Best Practices for Constructing Size-Structured Population Dynamics Models used for Stock Assessments

Lee Cronin-Fine

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2021

Reading Committee:

André E. Punt, Chair

Timothy E. Essington

William T. Stockhausen

Program Authorized to Offer Degree:

Quantitative Ecology and Resource Management

University of Washington

Abstract

Best Practices for Constructing Size-Structured Population Dynamics Models used for Stock Assessments

Lee Cronin-Fine

Chair of the Supervisory Committee:

André E. Punt

School of Aquatic and Fisheries Science

All models are predicated on assumptions which makes them simplified versions of reality. An important goal of stock assessment scientists is to expand the capability of stock assessment models to improve their ability to estimate the size of fished populations and how they will change over time. Several types of population dynamics models can be used for stock assessment, with age-structured models, which track the population by age, being the most popular. Unfortunately, there are valuable fished species such as crabs and lobsters that are difficult to age. Size-structured models that track the population by size are good alternatives. The objective of this dissertation is to improve the performance of size-structured models used for stock assessments. Three aspects of size-structured models are explored: growth, selectivity, and natural mortality. Size-transition matrices define growth in size-structured models. They are constructed from an underlying growth curve, typically the von Bertalanfy growth curve, defined by three parameters: the growth rate (k), the asymptotic height (L_∞), and the variation in the size

increment. Most assessments assume individuals follow a single growth curve with process error, which is unrealistic. A new size-transition matrix construction method that allows L_∞ and k to vary among individuals through numerical integration was developed and compared with methods that allow individuals to follow (a) a single growth curve with process error, or (b) one of three growth curves, each with process error. The number of size-classes in the size-transition matrix and how the data are generated heavily dictate performance. Not accounting for temporal variation in selectivity can lead to biased estimates of abundance and mortality. Simulations suggest that discrete time blocking of selectivity can adequately capture time-varying selectivity as this could reduce the number of estimated parameters and hence the variance of estimated quantities. As for likelihood functions for size-composition data, the results reveal that multinomial, Dirichlet-multinomial and multivariate normal are all valid options. Natural mortality (M) has a strong influence on stock assessment model outputs including estimates of spawning stock biomass, MSY and fishing mortality. Estimating M is difficult since it is confounded with several factors including catchability, recruitment, and growth. Simulation shows that terminal molt does not affect the ability to estimate M. However, estimating growth simultaneously with M has a negative impact on the ability to estimate M but a positive effect on the quality of the estimates for spawning stock biomass.

Table of Contents

List of Figures

List of Tables

Acknowledgements

There is a plethora of people I wish to acknowledge. First, I am eternally grateful to the chair of my supervising committee, André Punt. You have been a fantastic advisor throughout this entire experience. You have continued to offer guidance, answer questions, and help me when I was stuck. I could not have asked for a better advisor. Thank you for your support and all you have done for me during my PhD.

To my entire committee, Tim Essington, John Skalski, William Stockhausen, Michael Dalton and Phil Levin, thank you for your continued support. You all have been valued members of my committee. I appreciate the advice you have given me over the years on how I could improve my work. You all provided thoughtful questions during my committee meetings and ensure that my research did not go off the rails.

I enrolled at the University of Washington as a QERM student and joined the Punt lab in the SAFS department a year later. Both the QERM and SAFS communities were very welcoming. These communities focus more on support and collaboration than competition. If I ever had problems understanding a certain concept or difficulties with my code there would always be multiple people that I could turn to for help. I would especially like to acknowledge my fellow lab-mates in the Punt lab, past and present, and my QERM cohort, Elliot Koontz, for discussion concepts with me when I was stuck and helping me practice for presentations.

Finally, I would like to acknowledge my funding sources, North Pacific Research Board, Joint Institute for the Study of the Atmosphere and Ocean (JISAO) and NOAA under the Cooperative Institute for Climate, Ocean and Ecosystem Studies.

Dedication

I dedicate my dissertation to my loved ones, especially my wife Nisha Singh.

Introduction

The world's oceans provide a plethora of ecosystem resources, of which fisheries are an important one given the food and jobs they support (NMFS, 2018). In the past, some scientists argued that the world's fisheries were inexhaustible (e.g., Huxley, 1884). Unfortunately, it is now generally recognized that fisheries, as with all the ocean's resources, are limited and require proper management. Over the years, the United States has enacted laws that outline how to manage fisheries within US waters. Currently, the Magnuson-Stevens Fisheries Conservation and Management Reauthorization Act (MSFCA) is the key law governing federally managed US fishery stocks. National Standard 2 of the MSFCA requires "that the national fishery conservation and management program utilizes, and is based upon, the best scientific information available" (Federal Register, 2009). Determining the number of individuals in a fished population and how many can be removed to achieve optimum yield is, however, difficult (Maunder and Piner 2015). This is because it is not possible to directly census marine organisms. Therefore, fisheries management often uses stock assessments based on population dynamics models to estimate the sizes of populations and the optimal amount that can be harvested from each.

All models are simplified versions of reality. They are used to understand the underlying drivers for population dynamics and make predictions on future behavior based on prior knowledge and data collected for monitoring purposes. All models are predicated on assumptions, which means no model truly reflects the real world. For example, some models assume that natural mortality is independent of time, age and size. Such assumptions can limit a model's ability to capture the true dynamics, resulting in inaccurate prediction of current biomass and how the population will behave given candidate management actions. A goal of stock assessment scientists

is to expand the capabilities of stock assessment methods to improve the ability to estimate the sizes of populations of fished species and how they will change over time.

A variety of models are available for conducting a stock assessment. Many stock assessments use a version of an age-structured population dynamics model and hence track population abundance by age. Unfortunately, important fished species such as crabs, lobster and prawns have exoskeletons, which molt on an irregular basis making it difficult to age individuals. Scientists have developed alternative models, such as surplus production models, which are applicable for hard-to-age species (Smith and Addison 2003). However, surplus production models do not account for size- and/or age-structured dynamics, which can lead to biased estimates of, for example, the biomass associated with Maximum Sustainable Yield (B_{MSY}) (Punt and Szuwalski, 2012). Size-structured population dynamics models provide a viable option for the stock assessment of hard to age species. These models assume that population dynamic processes, such as mortality, can vary with size and can use a wide variety of data types for parameter estimation purposes (Punt et al., 2013). They have been adapted for use in many fisheries, including the crab stocks in the Bering Sea and Aleutian Islands region of Alaska (e.g., Szuwalski and Turnock, 2016).

Fisheries that rely on stock assessments based on size-structured models can be valuable. For example, the ex-vessel commercial landings for the crab stocks in the Gulf of Alaska and Bering Sea/Aleutian Islands region of Alaska was 31.9 million lbs in 2018, which resulted in an estimated ex-vessel revenue of $169 million (Garber-Yonts and Lee, 2020). These fisheries provide income for the fishing vessel's crew, captain and vessel owners, processor plant workers and to local communities. Each year, a total allowable catch (TAC) is set for US federally managed species which is based on the overfishing limit (OFL) determined using stock assessments. The OFL is

the calculated catch limit associated with the maximum sustainable yield (MSY). Improving the performance of stock assessments based on size-structured models can potentially increase the accuracy of OFL estimation and better achieve optimum yield.

The equation for a size-structured population dynamics model is:

$$N_{y,j} = \sum_{i \le j} X_{i,j} N_{y-1,i} e^{-Z_{y-1,i}} + R_{y,j} \quad \text{(I.1)}$$

where $N_{y,j}$ is the number of individuals in size-class j at the start of year y, $X_{i,j}$ is the probability of growing from size-class i to size-class j in one year, $Z_{y,i}$ is the total mortality during year y for individuals in size-class i and $R_{y,j}$ is the number of recruits during year y to size-class j. Within this general equation is a multitude of assumptions. Examples include, but are not limited to, is there individual variation in growth, is there a stock-recruit relationship, does natural mortality vary over time or by sex or by maturity stage, and what is the shape of the selectivity curve. The results of stock assessments depend not only on the assumptions underlying the population dynamics model but also on how the data are used for parameter estimation, i.e. the likelihood function. The goal of this thesis is to examine and identify best practices for the assumptions used in stock assessments based on size-structured models so that they can better capture the true underlying dynamics.

A key component of a size-structured population dynamics model is the size-transition matrix. This matrix specifies the probability of an individual growing from one size-class to another after a certain period of time. Equation I.2 is an example of a size-transition matrix with four size-classes:

$$\begin{bmatrix} x_{11} & x_{12} & x_{13} & x_{14} \\ 0 & x_{22} & x_{23} & x_{24} \\ 0 & 0 & x_{33} & x_{34} \\ 0 & 0 & 0 & x_{44} \end{bmatrix} \quad \text{(I.2)}$$

where $x_{i,j}$ represents the probability of an individual transitioning from size-class i to size-class j. In Equation I.2, the rows represent the starting size-class, and the columns represent the ending size- class after one time step. An underlying growth curve is used to determine the values within the matrix. An important assumption within the size-transition matrix is how to account for individual variation in growth. One way this is accomplished is to assume that all individuals follow the same growth curve, and that deviations from this curve are due to process error. However, Punt et al. (2016) showed that assuming there is only a single growth curve when there is individual variation in growth can result in biased estimates of growth parameters. A group of alternative methods allows certain growth parameters to vary among individuals, thus creating separate growth curves (Punt et al., 2009). A variety of growth equations, each with a separate set of parameters, is available for constructing a size-transition matrix (e.g., Punt et al., 2016; Francis 1988; Turnock and Rugolo 2013). For my work, I assume that expected size follows the von Bertalanffy growth curve, which is commonly used in stock assessments (Punt et al., 2016):

$$l_j = l_i + (L_\infty - l_i)(1 - e^{-kt}) \tag{I.3}$$

where L_∞ is the asymptotic size, k is the growth rate and t is the number of time steps. Punt et al. (2009) evaluated the performance of several methods for constructing a size-transition matrix based on the von Bertalanffy growth curve and determined that allowing k to vary among individuals to be the most robust. Punt et al. (2009) did not evaluate a method that allowed both k and L_∞ to vary among individuals due to the complexity of the equations and the correlation between parameters k and L_∞. Chapter 1 of this thesis develops a method for constructing size-transition matrices that allows both k and L_∞ to vary among individuals.

Chapter 2 of this dissertation compares the method developed in Chapter 1 with two alternative methods for constructing a size-transition matrix. The alternative methods differ in terms of their

assumptions about how to account for individual variation in growth. One alternative method assumes all individuals follow a single growth curve with deviation due to process error. The other alternative method assumes individuals can follow one of three curves through the 'platoon' method, which has been used to model finfish growth (e.g., Methot and Wetzel, 2013; Punt *et al.*, 2001, 2017). The goal is to determine if the method from chapter 1 can outperform existing methods for constructing size-transitions matrices and provide guidance on how to include individual variation in growth in size-structured models.

Chapter 3 explores two aspects of selectivity within size-structure models. Within stock assessments that do not assume a spatial structure, the concept of selectivity refers to the availability of an animal, of a particular size or age, to the fishing gear and the probability of that animal being caught, given that it encounters the fishing gear. Selectivity ranges in value from 0 to 1, and is modeled using various curves, such as, but not limited to, the logistic, double logistic and double normal (e.g., Methot and Wetzel, 2013; Hulson and Hanselman, 2014). An entire fishery can have a single selectivity curve, or "fleets" within a fishery can have separate curves. There can be considerable inter-annual variation in selectivity-at-age (Sampson and Scott, 2012). A variety of factors can cause this, such as changes over time in the use of fishing gear types, fish behavior in relation to fishing gear and the spatial distribution of fish (Sampson, 2014). Ignoring variation in selectivity can lead to biased estimates of abundance and mortality rates (Gudmundsson and Gunnlaugsson, 2012) as well as optimistic projections for stock abundance (Walters and Maquire, 1996). Previous studies that explored the use of time-varying selectivity focused on age-structured models (e.g., Martell and Stewart, 2014; Linton and Bence, 2011). Little attention has placed on time-varying selectivity in relation to size-structured models. Therefore,

the first part of Chapter 3 explores how best to account for time-varying selectivity within size-structured models.

When estimating the selectivity curve, a key source of information is the size-composition data. These data, collected from fisheries and/or surveys, are used to specify the likelihood function that is used to estimate the parameters of the population dynamics and observation model, including those upon which selectivity is based. Typically, stock assessments use a multinomial or multinomial-based likelihood for composition data because these functions are assumed to closely match the random sampling design used to collect the data (Fournier and Archibald, 1982), or (more often) because it is conventional. Maunder (2011) summarizes potential likelihood functions for size- and/or age-composition data, which include the multinomial, Dirichlet-multinomial and multivariate normal.

Simulation studies exploring selectivity usually generate size- and/or age-composition data in a fashion similar to the likelihood function used when including these data in the assessment. This limits the generality of such studies because the variability in the size- and/or age-composition data could differ from that implied by the assumed likelihood (Punt et al., 2014). Therefore, Chapter 3 also explores alternative likelihood functions for size-composition data to determine which is most robust.

The final chapter of this dissertation examines estimation of natural mortality within size-structured models. Natural mortality (M) accounts for all mortality that is not directly associated with fishing. M strongly influences the outputs from a stock assessment since it directly impacts the productivity of the stock (Clark, 1999; Kenchington, 2014; Punt et al., 2021). M is a difficult parameter to estimate since it is confounded with catchability, fishing mortality, recruitment and growth (Schnute and Richards, 1995; Gislason et al., 2010). For this reason, many stock

assessments pre-specify *M* to an externally estimated value (Wang and Ellis, 2005) However, studies are now suggesting that it is better to estimate *M* within stock assessments (e.g., Johnson et al., 2015; Lee et al., 2011; Punt et al., 2021). These studies have focused on age-structured models. Size-structured models do not use age data and have unique characteristics such as a size-transition matrix that could influence the ability of a stock assessment to estimate *M*. Therefore, the fourth chapter of this dissertation will explore the ability to estimate *M* within stock assessments based on size-structured models. This chapter aims to determine if it is possible to estimate *M* and understand what factors influence the ability to estimate *M* accurately.

The theme behind my dissertation is to evaluate and explore assumptions within stock assessments based on size-structure population dynamics models to improve the ability of such assessment methods to estimate quantities on which management decisions are based, such as the time-trajectory of spawning stock biomass. The first two chapters focus on developing a new method for constructing size-transition matrices that allows for individual variation in growth and testing its performance compared to previous methods. Chapter 3 investigates how best to incorporate time-varying selectivity within size-structured models and what likelihoods are best for size-composition data. Chapter 4 investigates natural mortality within size-structured population dynamics models and evaluates what factors influence the ability to estimate *M* accurately and precisely.

Chapter 1 : A Numerical Method for Modelling Individual Variation in Growth

1.1 Abstract

Stock assessment methods for many invertebrate stocks, include stocks of crabs in the Bering Sea and Aleutian Islands region of Alaska, are based on size-structured population dynamics models. A key component of these models is the size-transition matrix, which specifies the probability of growing from one size-class to another after a certain period of time. Size-transition matrices can be defined using three parameters, the growth rate (k), the asymptotic height (L_∞), and the variation in the size increment. Most assessments assume that all individuals follow the same growth curve. However, not accounting for individual variation in growth can result in biased estimators of growth parameters and it is unrealistic to assume that every individual has the same k or L_∞. Unfortunately, to date, the only way to compute the size-transition matrix when allowance is made for individual variation in growth is using simulation, which is both computationally very intensive and the resulting likelihood function is non-differentiable with respect to the estimated parameters. We outline an approach that uses a numerical approximation technique that allows k and L_∞ to vary among individuals. This approach is evaluated by comparing the approximated matrix to a "true" matrix created through simulation.

1.2 Introduction

Fisheries that target crustaceans (crabs, lobsters and shrimps) are widespread and highly valuable. For example, the crab fisheries in the Bering Sea and Aleutian Islands region of Alaska produced an estimated gross ex-vessel revenue of $246 million (Garber-Yonts and Lee, 2014). The management of most crab fisheries, including those in the Bering Sea, are based on stock

assessment models that estimate the size of the population. Assessments for finfish fisheries typically use some version of an age-structured population dynamics model, which rely on age-composition data. Obtaining such data for crustaceans is difficult since they have exoskeletons, which molt on an irregular basis and remove potential indicators of age. Surplus production models provide a potential alternative, but they can lead to biased estimates of population status (Punt and Szuwalski, 2012). Alternative methods are needed for crustacean and similar species.

Size-structured population dynamics models provide a viable option for the stock assessment of hard-to-age species. These models track groups within the population by size, instead of age, and allow population dynamic processes, such as mortality, to vary with size. Parameter estimation, with these models, uses a wide variety of data types, such as the size-composition of catch, and tag-recapture data (Punt et al., 2013). Many fisheries, including the crab stocks in the Bering Sea and Aleutian Islands region of Alaska, American lobster (*Homarus americanus*) in the Atlantic, Rainbow abalone (*Haliotis iris*) in New Zealand, and tiger and endeavor prawns in Australia (Szuwalski and Turnock, 2016; ASMFC, 2015; Marsh and Fu, 2017; Buckworth et al., 2015) use size-structured stock assessment methods.

A key component of a size-structured population dynamics model is the size-transition matrix. This matrix specifies the probability of an individual growing from one size-class to another after a certain period of time. Equation 1.1 is an example of a size-transition matrix with four size classes:

$$\begin{bmatrix} x_{11} & x_{12} & x_{13} & x_{14} \\ 0 & x_{22} & x_{23} & x_{24} \\ 0 & 0 & x_{33} & x_{34} \\ 0 & 0 & 0 & x_{44} \end{bmatrix} \tag{1.1}$$

where $x_{i,j}$ represents the probability of an individual transitioning from size-class *i* to size class *j*. In Equation 1.1, the rows represent the starting size-class and the columns represent the ending

size-class after one time step. A modeler can choose the ranges of sizes defining each size-class. Each row of the size-transition matrix sums to one, since individuals must fall into one of the size-classes after growing. Mortality is assumed to occur before or after growth.

Elements within the size-transition matrix can be determined using a population growth function, when such a function can be estimated. A variety of growth functions, each with a separate set of parameters and assumptions, are possible (e.g. Punt et al., 2016; Francis 1988; Turnock and Rugolo 2013). The von Bertalanffy growth curve (Equation 1.2) is commonly used in stock assessments (Punt et al., 2016).

$$l_j = l_i + (L_\infty - l_i)(1 - e^{-kt}) \quad (1.2)$$

where L_∞ is the asymptotic size, k is the growth rate and t is the number of time steps. Realistically, each individual in a population expresses its own growth trajectory. Thus, Equation 1.2 expresses the average size. Stock assessments should account for individual variation in growth. One way to accomplish this is to allow for deviations due to process errors around the average growth curve. However, Punt et al. (2016) showed that assuming there is only a single growth curve when there is individual variation in growth could bias estimates of growth parameters. Alternative methods allow growth parameters to vary among individuals. Punt et al. (2009) determined that allowing k to vary among individuals had the most robust outcomes. However, that study did not explore the benefits of jointly allowing k and L_∞ to vary due to the analytical complexity of such models, and the high correlation between the two parameters. Developing a method for constructing size-transition matrices that allows both k and L_∞ to vary among individuals might perform best when accounting for individual variation in growth.

Before incorporating individual variation in growth into a size-transition matrix, it is first necessary to understand how the values within the matrix are determined (Equation 1.3):

$$X_{i,j} = \int_{l_i^-}^{l_i^+} \int_{l_j^-}^{l_j^+} F(\theta, l_j, l_i) dl_j dl_i \quad (1.3)$$

where $X_{i,j}$ represents one of the values within the size transition matrix. The right hand side of Equation 1.3 is a two dimensional integral, with the first integral going over the initial size l_i (which ranges from l_i^- to l_i^+) and the second integral going over the ending size l_j of an individual (which ranges from l_j^- to l_j^+). *F* is the probability density function (pdf) for an individual's growth multiplied by the pdf for the distribution of individuals in the initial size-class. θ is the vector of the parameters that define the underlying growth curve, its variability and the initial size class pdf. A variety of equations is available for the pdf of an individual's growth including the normal, lognormal or gamma. The pdf options for the distribution of individuals within the starting size class include the uniform or truncated normal. Equation 1.3 assumes that all individuals follow the same growth curve with process error. This equation is difficult to solve analytically. A commonly used simplifying assumption is that all individuals in the starting size-class *i* have the same size, the midpoint of the starting size-class *i* ($\bar{l}_i$) (Punt et al., 1997). This reduces Equation 1.3 to:

$$X_{i,j} = \int_{l_j^-}^{l_j^+} F(\theta, \bar{l}_i, l_j) dl_j \quad (1.4)$$

which is more tractable because it requires integration in only one dimension. In addition, the function *F* no longer includes a pdf for the initial size-class. However, Equation 1.4 still assumes that all individuals follow the same underlying growth curve.

Allowing individuals to have unique L_∞ and/or *k* requires more complexity, as shown in the following integral:

$$X_{i,j} = \int_{l_i^-}^{l_i^+} \int_{l_j^-}^{l_j^+} \int_0^\infty \int_0^\infty F(l_i, l_j, L_\infty, k, \theta') dL_\infty dk dl_j dl_i \quad (1.5)$$

where θ' is a vector of parameters that includes the mean and variance for L_∞ ($\bar{L}_\infty$, σ_{L_∞}) and k ($\bar{k}$, σ_k). The function F now includes probability density functions for growth variation, the distribution of individuals within the initial size-class, and the distributions for L_∞ and k. There are several alternative probability distributions for L_∞ and k, such as gamma, lognormal or normal.

