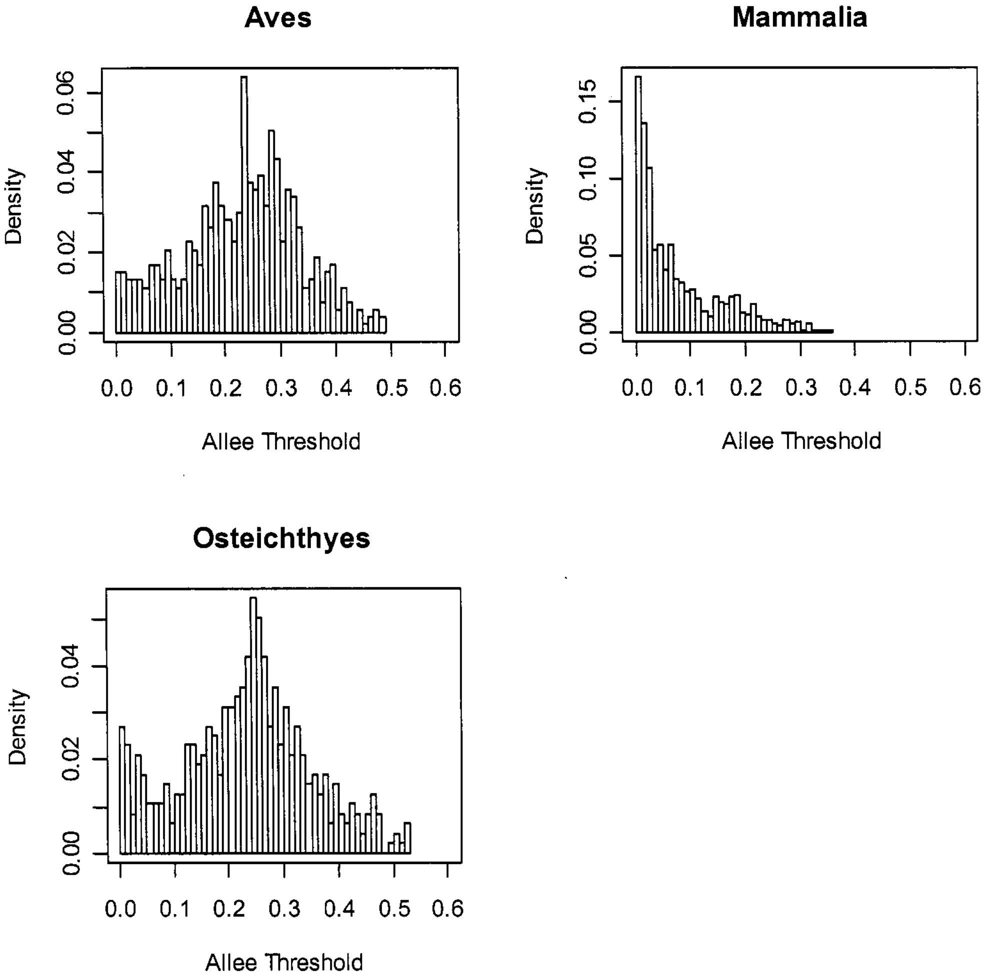

Figure 6.5. Estimated posterior medians of the Allee threshold as a proportion of carrying capacity (a / K). All records in the analysis are included, and the estimate is not averaged across models because the parameter a only appears in the Allee model.

References

Allee, W.C. 1938. The Social Life of Animals. W.W. Norton Inc, New York.

A'mar, Z.T. 2004. Quantifying error and uncertainty in fishery stock assessment models. M.Sc. Thesis, School of Aquatic and Fishery Sciences, University of Washington.

Akaike, H. 1973. Information theory as an extension of the maximum likelihood principle. Pp. 267-281 in B.N. Petrov and F. Csaksi, eds. 2nd International Symposium on Information Theory. Akademiai Kiado, Budapest, Hungary.

Anderson, D.R., K.P. Burnham, and G.C. White. 1994. AIC model selection in overdispersed capture-recapture data. Ecology, 75:1780-1793.

Andrewartha, H.G. and L.C. Birch. 1954. The Distribution and Abundance of Animals. University of Chicago Press, Chicago.

Angliss, R.P., and K.L. Lodge. 2002. Alaska marine mammal stock assessments, 2002. U.S. Department of Commerce, NOAA Technical Memorandum NMFSAFSC-133, 224p.

Barry, S.C., S.P. Brooks, E.A. Catchpole, and B.J.T. Morgan. 2003. The analysis of ring-recovery data using random effects. Biometrics, 59:54-65.

Begon, M., J.L. Harper, and C.R. Townsend. 1996. Ecology: Individuals, Populations, and Communities. Blackwell Science, Oxford.