The four dimensional integral in Equation 1.5 allows for individual variation in growth but it is difficult to solve analytically. Here, we develop a numerical integration technique to approximate Equation 1.5. We explored using either Simpson's rule or Gaussian quadrature. The performance of each numerical integration method is evaluated by comparing the approximated matrix to a (true) simulated matrix created under the assumption that individuals have their own growth curve.

1.3 Methods

1.3.1 Setting up the integral

We assumed that expected growth follows the von Bertalanffy growth curve (Equation 1.2) and animals vary in terms of k and L_∞. Allowing both k and L_∞ to vary among individuals leads to a problem in Equation 1.5. Specifically, the variation for l_j comes directly from three sources, the initial size, and the values of the two growth parameters (k and L_∞). We can reduce the dimensionality by constraining one parameter to depend upon the value of another parameter. This is accomplished by rearranging Equation 1.2 such that k is a function of L_∞ given l_i and l_j or vice versa. Equations 1.6a and 1.6b show the results of rearranging Equation 1.2:

$$k = -log\left(1 - \frac{l_j - l_i}{L_\infty - l_i}\right) \tag{1.6a}$$

$$L_\infty = l_i + \frac{l_j - l_i}{1 - e^{-k}} \tag{1.6b}$$

Equation 1.6a acts to restrict the integration range for *k*, as the upper and lower bounds now depend on L_∞ given l_i and l_j, and obviates the need to integrate over the ending size-class. The same is true for Equation 1.6b, except that L_∞ depends on *k*. Incorporating the transformation produced by Equations 1.6a and 1.6b into Equation 1.5 leads to:

$$\int_{l_i^-}^{l_i^+} \int_0^\infty \int_{-log\left(1-\frac{l_j^- - l_i}{L_\infty - l_i}\right)}^{-log\left(1-\frac{l_j^+ - l_i}{L_\infty - l_i}\right)} F(l_i, L_\infty, k) dk dL_\infty dl_i \tag{1.7a}$$

$$\int_{l_i^-}^{l_i^+} \int_0^\infty \int_{l_i + \frac{l_j^- - l_i}{1-e^{-k}}}^{l_i + \frac{l_j^+ - l_i}{1-e^{-k}}} F(l_i, L_\infty, k) dL_\infty dk dl_i \tag{1.7b}$$

where the first integral is over the range for the initial size-class, the second integral is over the range for L_∞ in Equation 1.7a (aka Equation-L_∞) and k in Equation 1.7b (aka Equation-*k*), and the third integral is over the range for k in Equation-L_∞ and L_∞ in Equation-*k*. Reducing the number of integrals from four to three simplifies the computational demands of Equation 1.5.

The function $F(l_i, L_\infty, k)$ in Equations-L_∞ and -*k* is the product of the probability density functions for l_i, L_∞ and k. It is assumed here that $l_i \sim U[l_i^-, l_i^+]$, $L_\infty \sim lognormal\left(\overline{L_\infty}, \sigma_{L_\infty}\right)$, and $k \sim lognormal\left(\bar{k}, \sigma_k\right)$. Therefore $F(l_i, L_\infty, k)$ is:

$$F(l_i, L_\infty, k) = \frac{1}{l_i^+ - l_i^-} \frac{1}{\sqrt{2\pi\sigma_{L_\infty}^2}} e^{\frac{(\ln(L_\infty) - \overline{L_\infty})^2}{2\sigma_{L_\infty}^2}} \frac{1}{\sqrt{2\pi\sigma_k^2}} e^{\frac{(\ln(k) - \bar{k})^2}{2\sigma_k^2}} \tag{1.8}$$

The values for $\overline{L_\infty}$, σ_{L_∞}, $\bar{k}$ and σ_k are all given.

1.3.2 Numerical Approximation

We used two numerical integration techniques, Gaussian quadrature and Simpson's rule, to approximate Equations-L_∞ and -*k*. The equation for Gaussian quadrature is:

$$\int_{-1}^{1} f(x) dx \approx \sum_{i=1}^{n} w_i f(x_i) \tag{1.9}$$

where x_i are the evaluation points, w_i are the weights associated with the evaluation points and n is the number of evaluation points. The number of evaluation points is user-specified. The larger the number of evaluation points the higher the accuracy. For this work, we used 16 and 32 evaluation points to explore how accuracy improves with the number of evaluation points. The evaluation points and associated weights are taken from Abramowitz and Stegun, (1972) where the evaluation points are determined based on the roots of a polynomial belonging to a class of orthogonal polynomials (Press et al., 2007). Gaussian quadrature uses evaluation points that are bounded between -1 to 1. When integrating over a range other than -1 to 1, the evaluation points were transformed to cover the desired range using the following equation:

$$y_i = \frac{x_i+1}{2}(y^+ - y^-) + y^- \tag{1.10}$$

where y_i is the evaluation point within the desired range, x_i the original evaluation point between -1 and 1, y^+ is the upper limit of y and y^- is the lower limit of y.

Simpson's rule is defined by the following equation:

$$\int f(x)dx \approx h\left[\frac{1}{3}f_1 + \frac{4}{3}f_2 + \frac{2}{3}f_3 + \cdots + \frac{4}{3}f_{N-1} + \frac{1}{3}f_N\right] \tag{1.11}$$

where h is the step size and N is the number of steps. Simpson's rule uses evaluation points that are evenly spaced over the desired range. Gaussian quadrature usually leads to higher accuracy with fewer function calls than Simpson's rule. We evaluate both numerical integration methods to determine whether the additional complexity of Gaussian quadrature is justified given the improvement in the approximation.

Both numerical approximation methods require knowledge on the bounds of the space explored. This is a problem for the integral over the growth parameter distribution that goes from zero to infinity in Equation-L_∞ and -k. One possible solution is to use an educated guess for the

bounds. Another is to let the bounds depend on the variance term from the growth parameter distribution. We determined bounds for L_∞ in Equation-L_∞ and k in Equation-k as follows:

$$\overline{L_\infty} +/- (s\sigma_{L_\infty}) \quad (1.12a)$$

$$\bar{k} +/- (s\sigma_k) \quad (1.12b)$$

where s is a variance multiplier.

1.3.3 Specifying the true size-transition matrix

The approximated matrices produced by the two numerical integration methods are compared to a "true" matrix that assumes every individual has their own growth curve. The "true" matrix is constructed through simulation. Randomly drawn values for L_∞ (lognormal distribution [$\bar{L}_\infty$, σ_{L_∞}]), k (lognormal distribution [$\bar{k}$, σ_k]) and l_i (uniform distribution [l_i^-, l_i^+]) are inserted into Equation 1.2 to determine l_j. This process repeats 30,000,000 times, with each new l_j recorded. The proportion of l_j's that fall within each size-class determines the values within the "true" size-transition matrix. The large number of replicates ensures that the "true" matrix values are accurate to at least the first three decimal places.

1.3.4 Evaluation

The purpose of this chapter is to determine whether this new method for constructing size-transition matrices is sufficiently accurate under a variety of scenarios. Therefore, we used five sets of values for $\bar{L}_\infty$, σ_{L_∞}, $\bar{k}$, and σ_k (Table 1.1) to construct the "true" and approximated matrices. The first parameter set was based on the size-transition matrix for Eastern Bering Sea (EBS) Tanner crab (*Chionoecetes bairdi*) males estimated in the 2015 stock assessment (Stockhausen 2015). This matrix has 37 size-classes with a first size of 25 mm and an upper limit for the final size-class of 210 mm. The width of each size-class in the assessment was 5 mm. We determined the "true" values for $\bar{L}_\infty$, σ_{L_∞}, $\bar{k}$, and σ_k by minimizing the difference between the matrix from the

assessment and that inferred by simulation using the von Bertalanffy parameters. The remaining four sets of parameter values were either an increase or decrease to the mean and variance parameters obtained by fitting the size-transition matrix from the assessment by 20%. Figure 1.1 provides a visual representation of the growth variation in each parameter set. The aim is to evaluate whether the numerical integration methods perform adequately under a variety of situations.

Each set of parameter values produced one "true" and six approximated size-transition matrices (two using 16-point Gaussian quadrature, where the range of integration for L_∞ depends on k and vice versa, two using 32-point Gaussian quadrature where the range of integration for L_∞ depends on k and vice versa and two using Simpson's rule, where the range of integration for L_∞ depends on k and vice versa). We explored the potential impact of the width of the size-classes on the accuracy of the numerical integration procedure by creating "true" and the approximated matrices with size-class widths of 5 and 10 mm. The first size and upper limit of the final size-class matched those used for the EBS Tanner crabs. We explored setting the variance multiplier value in Equations 1.12a and 1.12b to either 5, 15 or 20 to determine which was optimal. In total, we created 10 (= 5x2) "true" size-transition matrices that varied in growth parameter value and size-class width. Each "true" matrix was compared to 18(= 6x3) approximated size-transition matrices that differed in terms of the multiplier value and the numerical approximation method.

We used two methods to evaluate the accuracy of each approximated matrix. The first involved comparing values within an approximated matrix with those in the "true" matrix by subtracting the difference between the two matrices. Only the values located in the upper triangular section of size-transition matrices were used since we assume that individuals cannot shrink. Therefore, the values located in the lower triangular part of both the approximated and "true" matrices are all

zero. The relative error is not used because it could be biased by large differences in small true values. We summarized these differences using boxplots and the sum of the absolute errors.

The second method of evaluation compared the equilibrium size distribution for a hypothetical population with a mortality rate of $0.2yr^{-1}$, which is similar to the natural mortality rate assumed for EBS Tanner crabs ($0.23yr^{-1}$) (Stockhausen 2015), when using either the "true" or approximated size-transition matrix. Altering the natural mortality made the results more general since the population is hypothetical. We assume that all recruits enter this population via the first size-class. The equilibrium size structure was determined by taking the right eigenvector of the product of the size-transition matrix and a matrix with the exponentiated negative of the mortality rate on the diagonals:

$$N_i = \begin{cases} \frac{1}{(1 - X_{i,i}e^{-M})} & if\ i = 1 \\ \frac{\sum_{j=1}^{i-1} X_{j,i}N_j e^{-M}}{(1 - X_{i,i}e^{-M})} & if\ i \neq 1 \end{cases} \quad (1.13)$$

where N_i is the number of individuals in the equilibrium size structure for size class i, and M is the mortality rate. Each equilibrium size structure vector was normalized to sum to one to allow for comparisons. The differences between the equilibrium size structures produced using the approximated matrices and the "true" matrix are used to compare performances. This second approach evaluates the net impact of numerical errors in matrix values, which can accumulate through Equation 1.13.

1.4 Results

Figure 1.2 shows performance of the approximation methods for all scenarios when the variance multiplier is 15. The degree of accuracy varies with scenario and approximation method. Three methods clearly underperformed relative to the others: 16-point Gaussian quadrature using

Equation-*k* and Simpson's rule using Equation-L_∞ or -*k*. Of these, 16-point Gaussian quadrature using Equation-*k* had the poorest performance. There was a noticeable difference between having size class widths of 5mm and 10mm for the three poorest performing approximation methods. Having a size-class width of 10mm resulted in greater accuracy than a size-class width of 5mm. The other three approximation methods are accurate to at least three decimal places for all scenarios except for 16-point Gaussian quadrature using Equation-L_∞, which has accuracy to only two decimal places for six of the scenarios (Figure 1.3). 32-point Gaussian quadrature had the best performance in relation to matrix values for all scenarios, regardless of whether it was based on Equation-L_∞ or -*k*.

A similar result appears in Table 1.2, which shows the sum of the absolute value of the values used to create Figure 1.2. The smallest sums for all scenarios occurred with 32-point Gaussian quadrature using either Equation-L_∞ or –*k*. It should be noted that the sums are very low and the "true" matrix accuracy is only to three decimal places. Therefore, Table 1.3 show the same results as Table 1.2, except for a different true matrix simulated using the same parameter values that constructed the "true" matrix used for Table 1.2. The intention is to explore whether the errors in the fourth decimal place or lower influenced the results shown in Table 1.2. In Table 1.3, 32-point Gaussian quadrature outperformed the other approximation methods for all scenarios except one (parameter set 2 with a size-class width of 10mm) where 16-point Gaussian quadrature using Equation-*k* performed the best.

Figure 1.4 shows example heat plots of the differences between the true and approximated matrices. The heat plot showed a random dispersion of small errors across the size-transition matrix when the approximated method had accuracy to three decimal places (32-point Gaussian quadrature using Equation-k). However, when the accuracy was less than three decimal places, a

pattern appeared based on whether Equation-L_∞ or -*k* was used. Equation-*k* resulted in larger error values (relative to the remaining absolute error values) at smaller initial size-classes. Equation-L_∞ showed the reverse trend, with larger error values at higher initial size-classes. This trend is not impacted by size-class widths and appeared in all scenarios except scenario 2.

The results of the comparison of the equilibrium size structures produced from the "true" and approximated matrices when the variance multiplier is 15 (Figure 1.5) mirror those shown in Figure 1.2. Approximated matrices that had errors that exceeded the dashed lines in Figure 1.2, 16-point Gaussian quadrature using Equation-*k* and Simpson's rule using Equation-*k*, also exceeded the dashed lines in Figure 1.5. The same impact of the width of the size-classes shown in Figure 1.2 on these two methods also appears in Figure 1.5. Evaluation of the remaining four methods requires a zoomed in version of Figure 1.5 (Figure 1.6). Of the remaining four methods, only 16-point Gaussian quadrature using Equation-L_∞ exceeds the new dashed lines. There is no apparent impact from the width of the size-classes on the remaining four methods.

The value of the variance multiplier has a major impact on the accuracy of the approximated matrices. Setting the variance multiplier to 5 (Figure 1.7) led to the accuracy of the approximated matrix depending more on the parameter set and equation. The poorest performing approximation methods are those that use Equation-L_∞, with little distinction between Simpson's rule and Gaussian quadrature. In addition, sets with low mean growth parameters and high variance resulted in poorer performances. There was no apparent impact from the width of the size-classes. Setting the variance multiplier to 20 (Figure 1.8) causes performance to depend on approximation method. The poorest performing methods were 16-point Gaussian quadrature and the two approximation methods based on Simpson's rule. There was little distinction between parameter sets. However, increasing the size-class width from 5mm to 10mm improved performance.

1.4.1 Simpson's Rule vs Gaussian Quadrature

The level of accuracy in any numerical integration method depends on the number of evaluation points used. In the scenarios in Figures 1.2-1.8, Simpson's rule used 50 evaluation points for each growth parameter distribution and 10 for the initial size-class distribution. To better understand the difference between the numerical approximation methods, we considered two modified Simpson's Rules in which the number of evaluation points for Simpson's Rule was increased. One modification increased the number of evaluation points for the initial size-class distribution to 50. The other increased the number of evaluation points for the parameter distributions to 100 while keeping the number of points to represent the initial size-class distribution at 10. Figure 1.9 shows boxplots comparing 32-point Gaussian quadrature to these alternative implementations of Simpson's Rule for all scenarios when the variance multiplier is 15. Increasing the number of evaluation points for the initial size class distribution did not markedly improve the performance of Simpson's rule. There was a better performance when the number of evaluation points for the growth parameter distributions was increased. However, this increase in accuracy did not lead to Simpson's rule performing better than 32-point Gaussian quadrature.

1.5 Discussion

All numerical approximation methods perform adequately. However, 32-point Gaussian quadrature using either Equation-L_∞ or -k performed best. Larger discrepancies in size-transition matrix values resulted in larger differences for equilibrium size-structure, as expected. However, the error in equilibrium size-structure was only noticeable when the accuracy of the values of the matrix was less than two decimal places. For example, in Figure 1.2, Simpson's rule using Equation-L_∞ failed to achieve accuracy of three decimal places, but the calculated equilibrium size

structure was always accurate to three decimal places. Nevertheless, neither 16-point Gaussian quadrature using Equation-*k* nor Simpson's rule using Equation-*k* should be used. Table 1.3 shows that 16-point Gaussian quadrature using Equation-*k* performing the best for parameter set 2 with a size-class width of 10mm. This is most likely due to the low growth variability in parameter set 2, as seen in Figure 1.1, and not the performance of 16-point Gaussian quadrature using Equation-*k*. The remaining four approximation methods are viable options, with the 32-point Gaussian quadrature using either equations being the best overall.

All numerical approximation methods require four parameters to be known: mean and variance for the growth parameters L_∞ and k. In theory, these parameters can be estimated using tag-recapture data. However, the variance multiplier cannot be estimated and must be fixed. The three variance multiplier values in this paper explore the impact of having small to large integration bounds for growth parameters bounded between zero and infinity. Setting the multiplier to a small value can lead to the approximation range being too narrow to adequately cover the entire space. This dictates performance, removing the distinction between numerical approximation methods. This is apparent in Figure 1.7 since all methods that use the same equation perform roughly the same for scenario 2. Setting the multiplier to a larger value can result in the approximation range being too wide, also causing the numerical approximation method to not accurately explore the space being integrated. Figure 1.8 shows that all methods are impacted by the wide approximation ranges for all scenarios except for the 32-point Gaussian quadrature using either Equation-L_∞ or -*k*. This exception occurred because the approximation is based on more function evaluations and the generally better performance of Gaussian quadrature. Based on these results, it is best to set the multiplier to around 15.

In general, Gaussian quadrature has greater accuracy than Simpson's rule with fewer points. Increasing the number of evaluation points for either Gaussian quadrature or Simpson's rule improves performance. However, an increase of one evaluation point for Gaussian quadrature is not proportional to a one evaluation point increase for Simpson's rule. Figure 1.9 shows that a very large increase in evaluation points for the growth parameter distributions is needed for Simpson's rule to match the performance of the 32-point Gaussian quadrature. Increasing the number of evaluation points for the initial size-class distribution does not improve the performance of Simpson's rule because that distribution is assumed to be uniform, which is a simple distribution to approximate. Although it is possible for Simpson's rule to match Gaussian quadrature in performance, the increase in evaluation points also substantially extends the run time for Simpson's rule. The 32-point Gaussian quadrature has a faster run time and better accuracy and thus is a better option.

The two poorest performing methods, 16-point Gaussian quadrature using Equation-*k* and Simpson's rule using Equation-*k*, both use Equation-*k*. Increasing the number of evaluation points from 50 to 100 for the growth parameter distributions for Simpson's rule appears to remove the discrepancy between Equations-L_∞ and -*k* (Figure 1.9). This also is the case for Gaussian quadrature when the number of evaluation points increases from 16 to 32. The performances of Equations-L_∞ versus -*k* differ due to the range over which the growth parameters are approximated when the number of evaluation points is low. If the approximation range is much larger than the "true" distribution, the space cannot be adequately explored with a small number of evaluation points. In these situations, accuracy can be improved by either increasing the number of evaluation points or narrowing the approximation range to more precisely match the "true" distribution. The approximation range for Equation-L_∞ is narrower than that for Equation-*k* because Equation-L_∞

is in log-space. It is not possible for the value inside the log to become negative since it is undefined. Therefore, the algorithm must narrow the approximation range to ensure this does not happen. Increasing the number of evaluation points cancels out the benefit of Equation-L_∞ having a narrower approximation range.

It is apparent that performance of the numerical integration method is impacted by using Equation-*k* or -L_∞. When accuracy is low, Equation-*k* leads to larger errors (relative to the other errors in the matrix) in the smaller size classes while Equation-L_∞ does the reverse (Figure 1.4). The cause for this trend is not clear. However, these "large errors" are still small. Thus, while it is worthwhile to acknowledge this effect, the impact is minimal.

This work shows that it is possible to approximate, to at least a three decimal places of accuracy, a size-transition matrix constructed using individual variation in growth. The numerical approximation method requires knowing four parameters ($\overline{L_\infty}$, σ_{L_∞}, $\bar{k}$ and σ_k). The next steps are to compare this method for constructing size-transition matrices with other methods and see if it is a better option.

Table 1.1. The five sets of parameter values used to explore the performance of the methods for constructing size-transition matrices. The mean values are in normal space while the variance values are in log space.

	Parameter Values			
	$\bar{L}_\infty$	σ_{L_∞}	$\bar{k}$	σ_k
Set 1	128 mm	0.264	0.064 year^{-1}	0.72
Set 2	128 mm	0.176	0.064 year^{-1}	0.48
Set 3	160 mm	0.22	0.08 year^{-1}	0.60
Set 4	192 mm	0.264	0.096 year^{-1}	0.72
Set 5	192 mm	0.176	0.096 year^{-1}	0.48

Table 1.2. The sum of the absolute values of the differences between the "true" and approximated matrices for all scenarios. The numerical approximation method is identified using two symbols. The first indicates the use of either Gaussian quadrature (G32 for the 32-point method and G16 for the 16-point method) or Simpson's rule (S), while the second indicates the use of either Equation-L_∞ (Linf) or Equation-k (k). The bold numeric values under Approximation Method are the lowest sum for that row. The remaining values under Approximation Method are the ratios between that particular row and column value and the lowest value in that row.

Run		**Approximation Method**					
Parameter set	**Size-Class Width**	**G32-Linf**	**G32-K**	**G16-Linf**	**G16-k**	**S-Linf**	**S-k**
1	**5**	**0.00956**	1.013	1.365	17.236	21.0758	17.682
	10	1.359	**0.00345**	1.653	8.792	18.747	10.974
2	**5**	2.220	**0.00536**	4.783	5.356	47.408	4.372
	10	5.137	**0.00158**	42.752	1.364	22.442	15.613
3	**5**	1.054	**0.00986**	3.336	44.743	14.552	28.916
	10	1.421	**0.00392**	2.840	6.431	10.838	23.650
4	**5**	1.098	**0.01833**	2.271	56.138	6.845	52.940
	10	**0.00493**	1.254	1.850	21.341	7.269	29.734
5	**5**	**0.00964**	1.014	5.760	57.891	6.537	20.365
	10	1.645	**0.00390**	3.603	12.569	5.730	11.104

Table 1.3. As for Table 1.2, but with a separate "true" matrix simulated using the same parameter values used to construct the "true" matrix in Table 1.2.

Run		Approximation Method					
Parameter set	**Size-Class Width**	**G32-Linf**	**G32-K**	**G16-Linf**	**G16-k**	**S-Linf**	**S-k**
1	**5**	**0.00918**	1.053	1.493	17.796	21.861	18.536
	10	1.250	**0.00313**	1.674	9.754	20.471	12.057
2	**5**	2.485	**0.00492**	4.959	5.768	51.502	4.553
	10	4.865	1.178	36.728	**0.00184**	19.378	12.881
3	**5**	1.024	**0.01028**	3.084	42.982	13.875	27.680
	10	1.395	**0.00323**	3.489	7.679	12.870	28.834
4	**5**	1.061	**0.01858**	2.279	55.357	6.778	52.230
	10	**0.00577**	1.078	1.598	18.192	6.312	25.769
5	**5**	1.056	**0.01203**	4.414	46.207	5.287	16.284
	10	1.896	**0.00332**	4.318	14.836	6.576	12.905

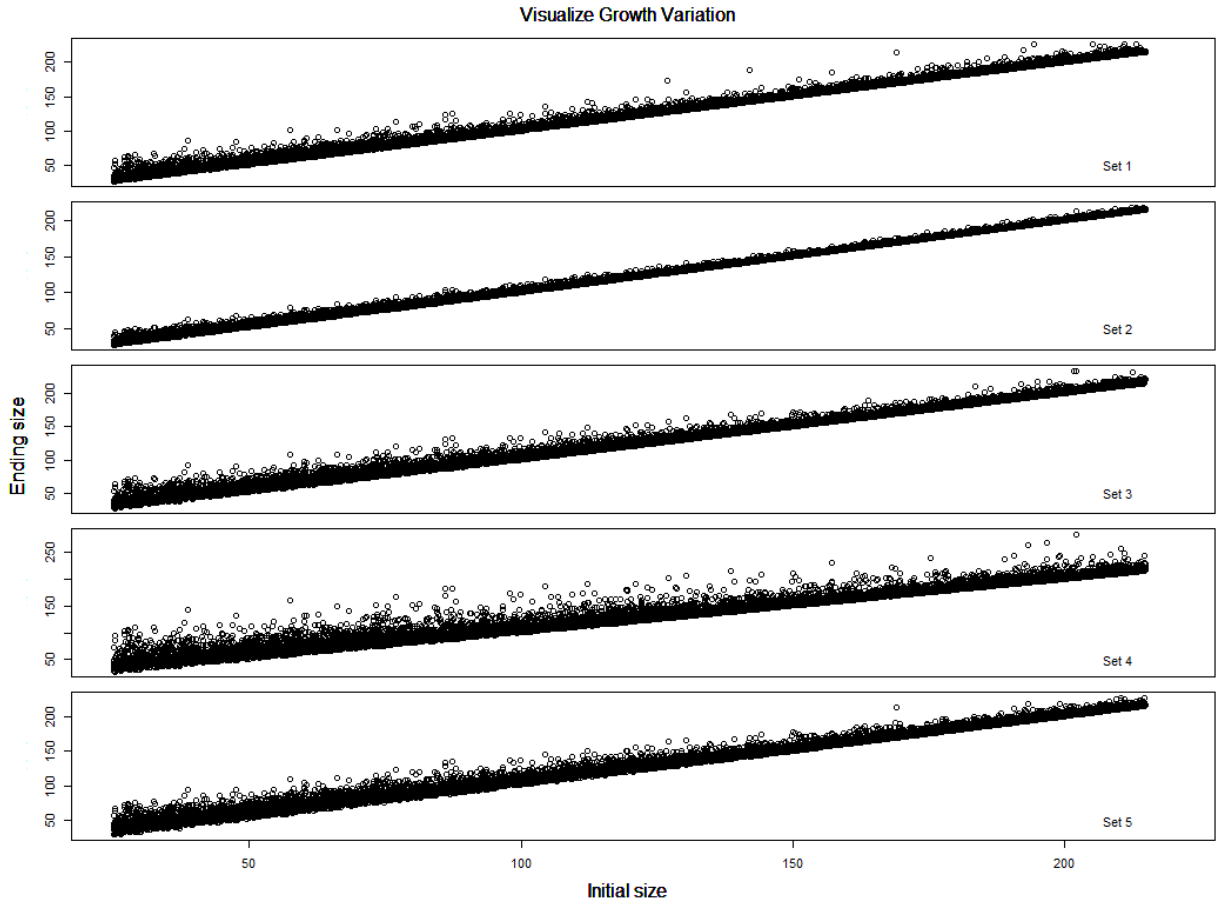

Figure 1.1. The variation in growth produced from the parameters sets listed in Table 1. The label in the bottom right corner represents the parameter set used. The x-axis is the initial size and the y-axis is the ending size after one time step. Each dot represents a single individual with 10,000 individuals in each plot.

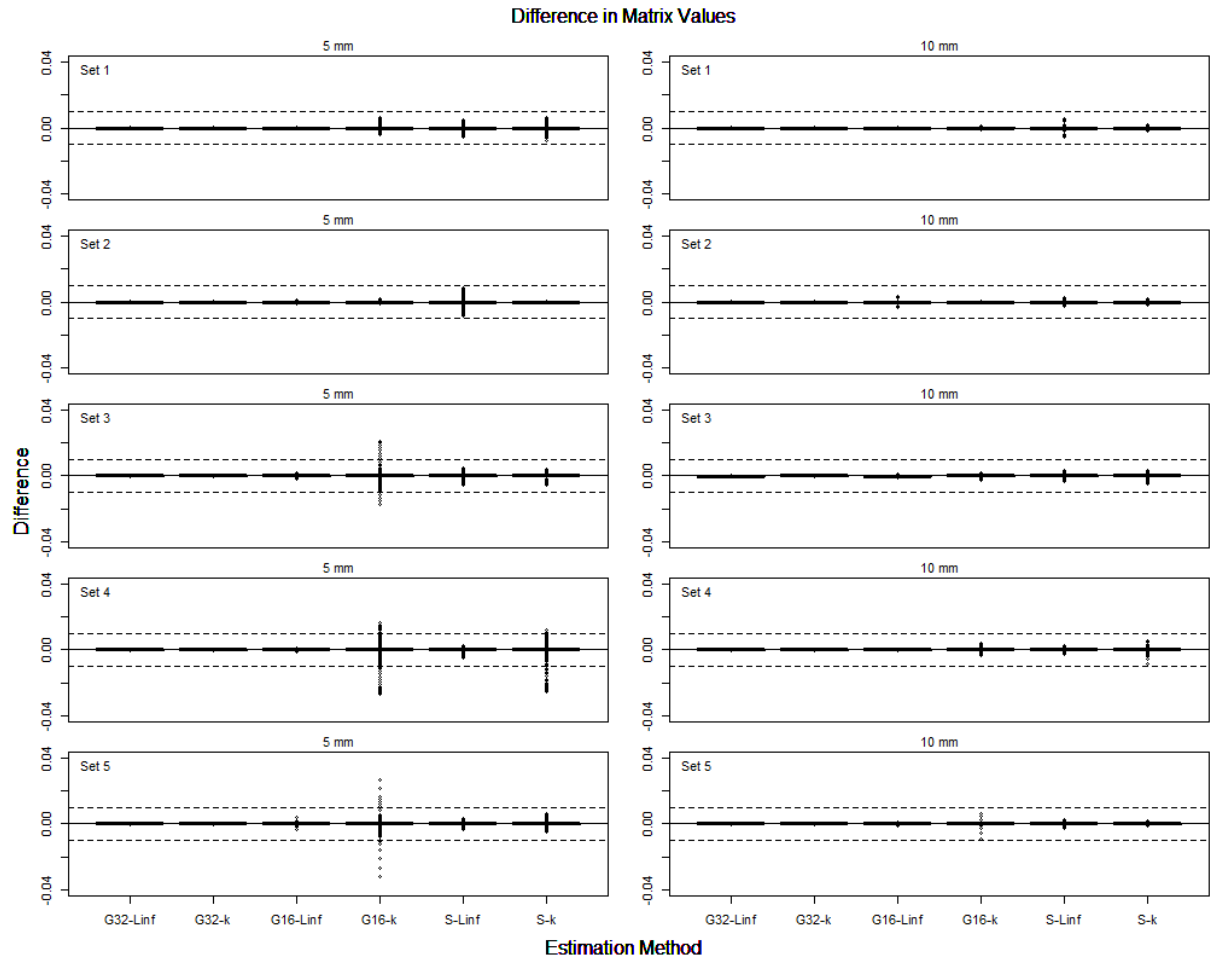

Figure 1.2. Boxplots of the differences between the values of the "true" and approximated matrices for each scenario. The title above each graph represents the width of each size class. The label in the top left corner represents the parameter set used. The x-axis indicates the numerical approximation method used. The first symbol defining the method represents either Gaussian quadrature (G32 for the 32-point method and G16 for the 16-point method) or Simpson's rule (S), while the second symbol represents either Equation-L_∞ (Linf) or Equation-k (k). The solid line represents the zero line. The dashed lines represent the bounds for two decimal places of accuracy.

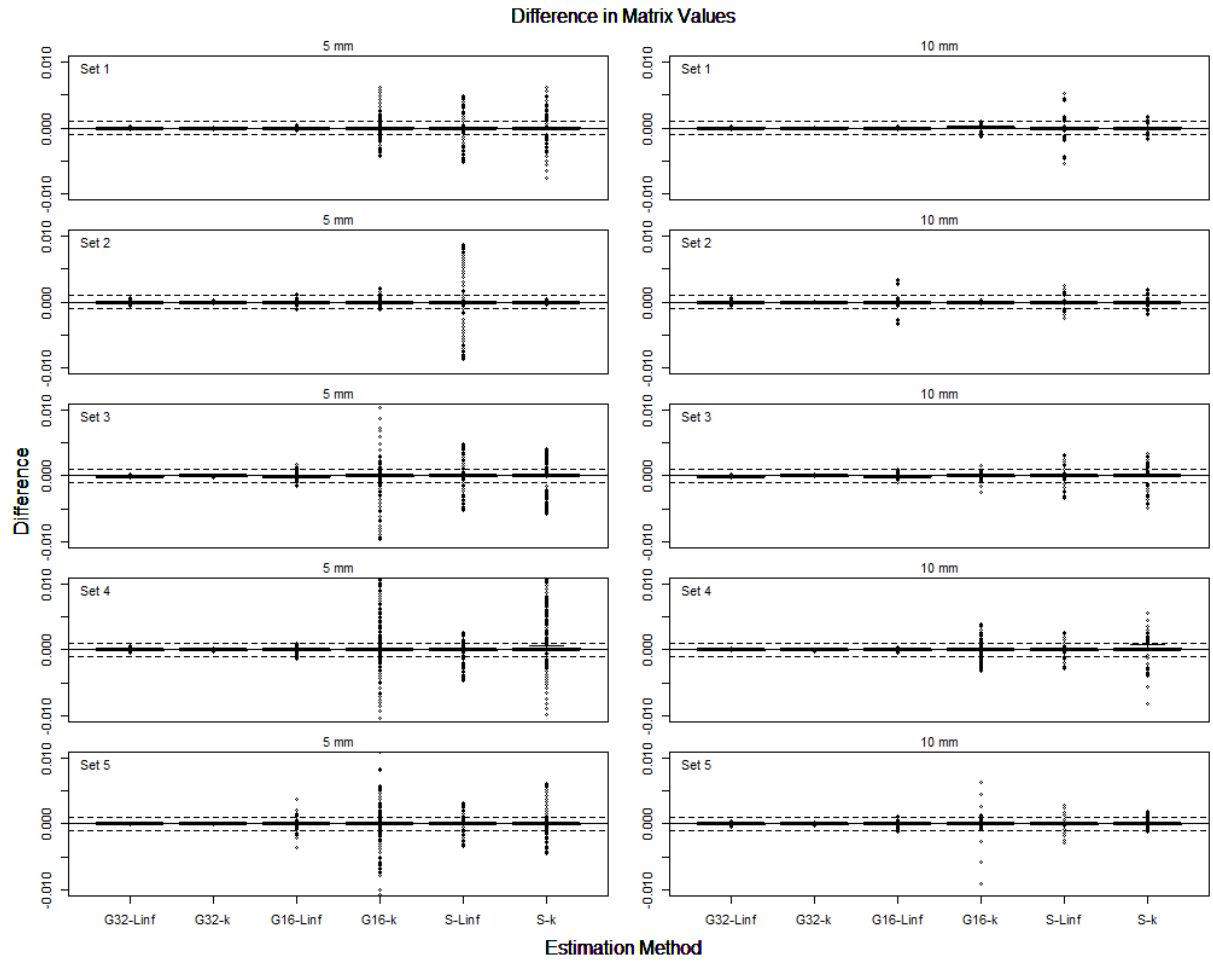

Figure 1.3. As for Figure 1.2, zoomed for increased clarity. The dashed lines represent the bounds for three decimal places of accuracy.

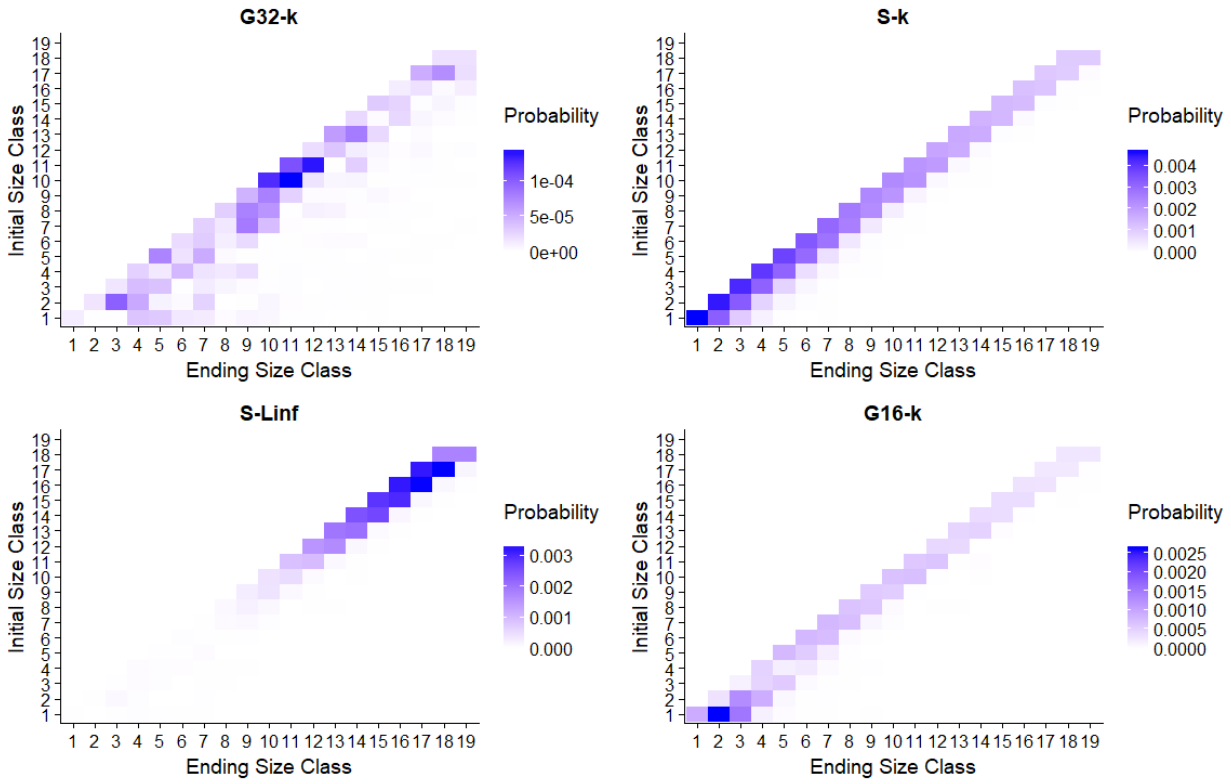

Figure 1.4. Heat plots of the absolute error between the true and approximate matrices for several approximation methods; 32-point Gaussian quadrature using Equation-k (G32-k), Simpson's Rule using Equation-k (S-k) and -L_∞ (S-Linf) and 16-point Gaussian quadrature using Equation-k (G16-k). The approximation method is shown above each graph. The true and approximated matrices are from scenario 3 with a size-class width 0f 10mm.

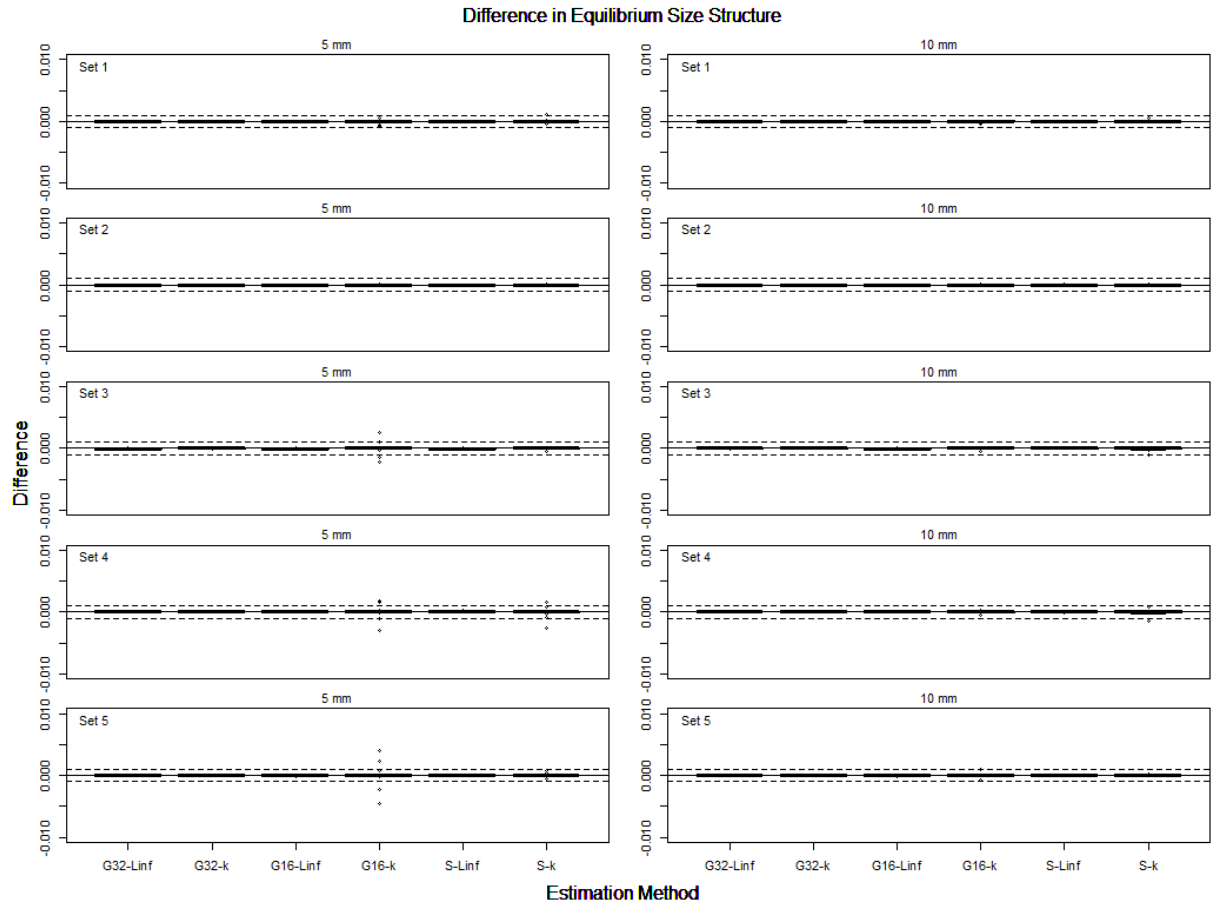

Figure 1.5. Boxplot of the difference between the equilibrium size structures produced from the "true" and approximated matrices for each scenario. The title above each graph represents the width of each size class. The label in the top left corner represents the parameter set used. The x-axis indicates the numerical approximation method used. The first symbol defining the method represents either Gaussian quadrature (G32 for the 32-point method and G16 for the 16-point method) or Simpson's rule (S), while the second symbol represents either Equation-L_∞ (Linf) or Equation-k (k). The solid line represents the zero line. The dashed lines represent the bounds for three decimal places of accuracy.