Benoit, H. P., E. McCauley, and J. R. Post. 1998. Testing the demographic consequences of cannibalism in *Tribolium confusium*. Ecology, 79:2839–2851.

Bernardo, J.M. and A.F.M. Smith. 2000. Bayesian Theory. Wiley, New York.

Bernardo, J.M. and R. Rueda. 2002. Bayesian hypothesis testing: a reference approach. International Statistical Review, 70:351-372.
Best, N. G., M. K. Cowles, and S. K. Vines. 1995. CODA Manual version 0.30. MRC Biostatistics Unit, Cambridge, United Kingdom.

Brown R.E., B.E. Wright, S.D. Riemer, and J. Laake. 2005. Trends in abundance and current status of harbor seals in Oregon:1977-2003. Marine Mammal Science, 21:657-670.

Burnham, K.P., G.C. White, and D.R. Anderson. 1995. Model selection strategy in the analysis of capture-recapture data. Biometrics, 51:888-898.

Burnham, K.P. and D.R. Anderson. 2002. Model selection and multimodel inference: a practical information-theoretic approach. Springer, New York.

Burnham, K.P. and D.R. Anderson. 2004. Multimodel inference: understanding AIC and BIC in model selection. Sociological Methods and Research, 33:261-304.

Cadigan, N.G. and R.A. Myers. 2001. A comparison of gamma and lognormal maximum likelihood estimators in a sequential population analysis. Canadian Journal of Fisheries and Aquatic Sciences, 58:560–567.

Calabrese, J.M. and W.F. Fagan. 2004. Lost in time, lonely, and single: Reproductive asynchrony and the Allee effect. American Naturalist, 164:25-37.

Calder, C., M. Lavine, P. Muller, and J.S. Clark. 2003. Incorporating multiple sources of stochasticity into population dynamics models. Ecology, 84:1395-1402.

Carlin, B.P., and T.A. Louis. 2000. Bayes and Empirical Bayes Methods for Data Analysis. Chapman and Hall, New York.

Case, T. 2000. An illustrated guide to theoretical ecology. Oxford University Press, Oxford, United Kingdom.

Celeux, G., F. Forbes, C.P. Robert, and D.M. Titterington. 2006. Deviance Information Criteria for Missing Data Models. Bayesian Analysis, *In press*.

Chen, Y., and D. Fournier. 1999. Impacts of atypical data on Bayesian inference and robust Bayesian approach in fisheries. Canadian Journal of Fisheries and Aquatic Sciences, 56:1525–1533.

Cherry, S. 1998. Statistical tests in publications of the wildlife society. Wildlife Society Bulletin, 26:947-953.

Chib, S. 1995. Marginal likelihood from the Gibbs output. Journal of the American Statistical Association, 90:1313-1321.

Christensen, V., and D. Pauly. 1992. ECOPATH II – A software for balancing steady-state models and calculating network characteristics. Ecological Modelling, 61:169–185.

Clark, J.S., and O.N. Bjørnstad. 2004. Population time series: process variability, observation errors, missing values, lags, and hidden states. Ecology, 85:3140-3150.

Courchamp, F., T. Clutton-Brock, and B. Grenfell. 1999. Inverse density dependence and the Allee effect. Trends in Ecology and Evolution, 14:405-410.

Cui, G., A.E. Punt, L.A. Pastene, and M. Goto. 2002. Bayes and empirical Bayes approaches to addressing stock structure questions using mtDNA data, with an illustrative application to North Pacific minke whales. Journal of Cetacean Research and Management, 4:123-134.

Davis, H.G., C.M. Taylor, J.G. Lambrinos, and D.R. Strong. 2004. Pollen limitation causes and Allee effect in a wind-pollinated invasive grass (Spartina alterniflora). Proceedings of the National Academy of Sciences, 101:13804-13807.

de Valpine, P. 2002. Review of methods for fitting time-series models with process and observation error and likelihood calculations for nonlinear, non-gaussian state-space models. Bulletin of Marine Science, 70:455-471.

de Valpine, P., and A. Hastings. 2002. Fitting population models incorporating process noise and observation error. Ecological Monographs, 72:57-76.

Dennis, B. 2002. Allee effects in stochastic populations. Oikos, 96:389:401.

Dixon, P., and J.H.K. Pechmann. 2005. A statistical test to show negligible trend. Ecology, 86:1751–1756.

Doncaster, P.C. 2006. Comment on "On the regulation of populations of mammals, birds, fish, and insects" III. Science, 311:1100.

Drechsler M. and M.A. Burgman. 2004. Combining population viability analysis with decision analysis. Biodiversity and Conservation, 13:115:139.

Ellison, A.E. 2004. Bayesian inference in ecology. Ecology Letters, 7:509-520.

Ellison, A.E. 1996. An introduction to Bayesian inference for ecological research and environmental decision-making. Ecological Applications, 6:1036-1046.

Erb, J.D., and M. Boyce. 1999. Distribution of population declines in large mammals. Conservation Biology, 13:199-201.

Farnsworth, M.L., J.A. Hoeting, N.T. Hobbs, and M.W. Miller. 2006. Linking chronic wasting disease to mule deer movement scales: a hierarchical Bayesian approach. Ecological Applications, 16:1026-1036.
Field, S. A., A. J. Tyre, N. Jonzen, J. R. Rhodes, and H. P. Possingham. 2004. Minimizing the cost of environmental management decisions by optimizing statistical thresholds. Ecology Letters, 7:669–675.

Firth, D. 1988. Multiplicative errors: Log-normal or gamma? Journal of the Royal Statistical Society B, 50:266-268.

Fonnesbeck, C.J. 2006. PyMC: Markov Chain Monte Carlo for Python, version 1.0. http://pymc.sourceforge.net. [Date accessed 17 February 2006.]

Forster, M.R. 2000. Key concepts in model selection: performance and generalizability. Journal of Mathematical Psychology, 44:205-231.

Fournier, D. 1996. An introduction to AD Model Builder for use in nonlinear modeling and statistics. Otter Research Ltd.

Fowler, C.W. and J.D. Baker. 1991. A review of animal population dynamics and extremely reduced population levels. Report of the International Whaling Commission, 41:545-554.

Fried, S.M. and R. Hilborn. 1988. In-season forecasting of Bristol Bay, Alaska, sockeye salmon (Oncorhynchus nerka) abundance using Bayesian probability theory. Canadian Journal of Fisheries and Aquatic Sciences, 45:850-855.

Fritz, L.W. and E.S. Brown. 2005. Survey- and fishery-derived estimates of Pacific cod (*Gadus macrocephalus*) biomass: implications for strategies to reduce interactions between groundfish fisheries and Steller sea lions (Eumetopias jubatus). Fishery Bulletin, 103:501-515.

Gascoigne, J.C. and R.N. Lipcius. 2004a. Allee effects driven by predation. Journal of Applied Ecology, 41:801-810.

Gascoigne, J.C. and R.N. Lipcius. 2004b. Allee effects in marine systems. Marine Ecology Progress Series, 269:49-59.

Gelfand, A.E. and D.K. Dey. 1994. Bayesian model choice: asymptotics and exact calculations. Journal of the Royal Statistical Society Series B, 56:501-514.

Gelman, A. B., J. S. Carlin, H. S. Stern, and D. B. Rubin. 1995. Bayesian data analysis. Chapman and Hall, New York.

Gerber, L.R. and R. Hilborn. 2001. Catastrophic events and recovery from low densities in populations of otariids: implications for risk of extinction. Mammal Review, 31:131–150.

Geweke, J. 1992. Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In Bayesian Statistics 4 (ed JM Bernardo, JO Berger, AP Dawid and AFM Smith). Clarendon Press, Oxford, UK.

Ghosh, J.K., and T. Samanta. 2001. Model selection – an overview. Current Science, 80:1135-1144.

Gilbert, J.R. and N. Guldager. 1998. Status of harbor and gray seal populations in northern New England. Final Report to: National Marine Fisheries Service, Northeast Fisheries Science Center, Woods Hole, MA. Under NMFS/NER Cooperative Agreement 14-16-009-1557. 13 pp.

Gilbert J.R., G.T. Waring, K.M. Wynne, and N. Guldager. 2005. Changes in abundance of harbor seals in Maine, 1981-2001. Marine Mammal Science, 21:519-535.

Gilks, W. R., S. Richardson, and D. J. Spiegelhalter. 1996. Markov chain Monte Carlo in practice. Chapman and Hall, New York.

Gilpin, M.E. and F.J. Ayala. 1973. Global models of growth and competition. Proceedings of the National Academy of Sciences, 70:3590-3593.

Good, I.J. 1958. Significance tests in parallel series. Journal of the American Statistical Association, 53:799-813.

Goodman, S.N. 1999. Toward evidence-based medical statistics 1 – the p-value fallacy. Annals of Internal Medicine, 130:995-1004.

Halley, J. and P. Inchausti. 2002. Lognormality in ecological time series. Oikos, 99:518-530.

Harding, K.C., T. Härkönen, and H. Caswell. 2002. The 2002 European seal plague: epidemiology and population consequences. Ecology Letters, 5:727–732.

Harding, K.C., T. Härkönen, and J. Pineda. Estimating quasi-extinction risk of European harbour seals: a reply to Lonergan and Harwood (2003). Ecology Letters, 6:894-897.

Hayes, J. P., and R. S. Steidl. 1997. Statistical power analysis and amphibian population trends. Conservation Biology, 11:273–275.

Heidelberger, P. and P.D. Welch. 1983. Simulation run length control in the presence of an initial transient. Operations Research, 31:1109-1144.