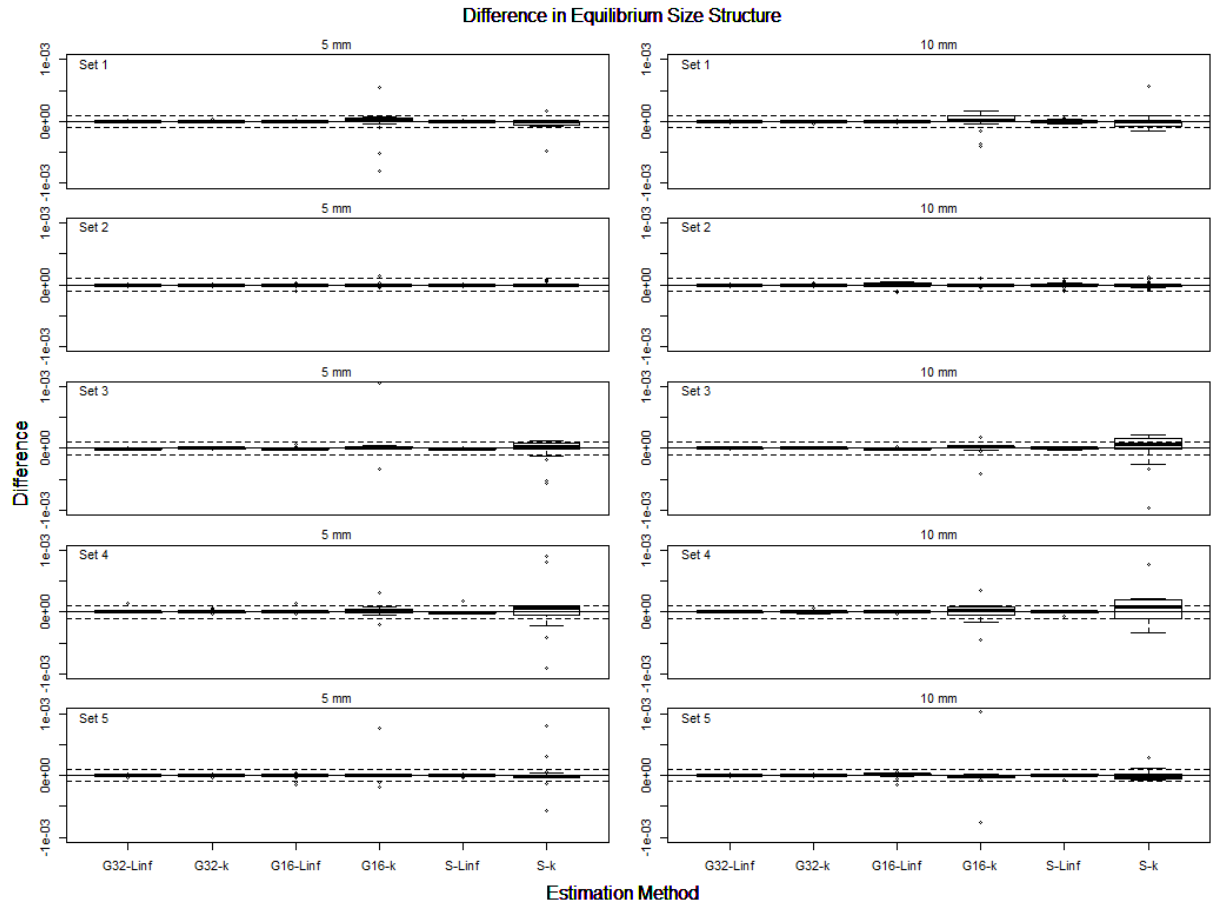

Figure 1.6. As for Figure 1.5, zoomed for increased clarity. The dashed lines represent the bounds for four decimal places of accuracy.

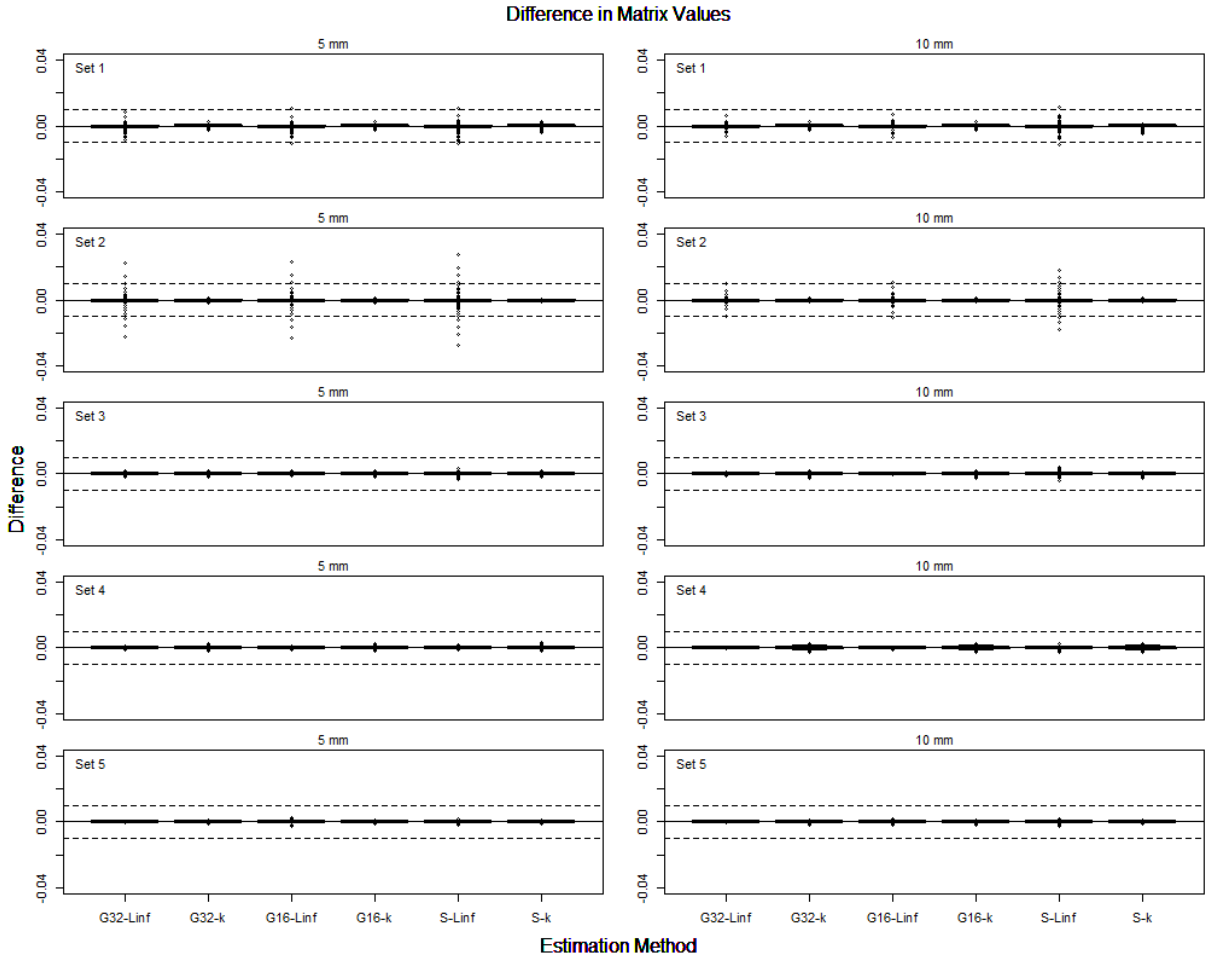

Figure 1.7. As for Figure 1.2, except the variance multiplier is 5.

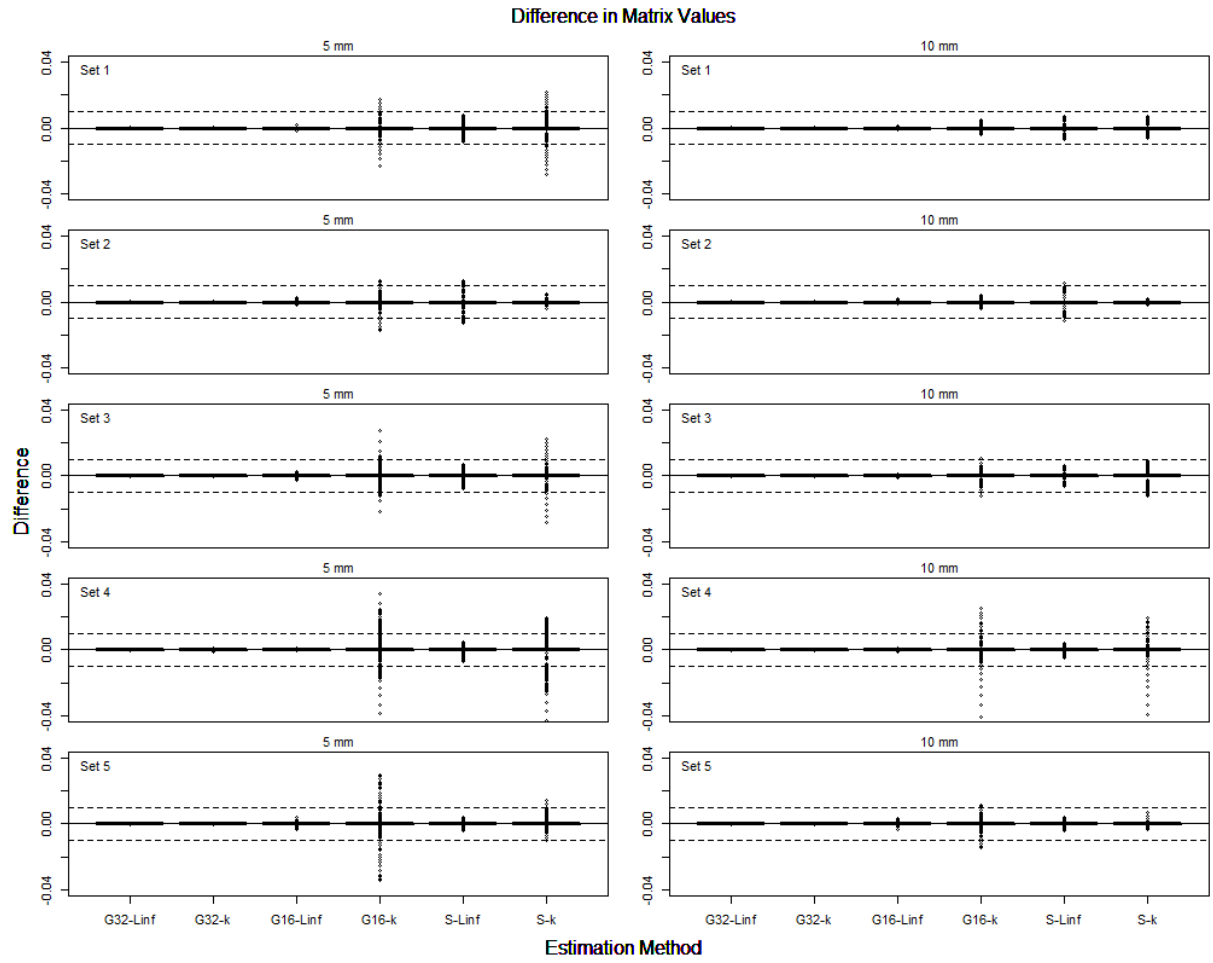

Figure 1.8. As for Figure 1.2, except the variance multiplier is 20.

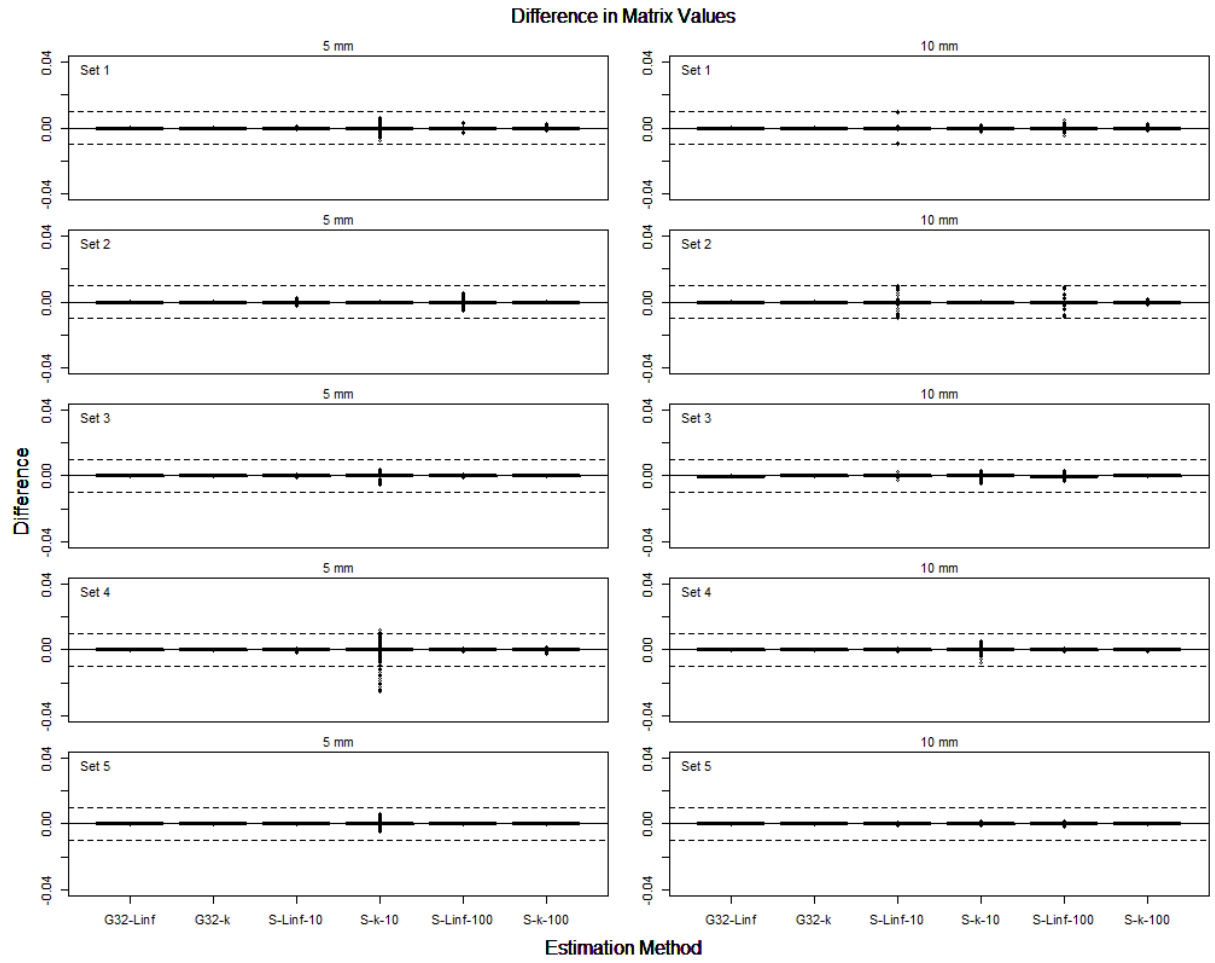

Figure 1.9. A modified version of Figure 1.2 with different numerical approximation methods. The x-axis indicates the numerical approximation method used. The first symbol defining the method represents either Gaussian quadrature (G32 for the 32-point method) or Simpson's rule (S), the second represents either Equation-L_∞ (Linf) or Equation-k (k) and the third represents the modified number of evaluation points for Simpson's Rule (50 points for initial size class distribution or 100 for growth parameter distributions). The solid line represents the zero line. The dashed lines represent the bounds for two decimal places of accuracy.

Chapter 2 : There is No Best Method for Constructing Size-Transition Matrices for Size-Structured Stock Assessments

2.1 Abstract

Stock assessment methods for many invertebrate stocks, including crab stocks in the Bering Sea of Alaska, rely on size-structured population dynamics models. A key component of these models is the size-transition matrix, which specifies the probability of growing from one size-class to another after a certain period of time. Size-transition matrices can be defined using three parameters, the growth rate (k), asymptotic size (L_∞), and variability in the size increment. Most assessments use mark-recapture data to estimate these parameters and assume that all individuals follow the same growth curve, but this can lead to biased estimates of growth parameters. We compared three approaches: the traditional approach, the platoon method, and a numerical integration method that allows k, L_∞ or both to vary among individuals, under a variety of scenarios using simulated data based on golden king crabs (*Lithodes aequispinus*) in the Aleutian Islands region of Alaska. No estimation method performed best for all scenarios. The number of size-classes in the size-transition matrix and how the data are generated heavily dictate performance. However, we recommend the numerical integration method that allows L_∞ to vary among individuals and smaller size-class widths.

2.2 Introduction

Crustaceans are a highly valuable fisheries resource. For example, the crab fisheries in the Bering Sea and Aleutian Islands region of Alaska produced an estimated gross ex-vessel revenue of $259.3 million in 2016 (Garber-Yonts and Lee, 2018). Crab fisheries, including those in the Bering Sea, use stock assessment models to support management. These models estimate the size and trend of a population as well as reference points, status relative to reference points, and provide

information needed to apply harvest control rules. Finfish assessments typically use some version of an age-structured population dynamics model, which rely on age-composition data. However, the molting of crustacean exoskeletons on an irregular basis removes potential indicators of age, making it difficult to obtain such data.

Stock assessment of hard-to-age species can be based on size-structured population dynamics models. These models track groups within the population by size instead of age, which allow population dynamic processes, such as mortality, to vary with size. A wide variety of data types are used to estimation parameters (e.g., size-composition of the annual catches, and tag-recapture data) (Punt *et al*., 2013). Many fisheries use size-structured stock assessment methods for assessment (e.g., crab stocks in the Bering Sea and Aleutian Islands region of Alaska, American lobster (*Homarus americanus*) in the Atlantic, rainbow abalone (*Haliotis iris*) off New Zealand, and tiger and endeavor prawns off Australia) (Szuwalski and Turnock, 2016; ASMFC, 2015; Marsh and Fu, 2017; Buckworth *et al.*, 2015).

A central aspect of a size-structured population dynamics model is the size-transition matrix ($\boldsymbol{X}$). This matrix specifies the probability of an individual growing from one size-class to another after a certain period of time. Equation 2.1 is an example of a size-transition matrix with four size-classes:

$$\boldsymbol{X} = \begin{bmatrix} X_{1,1} & X_{1,2} & X_{13} & X_{1,4} \\ 0 & X_{2,2} & X_{2,3} & X_{2,4} \\ 0 & 0 & X_{3,3} & X_{3,4} \\ 0 & 0 & 0 & X_{4,4} \end{bmatrix} \tag{2.1}$$

where $X_{i,j}$ represents the probability of an individual transitioning from size-class *i* to size-class *j* after one time step. In Equation 2.1, the rows represent the starting size-class and the columns the ending size-class. The lower triangle portion of the matrix is all zero since we assume individuals cannot shrink. The modeler chooses the range of sizes defining each size-class. Each row of the

size-transition matrix sums to one, since individuals must fall into one of the size-classes after growth.

The values within the size-transition matrix are determined using an underlying growth function. A variety of growth functions are available (e.g. Punt *et al.*, 2016; Francis, 1988; Turnock and Rugolo, 2013). Stock assessments commonly use the von Bertalanffy growth curve (Equation 2.2) (Punt *et al.*, 2016):

$$l_2 = l_1 + (L_\infty - l_1)(1 - e^{-kt}) \tag{2.2}$$

where l_1 is the initial size, L_∞ is the asymptotic size, k is the growth rate, t is the number of time steps and l_2 is the ending size after t time steps, i.e. expected size after growth is a linear function of current size. In this paper, the von Bertalanffy growth curve is always the underlying growth curve, although the estimation methods described below extend naturally to other functional forms between current and eventual size. The following function determines the values within the size-transition matrix:

$$X_{i,j} = \int_{l_i^-}^{l_i^+} \int_{l_j^-}^{l_j^+} F\left(l_{1i}, l_{2j}, L_\infty, k, \theta'\right) \mathrm{d}l_{2j} \mathrm{d}l_{1i} \tag{2.3}$$

where i represents the initial size-class (which ranges from l_i^- to l_i^+), l_{1i} is the initial size within size-class i, j represents the ending size-class (which ranges from l_j^- to l_j^+), l_{2j} is the resulting size within size-class j, F is the product of the probability distribution functions (pdfs) for growth variation and the distribution of individuals within the initial size-class, and θ' is a vector of parameters that define the pdfs. The parameters needed to construct the size-transition matrix are estimated using tag-recapture data, which consists of three components for each individual: the size-at-release (size when the individual is first tagged), the size-at-recapture, and the time-at-liberty (how long an individual was in the wild after being tagged).

A common issue when estimating size-transition matrices is how to account for individual variation in growth. The traditional method assumes that the underlying growth curve represents the average growth of the population, with deviation from the curve due to process error. Realistically, each individual in a population follows their own growth trajectory. Punt *et al.* (2016) and Sainsbury (1980) showed that assuming there is only a single growth curve when there is individual variation in growth could bias estimates of growth parameters.

A second ('platoon') method addresses the assumption that all individuals roughly follow the same growth curve by dividing the population into a predetermined number of groups (e.g., 3 or 5), where each group has its own size-transition matrix. Each size-transition matrix has its own growth curve by having either k or L_∞ (with expected growth based on the von Bertalanffy equation) vary among groups. There is still process error around each growth curve to account for individual variation within each platoon. This method has been used to model finfish growth (e.g., Methot and Wetzel, 2013; Punt *et al.*, 2001, 2017). It allows for multiple growth curves within the population, potentially improving the ability to account for individual variation in growth. However, within each group, individuals still follow one growth curve.

A third method to account for individual variation in growth is to allow individuals to have their own growth curves by allowing either k, L_∞, or both to vary for each individual. This is accomplished by adding additional integrals to Equation 2.3 depending on how many parameters vary. For example, if k and L_∞ vary Equation 2.3 becomes:

$$X_{i,j} = \int_{l_i^-}^{l_i^+} \int_{l_j^-}^{l_j^+} \int_0^\infty \int_0^\infty F\left(l_{1i}, l_{2j}, L_\infty, k, \theta'\right) \mathrm{d}L_\infty \mathrm{d}k \mathrm{d}l_{2j} \mathrm{d}l_{1i} \quad (2.4)$$

where F is now the product of the pdfs for growth variation, the distribution of individuals within the initial size-class and the growth parameter distributions. Several methods exist that allow either k or L_∞, but not both to vary among individuals when constructing a size-transition matrix (e.g.,

Troynikov, 1988; Wang *et al.*, 1995). We developed a new method that allows *k*, L_∞, or both to vary among individuals by constructing the size-transition matrix using numerical integration.