Hilborn, R. and C.J. Walters. 1992. Quantitative Fisheries Stock Assessment: Choice, Dynamics, and Uncertainty. Kluwer Academic Publishers, Boston.

Hilborn, R. and M. Mangel. 1997. The Ecological Detective: Confronting Models with Data. Princeton University Press, Princeton.

Hilborn, R., M. Maunder, A. Parma, B. Ernst, J. Payne, and P. Starr. 2003. Coleraine: a generalized age-structured stock assessment model, user manual version 2.0. Available at: http://www.fish.washington.edu/research/coleraine/body.html. [Date accessed 17 August 2005.]

Hobbs, N.T. and R. Hilborn. 2006. Alternatives to statistical hypothesis testing in ecology: a guide to self teaching. Ecological Applications, 16:5-19.

Hoeting, J.A., D. Madigan, A.E. Raftery, and C.T. Volinsky. 1999. Bayesian model averaging: a tutorial. Statistical science, 14:382-401.

Holmes, E.E., and W.E. Fagan. 2002. Validating population viability analysis for corrupted data sets. Ecology, 83:2379-2386.

Hurvich, C.M. and C.L. Tsai. 1995. Model selection for extended quasi-likelihood models in small samples. Biometrics, 51:1077-1084.

Jeffreys, H. 1935. Some tests of significance, treated by the theory of probability. Proceedings of the Cambridge Philosophical Society, 31:203-222.

Jeffries, S. J., R. F. Brown, H. R. Huber, and R. L. DeLong. 1997. Assessment of harbor seals in Washington and Oregon 1996. Annual report to the MMPA Assessment Program, Office of Protected Resources, NMFS, NOAA, 1335 East-West Highway, Silver Spring, MD 20910. Available at National Marine Mammal 12 Laboratory, 7600 Sand Point Way NE, Seattle, WA 98115.

Jeffries, S., H. Huber, J. Calambokidis, and J. Laake. 2003. Trends and status of harbor seals in Washington state, 1978-1999. Journal of Wildlife Management, 67:207-218.

Johnson, J.B. and K.S. Omland. 2004. Model selection in ecology and evolution. Trends in Ecology and Evolution, 19:101-108.

Kadane, J.B., and N.A. Lazar. 2004. Methods and criteria for model selection. Journal of the American Statistical Association, 99:279-290.

Kalman, R.E. 1960. A new approach to linear filtering and prediction problems. Transactions of the American Society of Mechanical Engineers, 82:35-45.

Kass, R. 1993. Bayes factors in practice. The Statistician, 42:551-560.

Kass, R. and A.E. Raftery. 1995. Bayes factors. Journal of the American Statistical Association, 90:773-795.

Kass, R. and L. Wasserman. 1996. The selection of prior distributions by formal rules. Journal of the American Statistical Association, 91:1343-1370.

Katona, S. K., V. Rough, and D. T. Richardson. 1993. A field guide to whales, porpoises, and seals from Cape Cod to Newfoundland. Smithsonian Institution Press: Washington, DC.

Keitt, T.H., M.A. Lewis, and R.D. Holt. 2001. Allee effects, invasion pinning, and species borders. American Naturalist, 157:203-216.

Kuha, J. 2004. AIC and BIC: Comparisons of assumptions and performance. Sociological Methods and Research, 33:188-229.

Lack, D. 1954. The natural regulation of animal numbers. Claredon Press, Oxford.

Lande, R. 1993. Risks of population extinction from demographic and environmental stochasticity and random catastrophes. American Naturalist, 142:911–927.

Lande, R., S. Engen, and B. E. Saether. 2003. Stochastic population dynamics in ecology and conservation. Oxford University Press, Oxford, United Kingdom.

Laud, P.W., and J.G. Ibrahim. 1995. Predictive Model Selection. Journal of the Royal Statistical Society Series B, 57:247-262.

Lewis, M., and P. Kareiva. 1993. Allee dynamics and the spread of invading organisms. Theoretical Population Biology, 43:141–158.

Liermann, M. and R. Hilborn. 1997. Depensation in fish stocks: a hierarchic Bayesian meta-analysis. Canadian Journal of Fisheries and Aquatic Sciences, 54:1976-1984.

Liermann, M. and R. Hilborn. 2001. Depensation: evidence, models and implications. Fish and Fisheries, 2:33-58.

London, J.M. 2006. Harbor seals in hood canal: predators and prey. Ph.D. dissertation, School of Aquatic and Fishery Sciences, University of Washington.

Lotka, A.J. 1932. The growth of mixed populations: two species competing for a common food supply. Journal of the Washington Academy of Sciences, 22:461-469.

Loughlin, T. R., G. A. Antonelis, J. D. Baker, A. E. York, C. W. Fowler, R. L.

DeLong, and H. W. Braham. 1994. Status of the northern fur seal population in the United States during 1992. In Fur seal investigations, 1992 (ed E.H. Sinclair). U.S. Department of Commerce, NOAA Technical Memo. NMFS-AFSC-45.

Ludwig, D. and C.J. Walters. 1981. Measurement errors and uncertainty in parameters for stock and recruitment. Canadian Journal of Fisheries and Aquatic Sciences, 38:711-720.

Madigan, D., and A.E. Raftery. 1995. Model Selection and Accounting for Model Uncertainty in Graphical Models Using Occam's Window. Journal of the American Statistical Association, 89:1535-1546.

Mangel, M., and C. Tier. 1994. Four facts every conservation biologist should know about persistence. Ecology, 75:607-614.

May, R. 1974. Biological populations with non-overlapping generations: stable points, stable cycles, and chaos. Science, 186: 645–647.

McAllister, M.K. and G.P. Kirkwood. 1998. Bayesian stock assessment: a review and example application using the logistic model. ICES Journal of Marine Science, 55:1031-1060.

McQuarrie, A.D.R. and C-L. Tsai. 1998. Regression and time series model selection. World Scientific Publishing Company, London.

Meyer, R., and R.B. Millar. 1999a. BUGS in Bayesian stock assessments. Canadian Journal of Fisheries and Aquatic Sciences, 56:1078-1087.

Meyer, R., and R.B. Millar. 1999b. Bayesian stock assessment using a state-space implementation of the delay difference model. Canadian Journal of Fisheries and Aquatic Sciences, 56:37-52.

Millar, R.B. and R. Meyer. 1999. Nonlinear state–space modeling of fisheries biomass dynamics using Metropolis–Hastings within Gibbs sampling. Tech. Rep. STAT9901. Department of Statistics, University of Auckland, Auckland, N.Z.

Myers, R., and R. Pepin. 1990. The robustness of lognormal-based estimators of abundance. Biometrics, 46:1185-1192.

Myers, R.A. and B. Worm. 2003. Rapid worldwide depletion of predatory fish communities. Nature, 423:280-283.

Myers, R.A., N.J. Barrowman, J.A. Hutchings, and A.A. Rosenberg. 1995.

Population-dynamics of exploited fish stocks at low population-levels. Science, 269:1106-1108.

Myung, I.J. 2000. The importance of complexity in model selection. Journal of Mathematical Psychology, 44:190-204.

National Marine Fisheries Service. 2000. Stock assessment of Phoca vitulina richardsi: Washington Inland Waters Stock.

National Marine Fisheries Service. 2000. Stock assessment of Phoca vitulina richardsi: California Stock.

NERC Center for Population Biology, Imperial College. 1999. The Global Population Dynamics Database [online]. Available from http://www.sw.ic.ac.uk/cpb/gpdd.html [accessed 01 June 2006].

Newby, T. C. 1973. Changes in Washington State harbor seal population, 1942-1972. Murrelet, 54:5-6.

Parkinson, E.A. 1990. Impaired school formation at low density: a mechanism for depensatory mortality in sockeye salmon. BC Fisheries Branch Fisheries Management Report, 99:1-17.

Pedregal, D. 1996. Bayesian analysis of time series (BATS). International Journal of Forecasting 12: 429–432.

Pinhiero, J.C. and D. Bates. 2000. Mixed effects models in S and S-Plus. Springer-Verlag, New York.

Pitcher, K.W., and D.C. McAllister. 1981. Movements and haul out behavior of radio-tagged harbor seals, Phoca vitulina. Canadian Field Naturalist, 95:292-297.

Pole, A., M. West, and J. Harrison. 1994. Applied Bayesian forecasting and time series analysis. Chapman Hall, London, United Kingdom.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. 1992. Numerical recipes in C: the art of scientific computing. Cambridge University Press, Cambridge, United Kingdom.

Pulliam, H.R. 1988. Sources, sinks, and population regulation. American Naturalist, 132:653-661.

Punt, A. E., and R. Hilborn. 1997. Fisheries stock assessment and decision analysis: the Bayesian approach. Reviews in Fish Biology and Fisheries, 7: 35–63.

Punt, A.E. 2003. Extending production models to deal with process error in the population dynamics. Canadian Journal of Fisheries and Aquatic Sciences, 60:1217-1228.

Raftery, A.E. 1995. Bayesian model selection in social research. Sociological Methodology, 25:111-169.

Raftery, A.E., D. Madigan, and J.A. Hoeting. 1997. Bayesian model averaging for linear regression models. Journal of the American Statistical Association, 92:179-191.

Raftery, A.E. 1999. Bayes factors and BIC: Comment on Weakliem. Sociological Methods and Research, 27:411–427.