For this paper, we conducted a simulation experiment to evaluate the performances of the three methods for constructing size-transition matrices. Past work also used simulation to evaluate methods for constructing size-transition matrices (e.g., Punt *et al.*, 2009; Wang *et al.*, 1995). Data generation was based on golden king crabs (*Lithodes aequispinus*) in the Aleutian Islands region of Alaska. We evaluated the performance of each method by estimating size-transition matrices using these methods and comparing them to a "true" matrix using the differences between the "true" and estimated matrix values and equilibrium size-structure.

2.3 Estimation methods

2.3.1 Method 1: Traditional

The first (traditional) method assumes that every individual in the population follows the same growth function and thus has the same growth parameter values. This assumption uses Equation 2.3 to determine the values within the size-transition matrix, but it can be difficult to solve. A common practice is to assume that, before growing, all individuals have a size equal to the midpoint of the initial size-class (Punt *et al.*, 1997). This modification changes Equation 2.3 to:

$$X_{i,j} = \int_{l_j^-}^{l_j^+} F\left(l_{2j}, \bar{l}_i, \theta'\right) \mathrm{d}l_{2j} \tag{2.5}$$

where $\bar{l}_i$ is the midpoint of the size-class *i*. The function *F* is the pdf for growth variation, with options such as the normal, lognormal, and the gamma distributions. This paper assumes a gamma distribution for the pdf because it is used as the error structure for assessments of golden king crab (Siddeek *et al.*, 2008). Therefore $F\left(l_{2j}, \bar{l}_i, \theta'\right)$ is:

$$F\left(l_{2j}, \bar{l}_i, \theta'\right) = \frac{(l_{2j} - \bar{l}_i)^{\alpha_i - 1}\, e^{-\beta_i (l_{2j} - \bar{l}_i)}}{\Gamma(\alpha_i)\beta_i^{-\alpha_i}} \tag{2.6}$$

where α_i and β_i are the parameters that define the gamma distribution for the growth increment of individuals that were initially in size-class *i*. The values for the parameters α_i and β_i are computed from the expected growth increment of an individual initially of size $\bar{l}_i$ (μ_i), which is determined using Equation 2.2, and the coefficient of variance for the growth increment (*CV*):

$$\mu_i = \frac{\alpha_i}{\beta_i} \qquad (\mu_i \cdot CV)^2 = \frac{\alpha_i}{\beta_i^2} \tag{2.7}$$

Therefore, there are only three estimable parameters: *k*, L_∞, and *CV* for this method.

The likelihood function is the sum over all recaptured individuals of the natural logarithm of the probability of an individual tagged in size-class *i* growing to the recapture size-class *j* given that it was at liberty for a time Δt (Moser *et al.*, 2002). These are the values within the size-transition matrix when accounting for time-at-liberty ($\boldsymbol{X}^{\Delta t}$ using matrix multiplication). Raising the size-transition matrix to the time-at-liberty accounts for growth during the intermediate time steps. Individuals tagged but not recaptured are not included since they do not provide information on growth. The likelihood function is not based on Equation 2.2 since the goal is create a size-transition matrix that fits the data. Selectivity is also included in the likelihood function to account for the probability of catching an individual of a particular size, i.e.:

$$\ln(L) = \sum_n \ln(S_{\bar{l}_{j,n}}[\boldsymbol{X}^{\Delta t_n}]_{\tilde{l}_{i,n},\tilde{l}_{j,n}}) \tag{2.8}$$

where $S_{\bar{l}_{j,n}}$ is the selectivity for the midpoint of the size-class into which individual *n* fell when it was recaptured, $\tilde{l}_{i,n}$ is the size-class of release for individual *n*, and $\tilde{l}_{j,n}$ is the size-class at recapture for individual *n*. The selectivity function needs to be pre-specified and is set, for this paper, to that used to generate the simulated data sets (outlined below). The time-at-liberties are integers representing years for this study because the assessment of golden king crab focuses on annual growth (Siddeek *et al.*, 2008).

2.3.2 Method 2: Platoon

The "platoon" method assumes that each individual follows one of several (for this paper three) growth curves. This framework is similar to finite mixture models in that the population is divided into sub-groups based on growth. For this paper, each growth curve has the same growth rate parameter (k), but a different asymptotic size (L_∞) (nominally "small", "medium", and "large"). There is a separate size-transition matrix for each growth curve, i.e.:

$$X_{p,i,j} = \int_{l_j^-}^{l_j^+} F_p\left(l_{2j}, \bar{l}_i, L_{\infty \cdot p}, k, \sigma_{within \cdot i}\right) \mathrm{d}l_{2j} \tag{2.9}$$

where $X_{p,i,j}$ is the probability of an individual in platoon p growing from size-class i to size-class j after one time step, $\sigma_{within \cdot i}$ is the variance about the platoon-specific growth increment when the initial size is $\bar{l}_i$, and $L_{\infty \cdot p}$ is the asymptotic size for platoon p. The values for $L_{\infty \cdot p}$ are determined from the median L_∞ for the entire population ($\widetilde{L_\infty}$) and the standard deviation for L_∞ (σ_L). The L_∞s for the "small" and "large" platoons are determined by adding or subtracting one standard deviation from $\widetilde{L_\infty}$. The L_∞ for the "medium" platoon equals $\widetilde{L_\infty}$. If there were five platoons, the additional two platoons would have L_∞'s that were +/- two standard deviations from $\widetilde{L_\infty}$. The pdf for growth variability within each platoon (F_p) is assumed to be a gamma distribution.

Taylor and Methot (2013) argued that the total variance in growth increment ($\sigma^2_{total \cdot i}$) can be broken down into two parts: the variance within platoons ($\sigma^2_{within \cdot i}$) and the variance among platoons ($\sigma^2_{between \cdot i}$). $\sigma^2_{within \cdot i}$ defines the variance in the growth increment from the underlying growth curve for each platoon while $\sigma^2_{between \cdot i}$ defines the variation in the expected growth increment among platoons. The values for the three variances are determined using ρ which equals $\sigma_{within \cdot i}/\sigma_{between \cdot i}$. Details on how this is done can be found in Appendix A1. The value of ρ controls the extent to which the platoons overlap. This parameter is pre-specified because it is

difficult to estimate (*sensu* Taylor and Methot, 2013). We explored the sensitivity to the pre-specified value for ρ by setting it to 0.75, 1.5, 2, 2.5 or 4.

Three parameters are estimated when using the platoon method; k, $\widetilde{L_\infty}$, and σ_L. Equation 2.8 cannot be used to estimate the parameters because the platoon structure is a theoretical concept. The likelihood function needs to account for the probability of an individual being in one of the three platoons. Therefore, the parameters are estimated using:

$$L = \sum_n \sum_p S_{\tilde{l}_{j,n}} \tau_p \left[\boldsymbol{X}_p^{\Delta t_n}\right]_{\tilde{l}_{i,n},\tilde{l}_{j,n}} \tag{2.10}$$

where τ_p is the proportion of recruits that settle to platoon p and $\boldsymbol{X}_p^{\Delta t_n}$ is the size-transition matrix for platoon p raised to the time-at-liberty. Multiplying each platoon's size-transition matrix by the proportion of new recruits to that platoon addresses the uncertainty related to which platoon a tagged individual belongs.

The values for τ_p are 16%, 68%, and 16% to the 'small', 'medium' and 'larger' platoons since we assumed that new recruits are randomly assigned to each platoon such that the distribution of recruits is approximately normal and mirrors how L_∞ is divided between platoons. This ensures that the average of each individual's size-transition matrix across all platoons is approximately the average size-transition matrix across all individuals if there were no platoons.

2.3.3Method 3: Numerical Integration

This method assumes that individuals follow their own growth curve. A single size-transition matrix is created, but instead of focusing on variation around a single curve, this method focuses on the variation in the values for the growth parameters by allowing either k, L_∞, or both to vary for each individual. For our work, we assume $L_\infty \sim \text{lognormal}\left(\overline{L_\infty}, \sigma_{L_\infty}\right)$, $k \sim \text{lognormal}\left(\bar{k}, \sigma_k\right)$ and that their values and distributions are independent. This is a simplifying assumption that ignores underlying ecological process that suggests the two growth parameters may be correlated

(e.g., Pilling *et al.*, 2002), although the method can be extended to allow for such correlation. The values within the matrix are determined using Equation 2.4. However, the pdf for growth variation cannot be written down in closed form when all individuals grow according to their own growth curves. The size of an individual after one time step depends on its growth parameters and initial size. Thus, there is no potential variation in the ending size of an individual. To calculate the probability of growing into a particular size-class, we force the range over which one of the growth parameters is integrated in Equation 2.4 to depend on the other. Three variants of this method (Methods 3a, 3b and 3c) differ in terms of whether k, L_∞, or both vary for each individual. The likelihood function used in Method 1 (Equation 2.8) is used for Method 3. Appendix A2 provide technical details on how the numerical integration method is implemented in this paper.

2.3.4 Simulation Evaluation

The methods described above, each with different assumptions about individual variation in growth, result in nine candidate methods (Table 2.1, with accompanied abbreviations that will be used henceforth). The formulae for the variation in the growth increment for each method are listed in Appendix Table A.3. We evaluated these methods using pseudo tag-recapture data generated using an individual-based operating model with biological characteristics mirroring golden king crab in the Aleutian Islands (100 simulated data sets for each scenario). The assessment of golden king crab is based on a linear relationship between initial and ending size, which matches the assumptions of von Bertalanffy growth curve. Previous analyses used similar techniques for simulating tag-recapture data to evaluate methods for estimating size-transition matrices (e.g., Kanaiwa *et al.*, 2005; Punt *et al.*, 2009). Two operating models are used. In the first operating model (OP1), each individual follows the same underlying growth curve, with deviations in growth increment due to process error. In contrast, individuals in the second operating model (OP2) have their own k and L_∞ values. The number of data points could impact the performance of each

estimation method. The golden king crab fishery has approximately 1,500 tag-recapture data points (Siddeek *et al.*, 2016). Thus, we considered pseudo tag-recapture data sets with 500, 1,500, 3,000 and 4,000 data points to explore the impact of sample size.

The width of the size-classes represents a trade-off between computational burden (quicker to have larger size classes) and biological precision (smaller size classes allow more detail to be captured). Szuwalski *et al.* (2014) showed that increasing size-class width could increase estimates of abundance. Therefore, the nine estimation methods constructed size-transition matrices with size-class widths of 5, 10, or 15mm for each data set. Changing the size-class width results in changes to the number of size-classes. A size-class width of 5mm has 18 size-classes, 10mm has 9, and 15mm has 6.

We simulated tag-recapture data by representing aspects of golden king crab life history, such as natural mortality and removal by the fishery, as a series of Bernoulli trials (Appendix Figure A.1). The first step is to obtain a value for the size-at-release for an individual. That individual is projected forward until it is recaptured, dies in the wild, or it is at liberty for more than the maximum time-at-liberty for golden king crab (7 years). For every yearly time step, the individual must first survive the year. The most recent assessment of golden king crab assumed that natural mortality (*M*) is the same for all individuals regardless of size (*i.e.*, 0.18yr^{-1}; Pengilly, 2016). If the individual survives, it grows according to Equation 2.2, where *t* equals 1. In OP1, all individuals follow the same growth curve, with variation in the growth increment due to process error, which is gamma distributed, i.e.:

$$\begin{aligned} \overline{GI} &= (L_\infty - l_1)(1 - e^{-kt}) \\ GI &\sim \Gamma(\alpha_1, \beta_1) \\ l_2 &= l_1 + GI \end{aligned} \tag{2.11}$$

where $\overline{GI}$ is the mean growth increment for an individual with initial size l_1, *GI* is the growth increment with process error and α_1 and β_1 define the gamma distribution whose values are determined using Equation 2.7 with $\overline{GI}$ replacing μ_i. In OP2, each animal follows its own growth curve, with k and L_∞ randomly chosen from lognormal distributions.

After growing, the individual encounters the fishery. The probability of an individual being caught is dictated by the selectivity curve and fishing effort (q). An individual is considered caught by the fishery (and hence included in the tag-recapture data set) if a random draw from $U[0,1]$ is less than the selectivity corresponding to the size of the individual multiplied by fishing effort. The assessment for the golden king crab assumes logistic selectivity:

$$S = \left[1 + \exp\left(-\ln(19)\frac{l_2 - \theta_{50}}{\theta_{95} - \theta_{50}}\right)\right]^{-1} \quad (2.12)$$

where θ_{50} is the size-at-50%-selectivity and θ_{95} is the size-at-95%-selectivity. We assumed that the values for the selectivity parameters are known when applying the estimation methods.

Four variants of the operating models are considered depending on how the initial sizes for individuals are selected (using actual data or a uniform distribution) and the maximum time-at-liberty (1 or 7 years). Tag-7 is a base-case as it is the closest representation of the actual gold king crab tag-recapture data. Uniform-7 explores the impact of the tagged animals representing a larger range of sizes compared to the distribution in Tag-7 (Appendix Figure A.2). Tag-1 explores reducing the conflict between discrete (size-transition matrix) and continuous (von Bertalanffy) growth since the number of time steps has been reduced to one. Uniform-1 combines the factors that constitute Uniform-7 and Tag-1, and would be expected to lead to the best performance because it provides the most informative data (broad range of sizes-at-release) and avoids dealing with the distinction between continuous and discrete growth. We used 864 scenarios to explore

the impact of sample size, size-class width, initial size distribution and maximum time-at-liberty (Appendix Table A.4).

The parameters of the operating model, their associated values, and the information on which those values are based are given in Appendix Table A.5. Fishing effort was chosen so that the distribution of the time-at-liberty resembles the actual distribution of the time-at-liberty for golden king crab (Appendix Figure A.3).

2.3.5 Evaluation

Three methods are used to evaluate the performance of the estimation methods. The first determines how well each method mimicked the simulated data. This involved first creating a distribution for the simulated sizes-at-recapture. The model-estimate of this distribution was computed by: (a) grouping the simulated data by size-class-at-release and time-at-liberty, (b) multiplying the number of individuals in each of these groups by the row of $\mathbf{X}^{T}$ (where T is the time-at-liberty) representing their predicted size-classes-at-recapture, and (c) summing the predicted size-at-recapture distribution over all recaptures. We visually compared these values and the number of size-at-recaptures for each size-class from the tag-recapture data produced through simulation when estimating the size-transition matrix.

The second performance metric, the matrix diagnostic, compares the "true" and estimated values for the size-transition matrix. The "true" matrix for both operating models is constructed using simulation. For OP1, values for l_{1i} are randomly drawn (U[l_i^-, l_i^+]) and inserted into the first line of Equation 2.11 (with t = 1) to determine the expected growth increment using the L_∞ and k values from OP1. This value, and the CV from OP1 are used to determine a gamma distribution from which a random value is drawn, added to l_{1i} (representing the ending size) and recorded. This process is repeated 30,000,000 times. The proportion of ending sizes that fall within each size-class determines the "true" size-transition matrix. In contrast, the creation of the "true" matrix for

OP2 draws new values for L_∞ (lognormal distribution [$\bar{L}_\infty$, σ_{L_∞}]) and k (lognormal distribution [$\bar{k}$, σ_k]) for each l_{1i}. These values are inserted into Equation 2.2 (with t = 1) to determine the ending sizes. The remaining steps match those used for constructing OP1's "true" size-transition matrix after obtaining the ending sizes.

The matrix diagnostic is the sum of the absolute differences between the estimated and "true" values of the size-transition matrix divided by the number of positions in the upper triangular portion of the size-transition matrix times 100 (equation shown in Appendix A3). This diagnostic is not based on relative errors because some of the entries of the size-transition matrix can be very small. Relative errors can make the differences between small values seem inappropriately important. When comparing the platoon method to the "true" matrix, the values within each platoon matrix are multiplied by the proportions of recruits added to that platoon and summed up to get a single estimated value. Differences in the matrix diagnostic below the second decimal place, which represents the 4th decimal place for the values in the size-transition matrix, were not considered 'appreciable' since the "true" matrix values were only repeatable to the third decimal place.

The third performance metric, the equilibrium diagnostic, compares the equilibrium size distribution between two theoretical population, each with a mortality rate of 0.18yr^{-1} (assumed rate of natural mortality for Aleutian Islands golden king crab) that used either the "true" or estimated size-transition matrix. We assumed that all recruits in these theoretical populations enter the first size-class. The equilibrium size-structure is determined by taking the right eigenvector of the product of the size-transition matrix and a matrix with the exponentiated mortality rate on the diagonals. The estimated and "true" equilibrium size-structures are compared using the average of the sum of the absolute error times 100 for each simulation (equations shown in Appendix A3).

2.4 Results

The estimation methods were able to mimic the simulated data sets well (Appendix Figure A.4). Fits improved with a smaller number of size-classes. Increasing the sample size also improved fits (Appendix Figure A.5). The simulated distribution of sizes-at-recapture depend on whether the initial sizes were generated from the actual tagging data versus a uniform distribution. However, this did not affect the estimation methods' ability to mimic the data (Appendix Figure A.6).

No estimation method was consistently among the best or worst performing for all scenarios for either the matrix or the equilibrium diagnostic (Figures 2.1 and 2.2; see Appendix Table A.6 for a tabular summary) except for Plat 0.75, which was always the worst for the equilibrium diagnostic (Figure 2.2). Estimation method performance was not consistent between diagnostics. For example, Plat 0.75 and Plat 1.5 performed well for the matrix diagnostic (top panels of Figure 2.1), especially when there was individual variation in growth (OP2; upper left panel of Figure 2.1) but they almost were never among the best methods for the equilibrium diagnostic (bottom panels of Figure 2.1).

Boxplots of the matrix and equilibrium diagnostics were used to explore the differences in performance among estimation methods (medians of the boxplot distributions listed in Appendix Tables A.7 – A.10). In general, increasing sample size reduced the among-simulation variation in the diagnostics while differences in performance between estimation methods remained relatively consistent. Consequently, our boxplots only show results for a sample size of 1,500 (this is approximately the tag-recapture sample size for golden king crab; see Appendix Figures A.7-A.10 for the full set of boxplots showing all sample sizes).

2.4.1 Ability to match the true size-transition matrix (matrix diagnostic)

Figures 2.3 and 2.4 show the distributions for the matrix diagnostic for OP1 and OP2 respectively. Low values imply better performance. In general, more size-classes resulted in lower values with a few exceptions such as Uniform-1, OP2, Plat 0.75 when comparing 6 versus 9 size-classes. The best performing method varied based on scenario and number of size-classes. Estimation methods performed best when the tag-recapture data were generated from a uniform distribution with a few exceptions such as when comparing Uniform-7 to Tag-1 for the numerical integration methods using 18 size-classes for OP2. The impact of the maximum time-at-liberty varied based on the initial size distribution. With a Uniform distribution, a 1 year time-at-liberty generally performed better than 7 years at liberty. However, the difference in performance between the maximum time-at-liberties was very small for 9-size-classes and OP1. The 7 year maximum time-at-liberty performed better for most scenarios using actual tagging data for OP1. However, for OP2, the difference between time-at-liberties depended on the number of size-classes with 7 years doing better with fewer size-classes and 1 year better with more size-classes.

2.4.2 Calculation of the equilibrium size-structure (equilibrium diagnostic)

Figures 2.5 and 2.6 shows the distributions for the equilibrium diagnostic for OP1 and OP2 respectively. The lowest value (i.e., best) occurred with more size-classes, although the best performing method again depended on the number of size-classes and scenario. A uniform initial size distribution generally resulted in a better performance except when comparing Tag-1 and Uniform-7 with OP2 and 18 size-classes. The influence of the maximum time-at-liberty varied based on the initial size distribution. A 1 year time-at-liberty performed best on average with a uniform distribution, with the exception of OP1 with 6 size-classes where there was barely any difference between the maximum time-at-liberties. When the tagging data were based on the actual tagging data, a 7 year maximum time-liberty performed better with fewer size-classes but the

difference between the maximum time-at-liberties decreased as the number of size-classes increased. This trend resulted in 1 year time-at-liberty performing best when there were 18 size-classes, with OP2.

2.4.3 The poor performance of the Plat 0.75 method

The poor performance of Plat 0.75 for the equilibrium diagnostic (Figure 2.2) was unexpected and relates to ρ. Small ρ values imply a large $\sigma_{between_i}$ and small σ_{within_i}, leading to less overlap in growth curves between platoons. Determining the equilibrium size-structure for the platoon method involves calculating the equilibrium size-structure for each platoon then combining them using the proportion of recruits to each platoon. The forced separation of growth trajectories with small ρ values is amplified when the equilibrium size-structure is calculated, leading to larger differences between the "true" and estimated equilibrium size-structures. Reversing the order of operation by combining each platoon's size-transition matrix first improves the ability to estimate the equilibrium size-structure. However, this ignores the platoon growth structure. Increasing ρ produces a similar effect because there is more overlap between the size-transition matrices (Appendix Figure A.11). This explains why platoon methods with higher ρ values occasionally perform best in terms of the equilibrium diagnostic (Figure 2.1). Unfortunately, estimating ρ is not an option because it is strongly negatively correlated with σ_L (results not shown).