Reckhow, K.H. 1990. Bayesian inference in non-replicated ecological studies. Ecology, 71:2053-2059.

Ricker, W .E. 1954. Stock and recruitment. Journal of the Fisheries Research Board of Canada 11: 559–623.

Rivot, E., E. Prevost, E. Parent, and J.L. Bagliniere. 2004. A Bayesian state-space modelling framework for fitting a salmon stage-structured population dynamic model to multiple time series of field data. Ecological Modelling, 179:463-485.

Ross, J.V. 2006. Comment on "On the regulation of populations of mammals, birds, fish, and insects" II. Science, 311:1100.

Rubin, D. 1988. Using the SIR algorithm to simulate posterior distributions. In Bernardo JM, Degroof MH, Lindleyand DV, Smith AFM (eds) Bayesian Statistics, 3rd edition. Oxford University Press, Oxford.

Saether, B.E., S. Engen, and E. Matthysen. 2002. Demographic characteristics and population dynamical patterns of solitary birds. Science, 295:2070-2073.

Schervish, M.J. 1996. P-values: what they are and what they are not. American Statistician, 50:203-206.

Schreiber, S.J. 2003. Allee effects, extinctions, and chaotic transients in simple population models. Theoretical Population Biology, 64:201-209.

Schwarz, G. 1978. Estimating the dimension of a model. Annals of Statistics, 6:461-464.

Seber, G.A.F. 1982. The estimation of animal abundance and related parameters. New

Jersey: Blackburn.

Shono, Hiroshi. 2000. Efficiency of the finite correction of Akaike's Information Criteria. Fisheries Science, 66:608-610.

Sibly, R. M., D. Barker, M. Denham, J. Hone, and P. Pagel. 2005. On the regulation of populations of mammals, birds, fish and insects. Science, 309:607–610.

Spiegelhalter, D.J., N.G. Best, B.P. Carlin, and A. van der Linde. 2002. Bayesian measures of complexity and fit (with discussion). Journal of the Royal Statistical Society B, 64:583-540.

Spiegelhalter, D., A. Thomas, N. Best, and D. Lunn. 2003. WinBUGS user manual 1.4. MRC Biostatistics Unit, Institute of Public Health, Cambridge, United Kingdom. Available at: http://www.mrc-bsu.cam.ac.uk/bugs. [Date accessed 17 August 2005.]

Stanley, H.F., S. Casey, J.M. Carnahan, S. Goodman, J. Harwood, and R.K. Kayne. Worldwide patterns of mitochondrial differentiation in the harbor seal (P. vitulina). Molecular Biology and Evolution, 13:368-382.

Staples, D.F., M.L. Taper, and B. Dennis. 2004. Estimating population trend and process variation for PVA in the presence of sampling error. Ecology, 85:923-929.

Stone, M. 1979. Comments on model selection criteria of Akaike and Schwartz. Journal of the Royal Statistical Society B, 41:276-278.

Taylor, B.L., and T. Gerrodette. 1993. The uses of statistical power in conservation biology: the Vaquita and Northern Spotted Owl. Conservation Biology, 7:489-500.

Thomas, L., S.T. Buckland, K.B. Newman, and J. Harwood. 2005. A unified framework for modelling wildlife population dynamics. Australia and New Zealand Journal of Statistics, 47:19-34.

Thomas, L., J. L. Laake, S. Strindberg, F. F. C. Marques, S. T. Buckland, D. L Borchers, D. R. Anderson, K. P. Burnham, S. L. Hedley, J. H. Pollard, J. R. B. Bishop, and T.A. Marques. 2005. Distance 5.0. Research Unit for Wildlife Population Assessment, University of St. Andrews, United Kingdom. Available at: http://www.ruwpa.st-and.ac.uk/distance. [Date accessed 17 August 2005.]

Towell, R.G., R.R. Ream and A.E. York. 2006. Decline in northern fur seal (Callorhinus ursinus) pup production on the Pribilof Islands. Marine Mammal Science, 22:486-491.

Trites, A.W., P.A. Livingston, M.C. Vasconcellos, S. Mackinson, A.M. Springer and D. Pauly. 1999. Ecosystem change and the decline of marine mammals in the Eastern Bering Sea: testing the ecosystem shift and commercial whaling hypotheses. Fisheries Centre, University of British Columbia, Vancouver, Canada.

Turchin, P. 2003. Complex population dynamics. Princeton University Press, New Jersey.

Vaida, F., and S. Blanchard. 2005. Conditional Akaike information for mixed-effects models. Biometrika, 92:351-370.

ver Hoef, J.M. and K.J. Frost. 2003. A Bayesian hierarchical model for monitoring harbor seal changes in Prince William Sound, Alaska. Environmental and Ecological Statistics, 10:201-219.

Volterra, V. 1926. Variations and fluctuations in the numbers of individuals in animal species living together. In *Animal Ecology*, McGraw Hill, New York.