2.5 Discussion

No estimation method consistently performed best across all factors considered in the analyses. The number of size-classes, maximum time-at-liberty, and how size-at-release was generated influenced performance. In addition, the matrix and equilibrium diagnostic metrics occasionally produced contradictory conclusions in terms of which estimation method performs best. Although

no method performed best, the goal of this work was to determine which estimation method was most robust, and best for general use.

A surprising result is the conflicting outcomes between the matrix and equilibrium diagnostics. Specifically, Plat 0.75 and Plat 1.5 were amongst the best methods for the matrix diagnostic yet were almost never the best in terms of the equilibrium diagnostic. In fact, Plat 0.75 had the poorest performance for the equilibrium diagnostic for all scenarios. The difference in performance between the two metrics occurs in part because different size-transition matrices can produce the same value for the matrix diagnostic, but not for the equilibrium diagnostic. This is due in part because the diagonals of the size-transition matrix heavily influence the equilibrium diagnostic.

2.5.1 What influences performance?

A goal of this work was to determine which estimation method is the most robust; unfortunately, there is no simple answer to this question. In general, the lowest value for the matrix and equilibrium diagnostic occurred with Uniform-1, with the largest number of size-classes having the smallest values overall (top boxplot of Figures 2.3-2.6). When the "true" matrix had individuals following a common growth curve, the Fixed method performed best (Tables A.7 and A.9) and when individuals followed their own growth curves, Vary k performed best (Tables A.8 and A.10). Decreasing the number of size-classes resulted in other estimation methods performing best. A possible reason for this is that fewer size-classes leads to an increase in size-class width causing the source of growth variability to be more difficult to detect. In addition, the assumption that individuals are uniformly distributed within a size-class is violated to a greater extent for wide size-classes.

The best estimation method also depends on the time-at-liberty. Size-transition matrices assume that growth is a discrete process. Thus, the distribution of size after two or more years for a given initial distribution of sizes is the product of the size-transition matrix two or more times.

Likelihood Equations 2.8 and 2.10 rely on the assumption that the distribution of size within each size-class is uniform for all time steps, even post-growth. This is not the case, and the assumption becomes increasingly more violated each time the size-transition matrix is multiplied.

The numerical integration method assumes that initial size is uniformly distributed, while the traditional and platoon methods assume the initial size for individuals is the midpoint of the starting size-class. Consequently, the numerical integration estimation methods are more consistent with the operating model when the initial sizes are generated from a uniform distribution (at least for a time-at-liberty of 1 year). The actual tagging data have a bell-shaped distribution over the entire size range (Appendix Figure A.2), which could explain why having a greater number of size-classes performs better since the distribution of sizes within smaller size-class widths more accurately mimic a uniform distribution.

The bell-shaped distribution of tag-recapture data does not always adequately represent all size-classes. Some size-classes, typically the larger ones, can have fewer data points. This reduces the contrast in the data making it more difficult to determine the size-transition matrix values associated with the data poor size-classes.

2.5.2 Next steps

The next steps in this work relate to estimation methods and scenarios. Additional estimation methods could be considered such as allowing k rather than L_∞ to vary among platoons for the platoon method and allowing for other distributions of sizes within the initial size-class. The number of years of tagging data used to estimate the size-transition matrix had more impact than anticipated. Analyses examining the trade-off between using fewer tag-recaptures (perhaps only those tags at liberty for a year or two) and reducing bias versus using more tags and increasing bias but reducing variance should be conducted. This trade-off will likely depend on total sample size. Our analyses assumed that selectivity was known, but there will always be some uncertainty

in this regard. Future analyses should also examine sensitivity to errors in estimates of selectivity as well as assess how robust the results are to the parameters that determine growth and natural mortality. Many stock assessments that use size-structured models, including that for golden king crab, estimate the probability of a crab molting. Tests should be conducted to examine how including molt probability influences the performance of the estimation methods.

2.5.3 Conclusions

This work has shown that numerical integration methods for estimating a size-transition matrix are viable. However, it is difficult to say which estimation method is the best due to confounding influences from the number of size-classes, maximum time-at-liberty, and the distribution of sizes at tagging. If we assume that crabs in the wild follow their own growth curve then Tag-7 and Tag-1 are the more realistic cases (non-uniform distribution of initial sizes, actual tagging data used). Therefore, we recommend Vary Linf as the default with smaller size-class widths and possibly restricting the tagging data to only one year. Vary Linf is always among the best performing methods for Tag-1 for the matrix diagnostic (Figure 2.4; Appendix Figure A.12). For the equilibrium diagnostic, it is among the best performing when there are fewer size-classes (Figure 2.6; Appendix Figure A.13). Vary Linf was also never among the worst performing methods for Tag-1 (Figure 2.2). Plat 0.75 and Plat 1.5 dominate Tag-7, yet should be avoided as their performance, particularly in terms of estimating the equilibrium size-structure, is poor. However, other methods can be valid alternatives depending on which characteristics interest the analyst.

2.6 Acknowledgments

We would like to thank multiple people for their help and support throughout this study. Shareef Siddeek (Alaska Department of Fish and Game) helped obtain the tag-recapture data for Golden King Crabs as well as explaining details about how growth is modeled within the Golden King

Crab stock assessment. Tim Essington (University of Washington), William Stockhausen (Alaska Fisheries Science Center) and two anonymous reviewers provided valuable critiques and suggestions on how to improve the study and manuscript. We would also like to thank the North Pacific Research Board, project number 1603: North Pacific crab growth, for funding our research.

Table 2.1. Methods for constructing a size-transition matrix and associated estimable parameters.

Method	Method Abbreviation	Description	Parameters estimated
Method 1	Fixed	Individuals follow a common curve	k, L_∞, σ^2
Method 2a	Plat 0.75	Platoon with $\rho = 0.75$	k, $\widetilde{L_\infty}$, σ_L
Method 2b	Plat 1.5	Platoon with $\rho = 1.5$	k, $\widetilde{L_\infty}$, σ_L
Method 2c	Plat 2	Platoon with $\rho = 2.0$	k, $\widetilde{L_\infty}$, σ_L
Method 2	Plat 2.5	Platoon with $\rho = 2.5$	k, $\widetilde{L_\infty}$, σ_L
Method 2c	Plat 4	Platoon with $\rho = 4.0$	k, $\widetilde{L_\infty}$, σ_L
Method 3a	Vary k	Variation in k	L_∞, $\bar{k}$, σ_k
Method 3b	Vary Linf	Variation in L_∞	k, $\overline{L_\infty}$, σ_{L_∞}
Method 3c	Both	Variation in L_∞ and k	$\overline{L_\infty}$, σ_{L_∞}, $\bar{k}$, σ_k

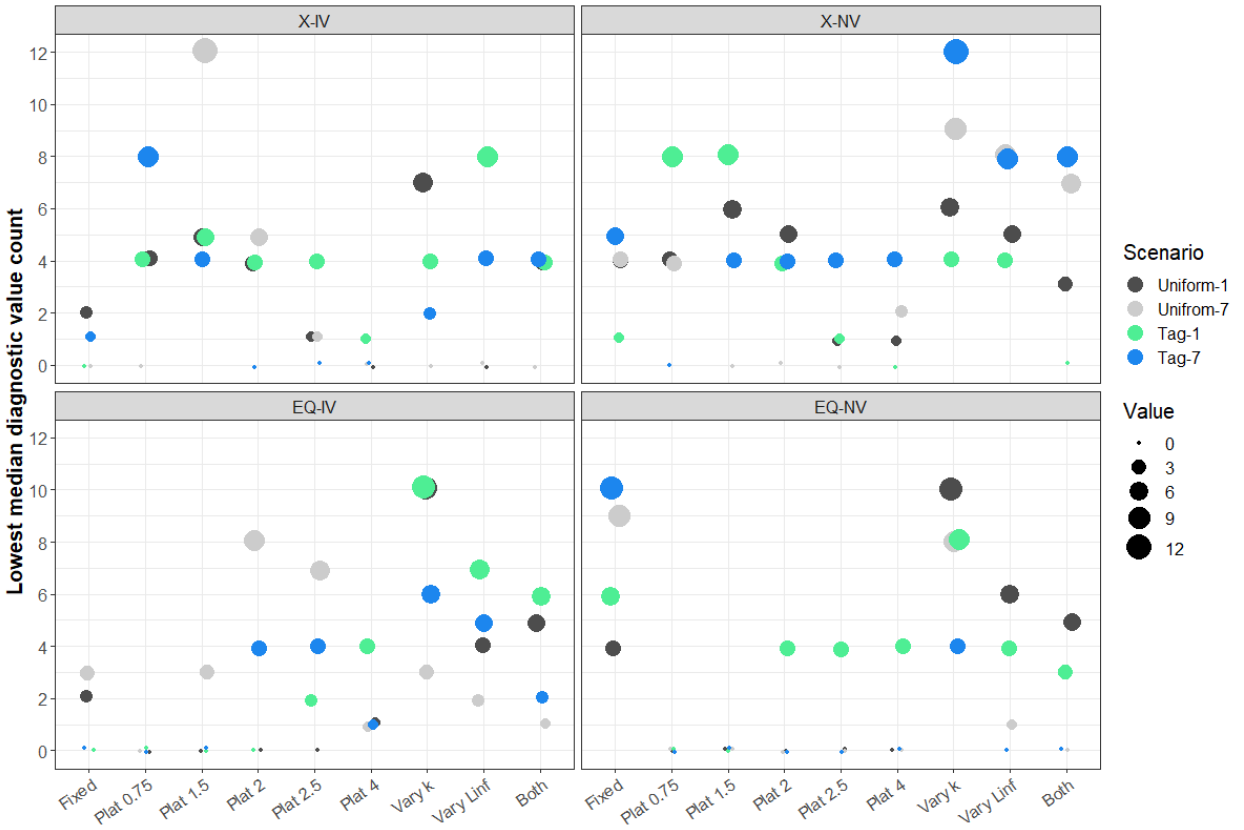

Figure 2.1. Counts for the number of times the median value, over 100 runs, of a particular diagnostic for an estimation method was among the lowest (meaning performed best) for a particular scenario. Methods were considered not appreciably different from the lowest value for the matrix diagnostic if the difference in median values was less than 0.1 from that of the lowest. For the equilibrium diagnostic, the discrepancy threshold was 0.01. The colors represent how the initial size was generated ("Uniform" – uniform distribution, "Tag" – actual tag-recapture data) and the maximum time-at-liberty (1 - one year or 7 - seven years). The counts arise from the options for sample size (4 options) and size-class width (3 options) with a maximum possible value of 12 for each color for each estimation method. The titles indicate the diagnostic (X – matrix diagnostic, EQ – equilibrium diagnostic) and whether there is individual variation in growth in the operating model (IV – Individual Variation; OP2, NV – No Individual Variation; OP1).

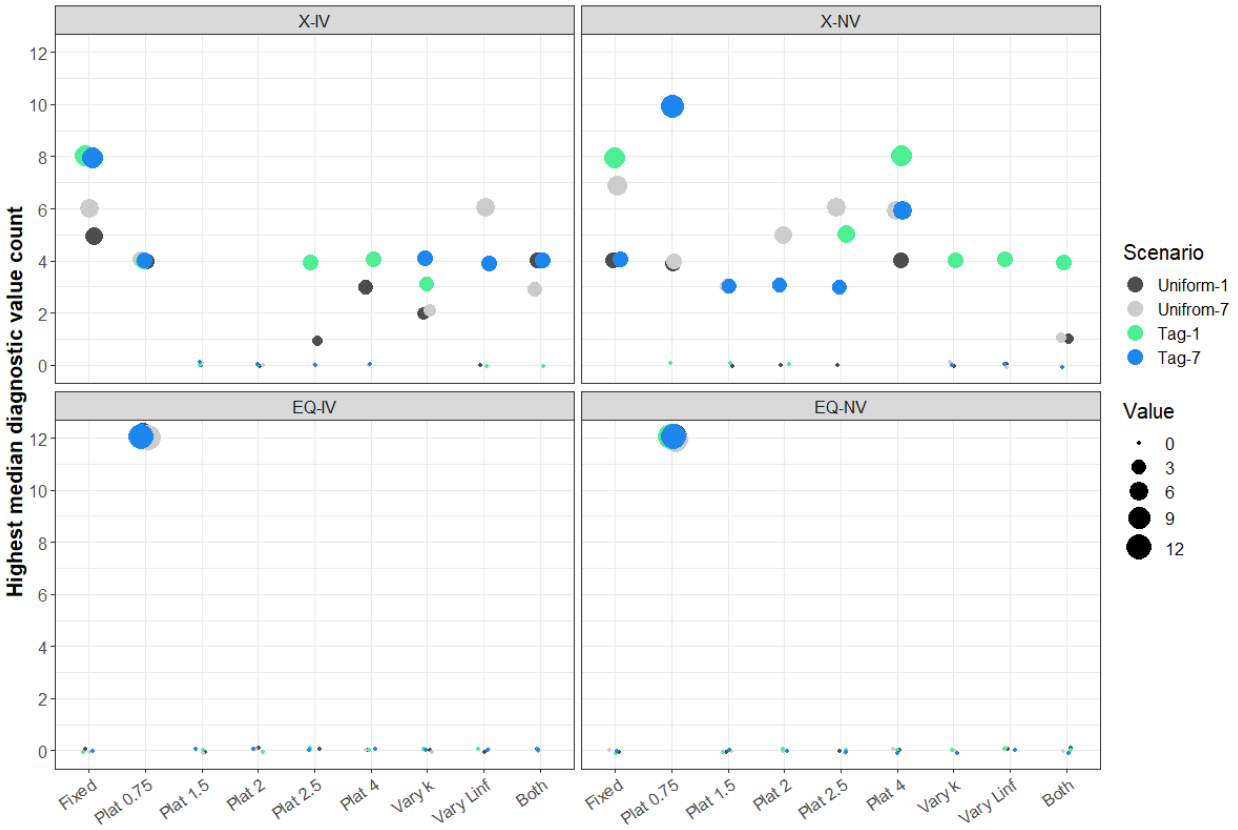

Figure 2.2. As for Figure 2.1 except the counts are of estimation methods that are among the highest (meaning performing poorly) median value.

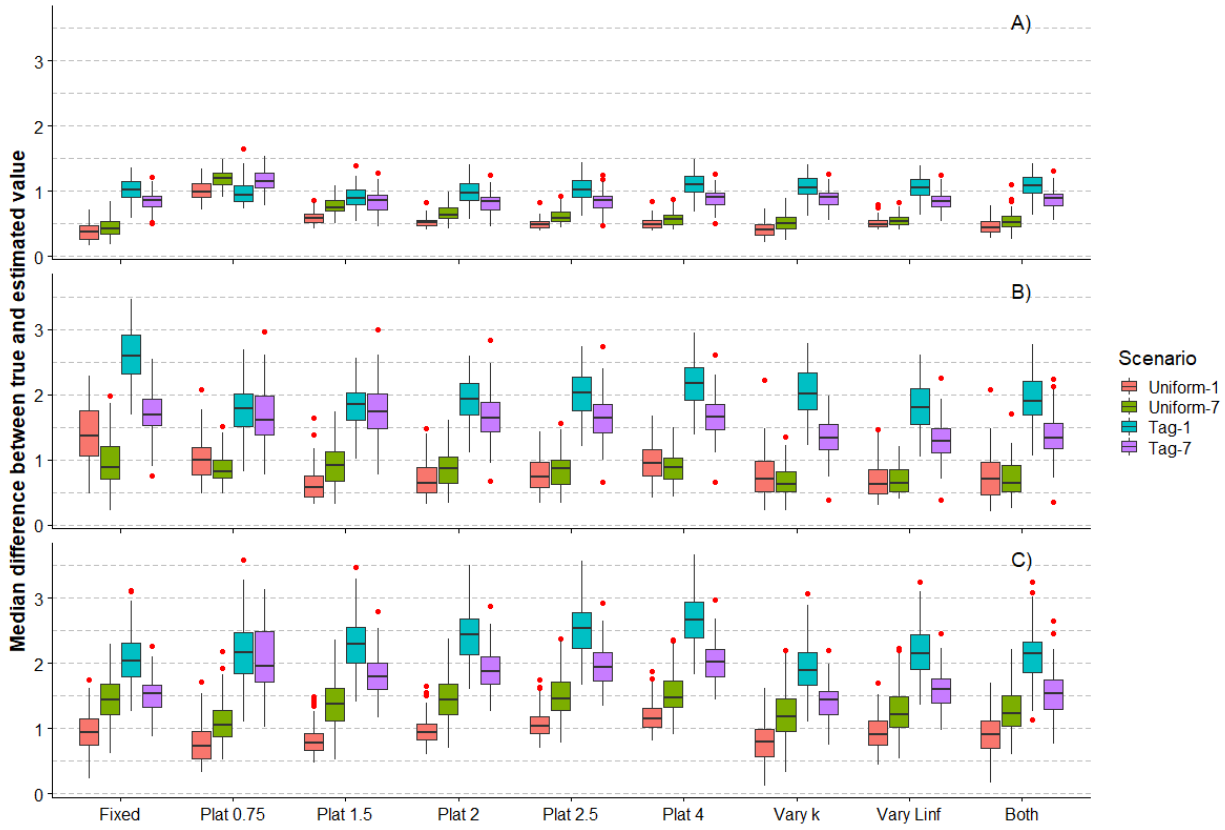

Figure 2.3. Boxplots of the matrix diagnostic for OP1. The colors indicate how the initial size was generated (Uniform – uniform distribution or Tag – actual tag-recapture data) and the maximum time-at-liberty (1 or 7 years). The letters in the upper right hand corner indicate the number of size-classes (A = 18, B = 9 and C = 6).

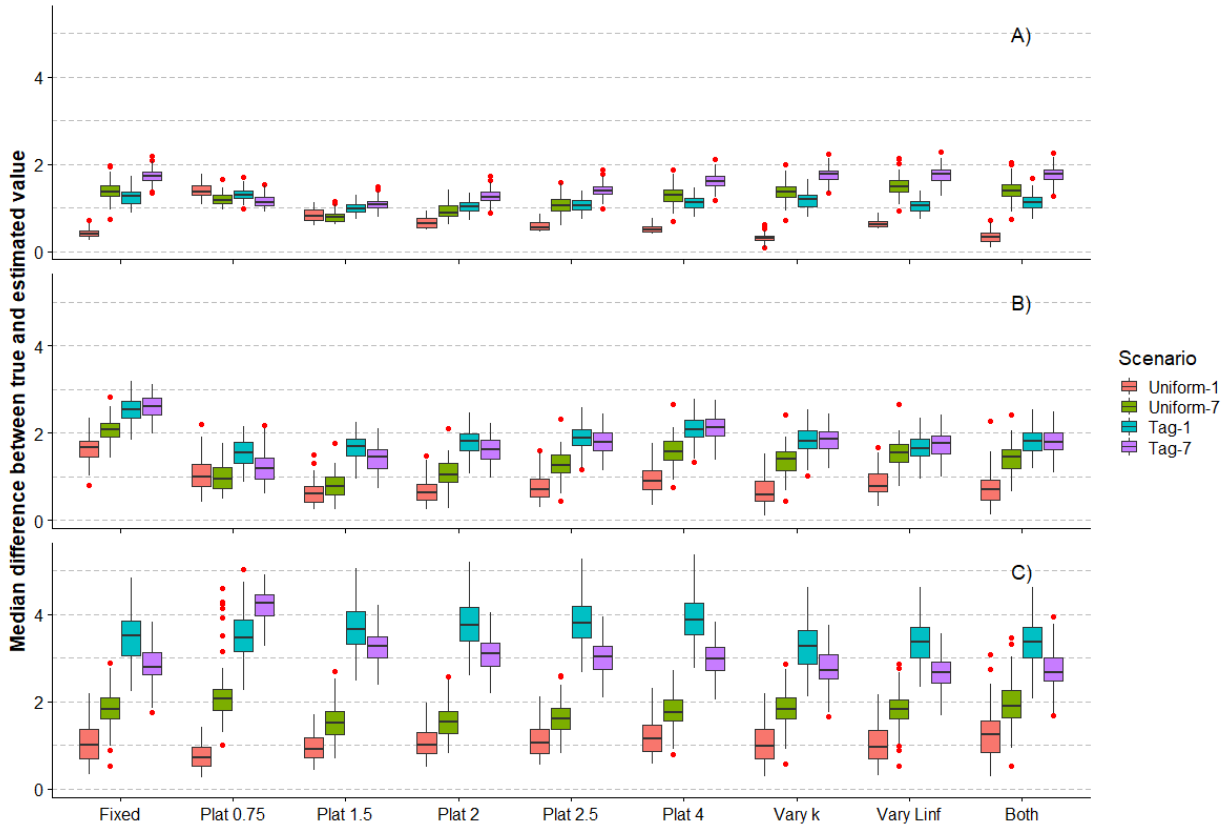

Figure 2.4. As for Figure 2.3, except that the operating model is OP2.