Wade, P.R. 2000. Bayesian methods in conservation biology. Conservation Biology 14: 1308–1316.

Walters, C.J., and D. Ludwig. 1981. Effects of measurement errors on the assessment of stock-recruitment relationship. Canadian Journal of Fisheries and Aquatic Sciences, 38:704-710.

Walters, C.J., and D. Ludwig. 1994. Calculation of Bayes posterior probability distributions for key population parameters. Canadian Journal of Fisheries and Aquatic Sciences, 51:713-722.

Ward, E.J. 2005. BEAST User Manual v. 1.0.7. Available at: http://students.washington.edu/warde/Beast.html. [Date accessed 16 February 2006].

Weiss, R. 1997. Bayesian sample size calculations for hypothesis testing. The Statistician, 46:185-191.

White, G. C. and K. P. Burnham. 1999. Program MARK: Survival estimation from populations of marked animals. Bird Study 46 Supplement, 120–138. Available at: http://www.cnr.colostate.edu/~gwhite/mark/mark.htm. [Date accessed 17 August 2005.]

Williamson, M. H. 1983. The land-bird community of Skokholm: ordination and turnover. Oikos, 41:378–384.

Wintle B.A., M.A. McCarthy, C.T. Volinsky, and R.P. Kavanagh. 2003. The use of Bayesian model averaging to better represent uncertainty in ecological models. Conservation Biology, 17:1579-1590.

York, A.E., and P. Kozloff. 1987. On the estimation of numbers of northern fur seal, Callorhinus ursinus, pups born on St. Paul Island, 1980-86. Fishery Bulletin, 85:367-375.

York, A.E. 1991. Sea surface temperatures and their relationship to the survival of juvenile male Northern Fur Seals from the Pribilof Islands. In Pinnipeds and El Niño (ed F. Trillmich and K.A. Ono). Springer-Verlag, Berlin.

York, A.E., and R. Towell. 1997. Can we return to estimating numbers of northern fur seals from subsamples of rookeries? In Fur seal investigations, 1995 (ed E.H. Sinclair). U.S. Department of Commerce, NOAA Tech. Memo. NMFS-AFSC-86.

York, A.E., R.G. Towell, R.R. Ream, J.D. Baker, and B.W. Robson. 2000. Population assessment, Pribilof Islands, Alaska, p. 7-26. In Fur seal investigations, 1998 (ed B.W. Robson). U.S. Department of Commerce, NOAA Tech. Memo. NMFS-AFSC-113.

York, A.E. R.G. Towell, R.R. Ream and C.W. Fowler. 2002. Population assessment, Pribilof Islands, Alaska , p. 7-32. *In* Fur seal investigations, 2000-2001 (ed B.W. Robson). U.S. Department of Commerce, NOAA Tech. Memo. NMFS-AFSC-134.

Young, T.P. 1994. Natural die-offs of large mammals: implications for conservation. Conservation Biology, 8:410-418.

Curriculum Vitae

Eric J. Ward

Date and Place of Birth

Born January 27, 1977 in San Jose, CA

Education

Ph.D. (anticipated Nov. 2006) University of Washington. Dissertation topic: Bayesian PVA, model selection, and decision analysis.

M.Sc. (Biology) Montana State University at Bozeman, Bozeman, MT (2003). Thesis: "Evaluating Trends and Biases in Shipboard Tuna Vessel Data Used in the Estimation of Dolphin Abundance".

B.S. (Evolutionary Biology and Ecology) University of California at San Diego, San Diego, CA (1999).

Employment

2003 – 2006: Graduate research assistant in the School of Aquatic and Fishery Sciences, University of Washington. Funding provided by the National Marine Fisheries Service – Washington State Sea Grant fellowship in population dynamics.

2003 – 2006: Undergraduate field trip leader in the Department of Biology, University of Washington. Field trips include marine biology, ornithology, conservation biology, and general ecology.

February – March 2005: Stock assessment scientist for the New Zealand Seafood Industry Council, Wellington, New Zealand. Conducted assessment for the orange roughy (Hoplostethus atlanticus) Challenger 7A stock.

July – August 2004: Scientific observer on the Alaskan Cetacean Ecosystem (ACE) cruise in Alaska. Participated in marine mammal line transect surveys (focusing on killer whales) along the Alaskan peninsula and Aleutian chain, in addition to a northern right whale tagging project in the Bering Sea.

July 2003: Scientific observer on the Stenella Abundance Research (STAR) cruise in the eastern tropical Pacific. Participated in 5-week marine mammal line transect survey from San Diego to Hawaii, conducting marine mammal observations and collecting oceanography data.

2000 – 2003: Graduate research assistant (RA) in the Department of Ecology, Montana State University. Funding provided by NMFS, Southwest Fisheries Science Center.