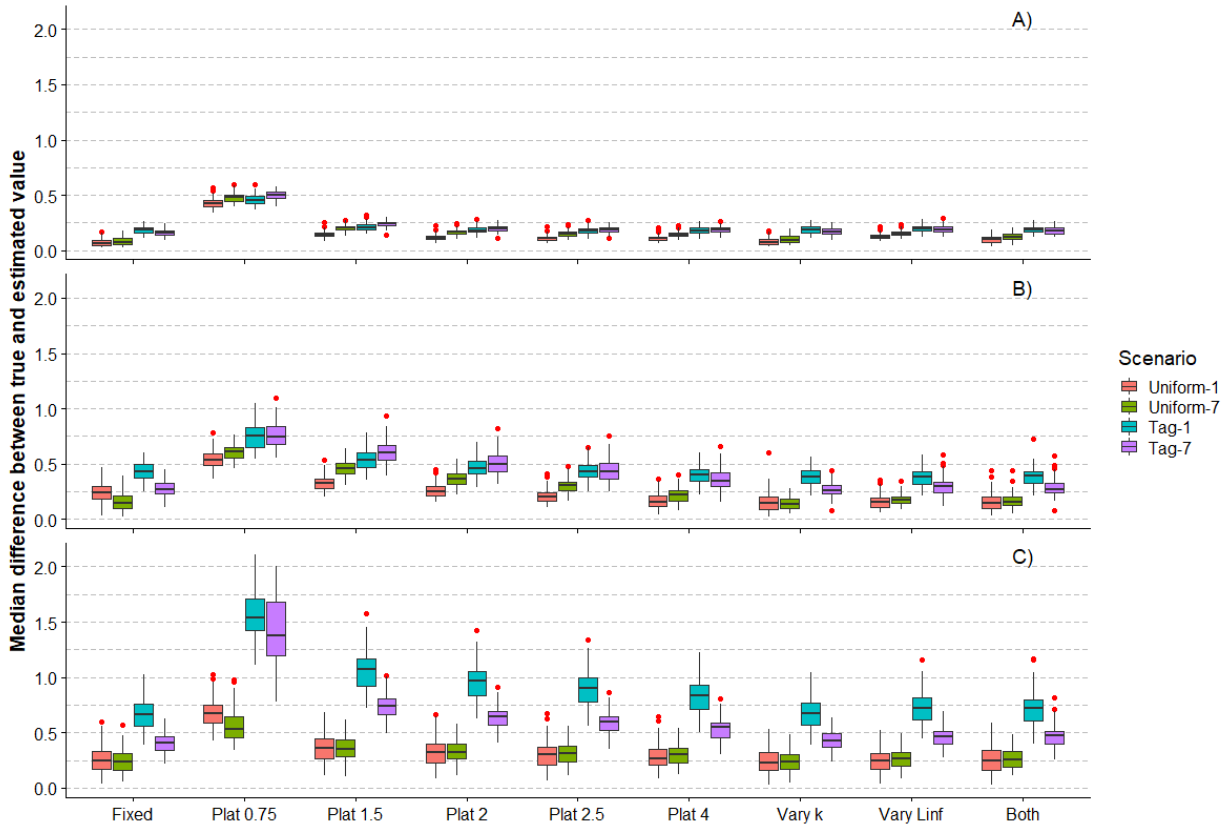

Figure 2.5. As for Figure 2.3, except that the metric is the equilibrium diagnostic.

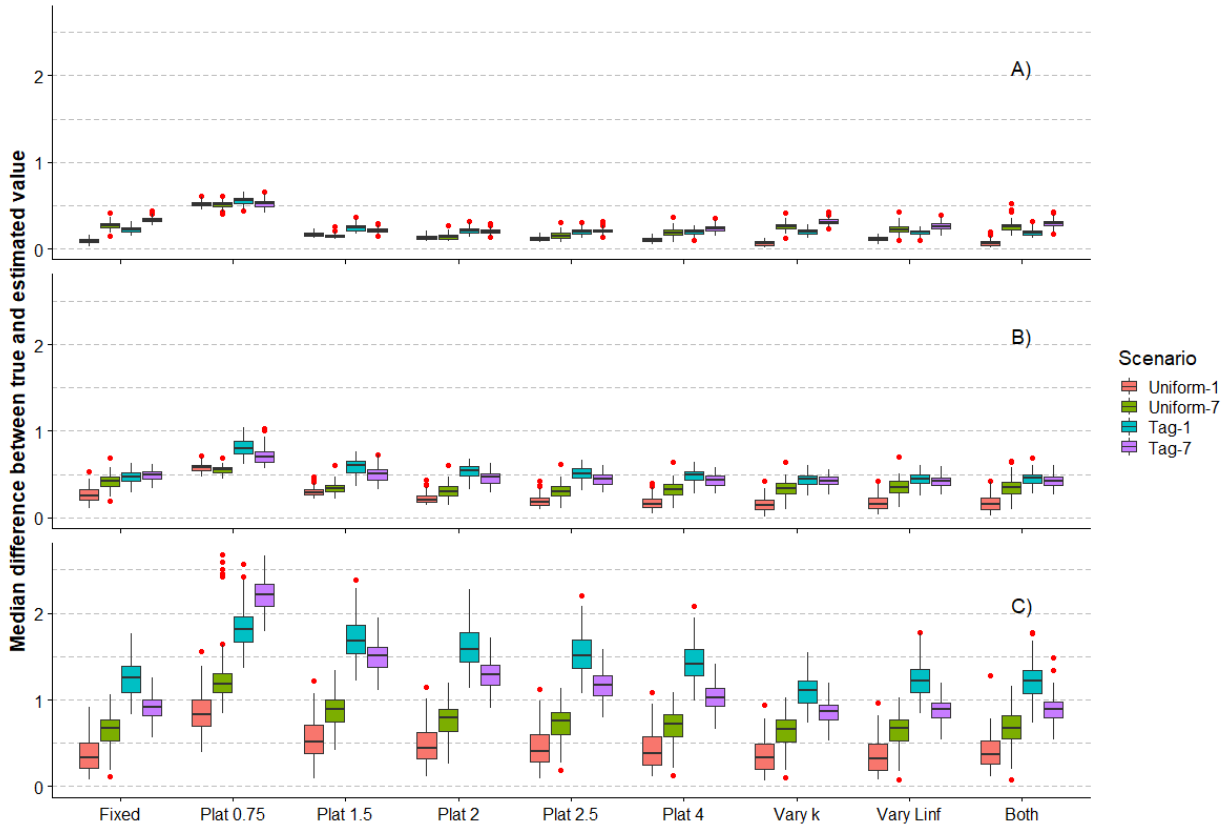

Figure 2.6. As for Figure 2.3, except that the metric is the equilibrium diagnostic and the operating model is OP2.

Chapter 3 : Modeling Time-Varying Selectivity in Size-Structured Assessment Models

3.1 Abstract

Within fisheries assessments, selectivity refers to the relative probability of a fish of a particular size or age being caught. Selectivity, typically modelled using a parametric function, is informed primarily by age- and size-composition data. This function can vary annually, and not accounting for this variation can result in biased estimates of abundance and mortality. Past simulation studies evaluating time-varying selectivity focused on age-structured models with little focus on size-structured models. In addition, the generating function for the size-composition data often mirrored the likelihood function. Unfortunately, this is not realistic and there is a need to determine which likelihood function is most robust. Therefore, this paper addresses two questions using simulation: (a) which is the best method for modeling time-varying selectivity and (b) which is the most robust likelihood function for size-composition data given time-varying selectivity. Both questions are addressed within the context of size-structured assessment methods. The results reveal that discrete time blocking can adequately capture time-varying selectivity. This could reduce the number of estimated parameters and hence the variance of estimated quantities. As for likelihood functions, the results reveal that the three likelihoods considered, multinomial, Dirichlet-multinomial and the multivariate normal, are all valid options. However, our preferred option is the multinomial because it has better estimates for desirable management quantities more often than the others.

3.2 Introduction

Formal stock assessments in Alaska and other regions rely on sophisticated population dynamics and observation models. For finfish populations, age-structured population dynamics models, which rely on age-composition data, are typically used because individual ages can be estimated

for many finfish species (Jones, 1986). A valuable fishery resource is crustaceans, which produced an estimated gross ex-vessel revenue of $184.12 million in 2017 in the Bering Sea and Aleutian Islands region of Alaska (Garber-Yonts and Lee, 2019). Unfortunately, crustaceans are difficult to age since they molt and have no permanent structure such as otoliths. Size-structured population dynamics models, which rely on size-composition data for parameter estimation, provide a good alternative for crustaceans.

An important component of a stock assessment model is selectivity, which refers to the relative probability of an individual of a particular size/age being caught, which is a function of gear selectivity and availability. A variety of functions can be used to model selectivity, including, but not limited to, the logistic, double logistic and double normal (Methot and Wetzel, 2013; Hulson and Hanselman, 2014). Sampson and Scott (2012) show that, in general, there can be considerable inter-annual variation in selectivity. Factors that can lead selectivity to vary over time include changes in fishing gear, fish behavior in relation to fishing gear, and the spatial distribution of fish (Sampson, 2014). Not accounting for temporal changes in selectivity can result in biased estimates of abundance and mortality rates (Gudmundsson and Gudmindsson, 2012, Radomski et al., 2005).

Two general approaches are used to account for time-varying selectivity. The first is to model selectivity in discrete time blocks. This approach groups years into “blocks”, with selectivity parameters constant within a block, but allowed to change between blocks. These blocks are usually selected based on known significant changes in fishing practices that are expected to impact the shape of the selectivity curve. The blocking method is used in the stock assessments for Eastern Bering Sea (EBS) Tanner crab (*Chionoecetes bairdi*) and EBS Pacific cod (*Gadus microcephalus*) (Stockhausen, 2015; Thompson, 2015), among many others. In both cases, the blocks are based on changes in fishing gear and fishing practices. However, determining when

changes in fishing behavior occurred is a difficult, and arguably subjective process (Martell and Stewart, 2014). The second approach for modeling time-varying selectivity is to allow for continuous changes in selectivity-at-age or -at-size. This can be accomplished by allowing selectivity parameters to follow a random walk (Gundmundsson, 1994) or to vary annually about mean values (random parameters). The stock assessment for EBS walleye pollock (*Gadus chalcogrammus*) uses a random walk process for selectivity (Ianelli et al., 2015).

Studies have evaluated methods for how best to model time-varying selectivity. For example, Martell and Stewart (2014) compared the performances of several estimation methods that differed in relation to whether time-varying selectivity is modeled as constant, using blocks or using a random walk, within the context of stock assessments based on age-structured population dynamics models. None of the estimation methods consistently outperformed the others. Overall, the conclusion was that it is more appropriate to assume selectivity varies with time than to assume that it is time-invariant. Linton and Bence (2011) used simulations to compare the performances of several estimation methods, again within the context of age-structured models, that use various methods to implement time-varying selectivity. These methods allow one, two or all four of the parameters of a double logistic selectivity curve to vary over time or for age-specific selectivity parameters to vary over time. Again, there was no single best approach. Both Martell and Stewart (2014) and Linton and Bence (2011) focused on age-structured models while using age-based selectivity. Size-structured models do not use age data, resulting in selectivity being assumed to be size-based. In addition, the population is tracked by size-class, and individuals are assumed to grow according to a size-transition matrix. For these reasons, the most appropriate way to model selectivity in age-structured models may not be the same for size-structured models.

Within stock assessments, size- and/or age-composition data are essential when estimating selectivity. These data, collected from fisheries and/or surveys, are used to configure a likelihood function which is used to estimate a variety of parameters, including those upon which selectivity is based. Typically, stock assessments use a multinomial or multinomial-based likelihood for composition data because these functions are assumed to closely match the random sampling design used to collect the data (Fournier and Archibald, 1982), or (more often) because it is conventional. The EBS Tanner crab and EBS Pacific cod stock assessments assume a multinomial likelihood for size- and/or age-composition data. However, there are arguably other, perhaps more appropriate, likelihood functions. For example, the stock assessment for EBS pollock uses a robust normal likelihood function. Maunder (2011) summarizes potential likelihood functions for size- and/or age-composition data, which include the multinomial, Dirichlet-multinomial and multivariate normal. Simulation studies exploring selectivity usually generate size- and/or age-composition data in a fashion similar to the likelihood function used when including these data in the assessment. This limits the generality of such studies because the variability in the size- and/or age-composition data could differ from that implied by the assumed likelihood (Punt et al., 2014). It is therefore desirable to vary the combinations of data generating and likelihood functions to determine which likelihood function is more robust.

This paper investigates two questions related to time-varying selectivity in size-structured models used for stock assessments. The first is the impact of assuming time-invariant, blocking or random parameters in relation to time-varying selectivity. The second evaluates three likelihood functions for size-composition data (the multinomial, Dirichlet-multinomial and multivariate normal) in relation to time-varying selectivity when the data are generated from functions that do not necessarily match the assumed likelihood. The multinomial likelihood was chosen because it

is conventional. The other two likelihoods can approximate a multinomial distribution, but allow for more variability, compared to the multinomial, and the Dirichlet-multinomial is used in assessments (Berger et al., 2019). The operating and estimation models are based on EBS Tanner crab. Model performance is determined using retrospective pattern analysis (which can be used to detect model misspecification), the ability to estimate quantities such as spawning stock biomass, which is of interest to management, and the ability of the likelihood function to estimate the effective sample size.

3.3 Methods

3.3.1 Overview

A simulation study has two components. The first is the operating model, which generates data sets, and represents the unknown 'truth'. There are typically multiple operating models given uncertainty about real world dynamics. The second component is the estimation methods, which estimate the parameters of a population dynamics model using the generated data sets and differ in terms of their parameterization (and perhaps population model and data). The aim of a simulation study is to determine which estimation method performs best at estimating quantities of management interest given the set of operating models.

This paper presents results from two simulation studies. The first examines how best to model time-varying selectivity. Five operating models are used to generate data. They differ in the degree to which selectivity varies over time: time-invariant, low variability, medium variability, high variability and very high variability. The data generated by each operating model are analyzed using three estimation methods, each with its own assumptions about time-varying selectivity: time-invariant (EM_{invar}), blocking (EM_{block}) and random parameters (EM_{RP}). A random walk was not used as an estimation method since EM_{RP} performs almost identically when the operating

model uses either random walk or random parameters (Appendix Figure B.1). Details about the operating and estimation methods are given below.

The second simulation study, which explores the performance of different likelihood functions for size-composition data in relation to time-varying selectivity, involves six operating models. The operating models differ in terms of whether selectivity is time-invariant or time-varying and the probability distribution (multinomial, Dirichlet-multinomial or multivariate normal) used to generate the size-composition data. The generated data are analyzed using the EM_{invar} or EM_{RP} estimation methods that differ in terms of the likelihood function assumed for the size-composition data. EM_{block} is not included since the goal of the second study is to focus on likelihoods. Details about the operating models and estimation methods are given below.

For both simulation studies, each combination of operating model and estimation method is simulated 100 times to allow broad trends in the performance of the estimation methods to be detected.

3.3.2 Operating model

The operating model, a two-sex size-structured population dynamics model with 32 size-classes, is loosely based on a simplified version of the actual stock assessment for EBS Tanner crab. There are several differences between the operating model and the model on which the actual stock assessment is based. For example, the operating model generates tagging data while the actual assessment does not, and the actual assessment has terminal molt while the operating model does not. These changes were made to draw generic conclusions and focus on selectivity. Each size-class is 5mm wide, with the smallest size beginning at 25mm and the last size-class being a plus group, i.e. all individuals larger than the lower limit of last size-class (180mm) are placed within the last size-class. The actual stock assessment for EBS Tanner crab has four fishery fleets, but only the directed fishery and the snow crab fishery take an appreciable number of Tanner crabs.

Consequently, the operating model includes three fleets: two fishery fleets (directed and snow crab), and a survey. Males can be retained or discarded (usually the smaller ones) in the directed fishery while all females are discarded. In the snow crab fishery, all crabs (both sexes) are discarded.

The population dynamics are governed by:

$$N_{y,s,i} = \sum_{j \le i} X_{s,j,i} N_{y-1,s,j} e^{-Z_{y-1,s,j}} + R_{y,s,i} \tag{3.1}$$

where $N_{y,s,i}$ is the number of animals at the start of year *y* of sex *s* in size-class *i*, $R_{y,s,i}$ is the number of recruits at the start of year *y* of sex *s* to size-class *i*, $X_{s,j,i}$ is the probability of individuals of sex *s* transitioning from size-class *j* to size-class *i* over a year (assumed time-invariant), and $Z_{y,s,i}$ is the total mortality during year *y* for animals of sex *s* in size-class *i*. The number of recruits is given by:

$$R_{y,s,i} = R_0 e^{\varepsilon_y - \sigma_R^2/2} P(i|a=0) \tag{3.2}$$

where R_0 is the expected number of recruits produced at unfished equilibrium, ε_y is the deviation during year *y* about the expected recruitment, σ_R is the degree of variation about expected recruitment, and $P(i|a=0)$ is the proportion of age-0 individuals of sex *s* that recruit to size-class *i*. Recruitment is assumed to be independent of spawning stock biomass, which is consistent with the assumption made by the EBS Tanner crab assessment. $P(i|a=0)$ is a gamma distribution integrated over each size-class range, approximated by:

$$P(i|a=0) = \frac{\bar{L}_i^{\alpha-1} e^{-\frac{\bar{L}_i}{\beta}}}{\sum_j \bar{L}_j^{\alpha-1} e^{-\frac{\bar{L}_j}{\beta}}} \tag{3.3}$$

where α and β are the parameters of the gamma distribution and $\bar{L}_i$ is the midpoint of size-class *i*. New recruits only join the first 10 size-classes i.e. 25mm - 75mm.

The total mortality for each year accounts for both natural mortality (assumed time-invariant) and fishing mortality:

$$Z_{y,s,i} = M_s + \sum_f F_{y,f}(\Omega_{f,s,i} + [1 - \Omega_{f,s,i}]\lambda_f)S_{f,y,i} \quad (3.4)$$

where $S_{f,y,i}$ is the selectivity for fleet *f* during year *y* on animals in size-class *i*, M_s is the natural mortality rate for animals of sex *s*, $F_{y,f}$ is the fully-selected fishing mortality for fleet *f* during year *y*, $\Omega_{f,s,i}$ is the proportion of animals of sex *s* in size-class *i* caught by fleet *f* that are retained, and λ_f is the mortality rate for individuals discarded by fleet *f*. The value for $\Omega_{f,s,i}$ for the directed fishery is determined using a logistic function:

$$\Omega_{dir,s,i} = \frac{1}{1+ e^{-\kappa_{dir,s}(\bar{L}_i - R50_{dir,s})}} \quad (3.5)$$

where $\kappa_{dir,s}$ is the slope parameter for the directed fishery for sex *s* and $R50_{dir,s}$ is the size-at-50%-retention for the directed fishery for sex *s*. The values for $\Omega_{f,s,i}$ for females in the directed fishery are zero because the fishery is male-only, and the values for $\Omega_{f,s,i}$ for both sexes in the snow crab fishery are zero because the entire catch is discarded.

All three fleets have logistic selectivity. The selectivity curve is defined by the inflection point and slope. Only the selectivity pattern for the directed fishery is time-varying (when the operating model has time-varying selectivity), with both parameters following a random walk. This means that the values for the selectivity parameters for year *y* depend on their values from year *y*-1. Previous studies explore time-varying selectivity in age-structured models using a random walk for selectivity parameters in the operating model (e.g., Martell and Stewart, 2014; Wilberg and Bence, 2006). Selectivity is therefore modelled as:

$$S_{f,y,i} = \frac{1}{1+ e^{-\theta_{f,y}(\bar{L}_i - L50_{f,y})}} \quad (3.6)$$

$$\theta_{f,y} = \theta_{f,y-1} + \delta\theta_{f,y}\,, \quad \delta\theta_{f,y} \sim N(0, \sigma^2_{\delta\theta}) \quad (3.7)$$

$$L50_{f,y} = L50_{f,y-1} + \delta L50_{f,y}\,, \quad \delta L50_{f,y} \sim N(0, \sigma^2_{\delta L50}) \tag{3.8}$$

where $\theta_{f,y}$ is the slope parameter for fleet *f* during year *y*, $L50_{f,y}$ is the inflection point for fleet *f* during year *y*, $\sigma^2_{\delta\theta}$ is the variance of the random walk in the slope parameter and $\sigma^2_{\delta L50}$ is the variance of the random walk for the inflection point. When the operating model has time-invariant selectivity, $\sigma^2_{\delta\theta}$ and $\sigma^2_{\delta L50}$ are set to zero. Figure 3.1 shows the selectivity curves for each fleet.

The catch is divided into retained and discarded:

$$C^R_{y,f,s,i} = \frac{F_{y,f}\Omega_{f,s,i}S_{f,y,i}}{Z_{y,s,i}} N_{y,s,i}(1 - e^{-Z_{y,s,i}}) \tag{3.9a}$$

$$C^D_{y,f,s,i} = \frac{F_{y,f}[1-\Omega_{f,s,i}]S_{f,y,i}}{Z_{y,s,i}} N_{y,s,i}(1 - e^{-Z_{y,s,i}}) \tag{3.9b}$$

where $C^R_{y,f,s,i}$ is the retained catch during year *y* by fleet *f* of sex *s* individuals in size-class *i*, and $C^D_{y,f,s,i}$ is the discarded catch during year *y* by fleet *f* of sex *s* individuals in size-class *i*. The weights of the landed ($\tilde{C}^R_{y,f}$) and discarded ($\tilde{C}^D_{y,f}$) catch by fleet *f* during year *y* are:

$$\tilde{C}^R_{y,f} = \sum_s \sum_i C^R_{y,f,s,i} w_{s,i} \tag{3.10a}$$

$$\tilde{C}^D_{y,f} = \sum_s \sum_i C^D_{y,f,s,i} w_{s,i} \tag{3.10b}$$

where $w_{s,i}$ is the mean weight of individuals of sex *s* in size-class *i*.