Selected Publications

Zerbini, A.N., Ward, E., Engel, M.E., Kinas, P.G. and Andriolo, A. 2006. A Bayesian Assessment of the Conservation Status of Humpback Whales (*Megaptera novaeangliae*) in the Western South Atlantic (Breeding Stock A). *In press.*

Ward, E., Zerbini, A.N., Engel, M.E., Kinas, P.G. and Andriolo, A. 2006. Estimates of population growth rates of humpback whales (*Megaptera novaeangliae*) in the wintering grounds off the coast of Brazil (Breeding Stock A). *In press.*

Ward, E.J. 2006. A new BEAST: Bayesian Software Tools for Ecological Trend Analysis. Wildlife Society Bulletin (*In press*).

Branch, T., A. Haynie, R. Hilborn, G. Fay, L.Flynn, J. Griffiths, K. Marshall, J. Randall, J. Scheurell, E.J. Ward, and M. Young. 2006. Fleet dynamics and fishermen behavior: lessons for fisheries managers. Can. J. Fish. Aquat. Sci. **63**:1647-1668.

Ward, E.J. 2005. Differences between fishery-dependent and fishery-independent estimates of single-species and mixed-species dolphin schools: implications for single-species stock assessments. Mar. Mamm. Sci **21**:189-203.

Ward, E.J., R. Hilborn. 2006. A state-space mixture model for estimating catastrophic events in time series data. (*In press, Can. J. Fish. Aquat. Sci.*).

Ward, E.J., R. Hilborn, Towell, R., and L. Gerber. 2006. Do northern fur seal populations experience catastrophes? (*In press, Can. J. Fish. Aquat. Sci.*).

Ward, E.J. 2006. Allee effects: really rare or just difficult to detect? (*In review*).

Reports

Ward, E.J. and A. Hicks. 2005. Assessment of the Challenger Plateau (ORH 7A) Orange Roughy Fishery for the 2004-2005 Fishing Year. Paper DWWG 02/05.

Ward, E.J. 2005. BEAST – Bayesian Ecological Analysis of Statistical Trends, User Manual v. 2.0.1. Available at http://www.beastsoftware.org.

Zerbini, A.N., Ward, E., Engel, M.E., Kinas, P.G. and Andriolo, A. 2006. A Bayesian Assessment of the Conservation Status of Humpback Whales (*Megaptera novaeangliae*) in the Western South Atlantic (Breeding Stock A). Paper SC/58/SH2.

Ward, E., Zerbini, A.N., Engel, M.E., Kinas, P.G. and Andriolo, A. 2006. Estimates of population growth rates of humpback whales (*Megaptera novaeangliae*) in the wintering grounds off the coast of Brazil (Breeding Stock A). Paper SC/58/SH14.

Teaching

Winter 2006 – Pre-doctoral instructor for Fish 497: Introduction to Scientific Programming. Duties included curriculum design, lectures, labs, and grading.

Spring 2006 – Teaching Assistant for Fish 458: Modeling for Conservation of Populations. Duties included organizing and running labs, occasional lectures, and grading.

Winter 2005 - Organized graduate student seminar for object-oriented programming at the University of Washington.

Selected Presentations

2006. New Methods For Incorporating Catastrophes Into Population Time Series. Annual Meeting of National Marine Fisheries Service Fellows in Population Dynamics, La Jolla, CA.

2005. Can Catastrophes be Included in Stock Assessments and PVAs? Quantitative Seminar, School of Aquatic and Fishery Sciences, University of Washington.

2005. Estimating Catastrophic Extinction with State Space Models. 90th Annual Meeting of the Ecological Society of America, Montreal, Quebec.

2005. Robustness of Bayesian Model Selection Criteria / State Space Modeling of Catastrophic Events in Population Time Series. Annual Meeting of National Marine Fisheries Service Fellows in Population Dynamics, Seattle, Washington.

2004. Approximate Bayes Factors in Biology. Annual Graduate Student Symposium, School of Aquatic and Fishery Sciences, University of Washington.

2004. Seeing Double: Differences Between Research and Tuna Vessel Sightings of Spotted Dolphins in the Eastern Tropical Pacific, 1980-2000. Society for Marine Mammalogy Northwest Regional Chapter.

2004. New Software for Detecting the Decline of Biological Populations. Annual Meeting of National Marine Fisheries Service Fellows in Population Dynamics, Woods Hole, Massachusetts.

2002. Model selection in a stochastic environment. 86th Annual Meeting of the Ecological Society of America, Madison, Wisconsin.

Awards

Recipient of National Marine Fisheries Service / Washington Sea Grant Fellowship in Population Dynamics. Research grant awarded for $120K for 3 years.

Memberships and Affiliations

American Fisheries Society

Ecological Society of America

Society for Marine Mammalogy

Society for Conservation Biology