A size-transition matrix (*X*) is used to determine the probability of an individual transitioning from one size-class to another (only increases in growth are allowed). The probability of molting each year is assumed to be one, which is appropriate for immature Tanner crab but not for mature crab which do not grow. However, it is an appropriate assumption here given the focus on selectivity estimation and likelihood function selection. The expected growth increment is a linear function of pre-molt size with three parameters, the intercept ($\tilde{a}$), the slope ($\tilde{b}$) and the coefficient of variation of the growth increment (σ_{SE}). The size-transition matrix is determined by:

$$X_{s,j,i} = \begin{cases} \int_{-\infty}^{\bar{L}_i + \Delta L/2} P\left(l \middle| \bar{L}_j, \tilde{a}, \tilde{b}, \sigma_{SE}\right) dl & if\ i = j \\ \int_{\bar{L}_i - \Delta L/2}^{\bar{L}_i + \Delta L/2} P\left(l \middle| \bar{L}_j, \tilde{a}, \tilde{b}, \sigma_{SE}\right) dl & if\ i \neq j\ and\ i \neq\ x^L \\ \int_{\bar{L}_i - \Delta L/2}^{\infty} P\left(l \middle| \bar{L}_j, \tilde{a}, \tilde{b}, \sigma_{SE}\right) dl & if\ i =\ x^L \end{cases} \quad (3.11)$$

where ΔL is the width of each size-class (5mm), x^L is the maximum size-class and $P\left(l \middle| \bar{L}_j, \tilde{a}, \tilde{b}, \sigma_{SE}\right)$ is the probability density function for size after growth for an individual of size L, i.e.:

$$P\left(l \middle| \bar{L}_j, \tilde{a}, \tilde{b}, \sigma_{SE}\right) = \frac{1}{\sqrt{2\pi(\sigma_{SE}[\tilde{a} + \tilde{b}\bar{L}_j])^2}} exp\left[-\frac{(l - [\tilde{a} + \tilde{b}\bar{L}_j])^2}{2(\sigma_{SE}[\tilde{a} + \tilde{b}\bar{L}_j])^2}\right] \quad (3.12)$$

The size-transition matrices used in the operating model are shown in appendix B material (Tables B.1 and B.2).

Table 3.1 lists the values for the biological parameters used in the operating models and whether these values are estimated by the estimation methods. The operating models start the simulated Tanner crab population in 1947 at the unfished equilibrium. The population is then projected with fishing and natural mortality rates and random recruitment, so the population is no longer at equilibrium by the start of the first year with catch (1968).

3.3.3 Data generation

Five data sets (catch, discard, index, size-composition and tagging) are generated by the operating model. Table 3.2 lists the quantities that define how precise the various data sources are. The catch data, in kilograms, are generated for both fishery fleets. Retained catches are assumed known without error while the discarded catches are not, because these are based on extrapolated observer reports. Estimates of discarded catches are log-normally distributed with a standard error of σ_d in log-space:

$$\tilde{C}_{y,f}^{D,obs} = \tilde{C}_{y,f}^{D} e^{\delta d_y - \sigma_d^2/2} \quad (3.13)$$

where $\tilde{C}_{y,f}^{D,obs}$ is the observed discard catch for fleet f during year y, δd_y is the deviation in the discard catch estimate for year y and σ_d is the standard error of δd_y.

The survey provides an index of the number of individuals by sex, subject to log-normal error, i.e.:

$$N_{y,s}^{obs} = [\sum_i N_{y,s,i} S_{sur,y,i}] e^{\delta s_{y,s} - \sigma_{sur}^2/2} \tag{3.14}$$

where $N_{y,s}^{obs}$ is the observed survey index, $\delta s_{y,s}$ is the observation error for sex s and year y, and σ_{sur} is the standard error of $\delta s_{y,s}$.

Size-composition data are generated for all fleets. Depending on the scenario, size-composition data are generated using one of three distribution functions: the multinomial, the Dirichlet-multinomial and the multivariate normal. The mean vector for the multivariate normal distribution is the vector of proportions of catch by size-class and the variance covariance matrix ($\mathbf{\Sigma}_{i,j}$) is defined by:

$$\mathbf{\Sigma}_{i,j} = \begin{cases} \sigma_M^2 P_i (1 - P_i) & if\ i = j \\ -\sigma_M^2 P_i P_j & if\ i \neq j \end{cases} \tag{3.15}$$

where σ_M^2 is the effective sample size and P_i is the proportion of the catch in size-class i. The variance-covariance matrix for the multivariate normal distribution is thus parameterized to approximate a multinomial distribution.

For the tagging data, individuals are tagged during the survey conducted in "2005". A total of 1,000 individuals (males and females) are randomly sampled from the survey catch. The time-at-liberty for each individual is randomly chosen between 1-5 years. The probability of recapturing a tagged individual in size-class j that was initially in size-class i when the time-at-liberty equals t ($Pc_{t,j,i}$) is determined by:

$$Pc_{t,j,i} = \frac{X_{s,j,i}^t S_{1,y,j}}{\sum_j X_{s,j,i}^t S_{1,y,j}} \tag{3.16}$$

The number of tagged individuals released in size-class *i* and recaptured in size-class *j* is randomly chosen using a multinomial distribution with mean given by the vector $Pc_{t,j,i}$.

3.3.4 Estimation method

The basic structure of the estimation methods mirrors the operating model. Each generated data set has an associated likelihood function used to estimate the designated parameters in Table 3.1 (Appendix B1). For the first simulation study, the estimation method and operating model differ in terms of how male selectivity is modeled in the directed fishery. In the estimation methods, time-varying selectivity is modeled as: a) time-invariant (EM_{invar}), b) blocking (EM_{block}) with multiple five-year blocks, i.e. selectivity parameters remain constant for five years, or c) allowing the selectivity parameters to vary every year using a random parameter process (EM_{RP}). EM_{block} and EM_{RP} involve estimating the selectivity parameters for the first year/block then estimating deviations to those parameters for the remaining years/blocks:

$$\theta_{dir,yb} = \theta_{dir,1} e^{\tau_{dir,yb}} \qquad L50_{dir,yb} = L50_{dir,1} e^{\varphi_{dir,yb}} \tag{3.17}$$

where $\tau_{dir,yb}$ is the estimated deviation of the slope for the directed fishery in year *y* / block *b* and $\varphi_{dir,yb}$ is the estimated deviation of the inflection point for the directed fishery in year *y* / block *b*. A penalty is placed on the deviations to limit how much the selectivity parameters can vary over time:

$$\frac{\sum_y \tau_{dir,yb}{}^2}{2\sigma_\tau^2} \qquad\qquad \frac{\sum_y \varphi_{dir,yb}{}^2}{2\sigma_\varphi^2} \tag{3.18}$$

where σ_τ is the standard error for $\tau_{dir,yb}$ and σ_φ is the standard error for $\varphi_{dir,yb}$. The value for σ_τ is 10 and σ_φ is 1.

There are six estimation methods for the second simulation, with the estimation method and operating model differing in terms of how male selectivity is modeled in the directed fishery and the assumed likelihood for the size-composition data. Selectivity is assumed to be either EM_{invar}

or EM_{RP}. The likelihood functions are either multinomial, Dirichlet-multinomial or multivariate normal (Appendix B1).

3.3.5 Experimental design

The first simulation study involved five operating models (Table 3.3). One operating model assumed time-invariant selectivity while the others allowed for time-varying selectivity (low, medium, high and very high). Each generated data set was analyzed using the three estimation methods for a total of 15 simulation scenarios (Figure 3.2). There were six operating models for the simulation study examining the effects of different likelihood functions for the size-composition data given time-varying selectivity (Table 3.3). The operating models differed in terms of how male selectivity was modeled in the directed fishery (time-invariant or medium time-varying) and the probability distribution used to generate size-composition data (multinomial, Dirichlet-multinomial, or multivariate normal). Each generated data set was analyzed using six estimation methods, resulting in 36 simulation scenarios (Figure 3.3).

The performances of the estimation methods are summarized in several ways. The first involves the estimation method's ability to approximate desired management quantities. Results are shown for R_0, the time-series of fully-selected fishing mortalities ($F_{y,f}$) and the spawning stock biomass (*SSB*). SSB is determined yearly by:

$$SSB_y = \sum_s \sum_i N_{y,s,i}\, Mat_{s,i} w_{s,i} \tag{3.19}$$

where $Mat_{s,i}$ is the proportion of individuals of sex *s* in size-class *i* that are mature. Maturity is a logistic curve defined by the slope parameter (τ) and the size-at-50%-retention (*Mat50*). Several crab assessments, such as that for EBS Tanner crab, use the mature male biomass (MMB), male portion of the spawning stock biomass, instead of spawning stock biomass. Therefore, results are shown for both MMB and SSB. Three management quantities based on spawning stock biomass, the unfished spawning stock biomass (SSB_0), the ratio between spawning stock biomass in 2016

and SSB_0 (SSB_{2016}/SSB_0), and the ratio SSB_{2016}/SSB_{1968} (1968 is the year when fishing began) were also assessed. These values are relevant to fisheries management since they could be used to determine the status of the stock and whether it is overfished. We use median relative error (MRE) and median absolute relative error (MARE) to quantify estimation performance:

$$MRE = 100 * median_i\left(\frac{E_i - T_i}{T_i}\right) \qquad MARE = 100 * median_i\left(\frac{|E_i - T_i|}{T_i}\right) \qquad (3.20)$$

where E_i is the estimated value and T_i is the true value. We focus on the median instead of the mean because there are occasional outlying estimates of biomass. Both simulation studies involve many scenarios. One way to interpret the MAREs is to determine which estimation method has the overall least worst MARE for a given management quantity. This value is determined by comparing the largest MARE (across operating model scenarios) from each estimation method for a given management quantity and finding the value that is the smallest from this group. For example, in the first simulation study there are three estimation methods each with their own worst MARE values for a given management quantity. The least worst MARE is the smallest from this group of three. The idea is that instead of focusing on which estimation performs the best; the focus is on identifying which estimation method performs the least poorly.

The second evaluation method involves determining whether the assessment exhibits retrospective patterns, which would indicate a mis-specified model, for spawning stock biomass and recruitment. Retrospective analysis involves removing the last year of data and rerunning the estimation method. This process is repeated several times, with each new estimation having one fewer year of data than the previous one (Hurtado-Ferro et al., 2015). Past studies use retrospective patterns to compare time-varying selectivity methods (e.g., Linton and Bence, 2011; Martell and Stewart, 2014). Linton and Bence (2011) identify retrospective patterns as the "best selection method" when compared to root mean square error (RMSE) and the deviance information criteria

(DIC). We use a modified version of Mohn's ρ (Mohn, 1999), which quantifies retrospective patterns, from Hurtado-Ferro et al. (2015) to evaluate the retrospective patterns. The modified equation for Mohn's ρ is:

$$\rho = \left. \sum_p \left(\frac{X_{y,p} - X_{y,ref}}{X_{y,ref}} \right) \right/ np \tag{3.21}$$

where $X_{y,p}$ is the value of the management quantity for year *y*, which is the last year in "peel" *p*, *ref* refers to the reference "peel", which is the estimation method that uses all years of data, and *np* is the number of times the process of removing a year of data is conducted (=5 for this study). "Peel *p*" refers to the estimation method run in which *p* fewer years of data are used. When conducting retrospective analysis, the goal is to not see retrospective patterns. Based on this, Hurtado-Ferro et al. (2015) suggest that values for Mohn's ρ higher than 0.20 or lower than -0.15 indicate a cause for concern. Appendix Figure B.2 provides examples of Mohn's ρ values that do and do not have a retrospective pattern according to this criterion.

A third evaluation method is applied solely to the second simulation study. The effective sample size helps determine the degree of independence in the data and how much information it provides about the model parameters. The multivariate normal and the Dirichlet-multinomial can estimate the effective sample size while the multinomial cannot. Studies have been conducted to determine ways to calculate the effective sample size for the multinomial likelihood to properly weight the data (Punt, 2017). The method of McAllister and Ianelli (1997) is used here and involves calculating a "multiplier" that is applied to the actual sample size to determine the effective sample size.

$$\left(\frac{1}{N_y} \sum_y \left(\left. E_y \right/ N_y \right)^{-1} \right)^{-1} \tag{3.22a}$$

$$E_y = \sum_i \widehat{P_{y,i}}\left(1 - \widehat{P_{y,i}}\right) \Big/ \sum_i \left(P_{y,i} - \widehat{P_{y,i}}\right) \quad (3.22b)$$

where E_y is the effective sample size for year *y*. The estimated (multivariate normal and Dirichlet-multinomial) and calculated (multinomial) effective sample sizes are compared to the actual effective sample size used to generate the data to determine the ability of each likelihood function to estimate/calculate it.

3.4 Results

3.4.1 How best to model time-varying selectivity

Table 3.4 lists the MRE and MARE values for each estimation method for the simulation study exploring how best to model time-varying selectivity. None of the three methods (EM_{invar}, EM_{block} and EM_{RP}) consistently outperform the others. For all scenarios, the three management quantities are close to unbiased since the MREs stayed between -10% and 10%, except for two EM_{RP} values. The estimates of SSB_{2016}/SSB_{1968} are, however, positively biased for all combinations of estimation method and operating model. Reasons for this are explored in Appendix B2.

As expected, the MAREs are larger than the MREs. MREs focus on whether the estimated values are larger or smaller than the true value, while MAREs quantify the scale of the difference. When the operating model has time-invariant selectivity, the MARE values for all estimation methods are similar with the percent difference less than 10%1 (Table 3.4; Appendix Table B.3, row "Time-Invariant"). This suggests that all estimation methods perform well and that the estimation of additional (unnecessarily) parameters when allowance is made for time-varying selectivity does not lead to an appreciable increase in estimation variance. EM_{invar} has the largest of the three worst values (bold values in Table 3.4) for two of the management quantities

1 A 10% difference in MARE is considered here to be noteworthy.

(SSB_{2016}/SSB_0 and SSB_{2016}/SSB_{1968}) implying it performs the worst for those management quantities. EM_{block} has the smallest worst MARE value for one of those quantities and is only 0.01 greater than the other best worst MARE. EM_{invar} has the smallest worst MARE and EM_{RP} the largest for the third management quantity (SSB_0).

We expected the MAREs to increase as the extent of time-varying selectivity increases in the operating model. This occurs for SSB_0 and SSB_{2016}/SSB_0, except for SSB_{2016}/SSB_0 when the operating model has medium time-varying selectivity. For both metrics, estimation methods perform similarly within operating models (percent difference less than 10%). However, there are several exceptions such as EM_{invar} which performs markedly worse for SSB_{2016}/SSB_0 when the operating model has high time-varying selectivity and EM_{invar} outperforming EM_{RP} for SSB_0 when there is very high time-varying selectivity. SSB_{2016}/SSB_{1968} has a different trend, with EM_{block} performing markedly better (> 10% difference, Appendix Table B.3) and EM_{RP} performing slightly better (< 10% difference, Appendix Table B.3) than the lowest MARE value for the time-invariant operating model, except when the operating model has medium time varying selectivity. In this situation both estimation methods perform slightly worse (< 10% difference, Appendix Table B.3) than their performance for the time-invariant operating model. EM_{invar} performs much worse as the level of time-varying selectivity increases in the operating model, except when there is low time-varying selectivity in which case performance is markedly better. This difference in performance between estimation methods is visually apparent from the relative errors in spawning stock biomass (Figure 3.4). Mature male biomass shows the same pattern (Appendix Figure B.5). All plots in Figure 3.4, regardless of operating or estimation model, show that spawning stock biomass is positively biased at the end of the time series. This result appears to be due to the generated size-composition data providing limited information for the cohorts that make up the

population during the last few years. Increasing the effective sample size for the survey size-composition data removes this bias (Appendix B2).

All estimation methods can estimate R_0 well but with positive bias, regardless of the operating model (Appendix Figure B.6). As for fully selected fishing mortality, EM_{invar} performs worse as the level of time-varying selectivity increases in the operating model (Appendix Figure B.7). EM_{block} and EM_{RP} remain constant for all operating models. The trends in fully-selected fishing mortality mirror those in spawning stock biomass.

The Mohn's ρ values for spawning stock biomass from the 100 simulations stay within the range suggested by Hurtado-Ferro et al. (2015) with a few outliers and little variability among estimation methods (Appendix Figure B.8; Table B.4). This suggests a lack of retrospective patterns for all estimation methods, hence Mohn's ρ does not distinguish among the estimation methods. However, the standard deviation in Mohn's ρ for spawning stock biomass for EM_{invar} becomes larger as the level of time-varying selectivity increases beyond medium level (Appendix Table B.4). This pattern is evident to some extent for the other estimation methods, but they do not exhibit as large of an increase as EM_{invar}.

The range of Mohn's ρ for recruitment is larger than that for spawning stock biomass, with some values outside the suggested range, implying noteworthy retrospective patterns (Figure 3.5). Within each operating model scenario, the difference between each estimation method's mean, median and standard deviation for Mohn's ρ is less than one percent (Appendix Table B.4). The only exceptions are that EM_{block} has a lower median than EM_{invar} when the operating model has high time-varying selectivity while the reverse occurs for standard deviations when the operating model has very high time-varying selectivity. Regardless, all estimation methods have some difficulties estimating recruitment since they all have Mohn's ρ values outside the suggested range.

3.4.2 Effects of different likelihood functions given time-varying selectivity

Table 3.5 lists the MREs and MAREs for the simulation study examining different likelihood functions for size-composition data in relation to time-varying selectivity. All MREs are between -10% and 10%, except for several SSB_{2016}/SSB_{1968} MREs (< 2% above 10%), suggesting that the estimated management quantities are close to unbiased. No estimation method consistently outperforms the others (Figure 3.6). This is surprising since we expected better performance when the operating model and estimation method match. With regards to the worst (largest) MARE, two of the three smallest values occur for estimation methods that use the Dirichlet-multinomial likelihood with time-varying selectivity. The largest worst MAREs all have the multivariate normal likelihood, with two of the three occurring for EM_{invar} (Table 3.5).

The percent difference from the lowest MARE value within each operating model scenario is less than 30% (Appendix Table B.5) with only eight exceptions. The average percent difference for each management quantity is less than 13%. This suggests that the three likelihoods perform similarly. The relative error in spawning stock biomass provides a visual representation of this (Figures 3.7 and 3.8). Mature male biomass shows the same pattern (Appendix Figures B.9 and B.10). With six operating models and three management quantities, there are a total of 18 chances for a likelihood to perform best in terms of MARE. Of these 18, the multinomial performs best the most (9 out of 18 times). Five of these occur when the estimation method has time-varying selectivity. The multinomial estimation method with time-varying selectivity has the most best values out of all estimation methods. There are also 18 chances for a likelihood to have the worst MARE. When disregarding whether there is time-varying selectivity, the multinomial has the fewest worst values, with two. All of these occur when the estimation method has time-invariant selectivity. The multinomial with time-varying selectivity has the fewest worst values out of all

estimation methods with a total of zero. Even though the multinomial has the best MAREs most often, there are still multiple occasions where either the Dirichlet or multivariate normal is the best.

All estimation methods perform similarly in terms of estimating R_0 regardless of whether there is time-varying selectivity when the generating function is multinomial or multivariate normal (Appendix Figure B.11). However, all estimation methods perform slightly worse when the generating function is multivariate normal (mean farther away from true value) compared to the multinomial generating function. When the generating function is Dirichlet-multinomial, estimation methods with multivariate normal likelihoods perform slightly better (mean closer to true value) than other estimation methods. However, the true R_0 value is within the 25-75 quantile for all estimation methods when the generating function is Dirichlet-multinomial. All estimation methods perform similarly well for fully-selected fishing mortality regardless of the generating function and whether there is time-varying selectivity in the operating model (Appendix Figures B.12 and B.13). The only exception is when the generating function is Dirichlet-multinomial and the estimation method has time-varying selectivity with a multivariate normal likelihood. In this case, there is significantly more variability regardless of whether the operating model has time-varying selectivity.

A majority of the Mohn's ρ values for spawning stock biomass stay within the suggested range, although several are outside the range, particularly for the multivariate normal likelihood when the operating model has a Dirichlet-multinomial generating function, irrespective of whether there is time-varying selectivity (Appendix Figure B.14). The largest number of values outside the range occur when there is time-varying selectivity in the operating model where the multivariate normal has 5-6 values outside the range. This suggests that the multivariate normal likelihood is

susceptible to retrospective patterns when the generating function is Dirichlet-multinomial and there is time-varying selectivity.

The distribution of Mohn's ρ for recruitment is wider than that for spawning stock biomass (Appendix Figure B.15). All estimation methods lead to Mohn's ρ values outside the suggested range, implying some noteworthy retrospective patterns regardless of the operating model. The range of values outside the range was 26-43 with no estimation method consistently having the highest or lowest count.

In relation to effective sample sizes, the performance of the Dirichlet-multinomial estimation depends on the generating function used. It estimates the effective sample size better than the multivariate normal when the generating function is Dirichlet-multinomial (Appendix Figure B.16). However, it goes to the upper estimation bounds when the generating function is either multinomial or multivariate normal (Appendix Figures B.17-B.18). The multivariate normal likelihood can produce an effective sample size estimate for all generating functions. However, the accuracy of the estimates varies between generating function, size-composition data source and whether the estimation method assumes time-varying selectivity (Appendix Figures B.19-B.20). The calculated multinomial effective sample size performs well. However, results vary between generating function, size-composition data source and whether the estimation method assumes time-varying selectivity. The performances in terms of estimating or calculating effective sample size do not vary when the operating model has time-varying selectivity.

3.5 Discussion

3.5.1 Best method for time-varying selectivity

No estimation method consistently outperforms the others for all scenarios. However, performance is better when some allowance is made for time-varying selectivity even when the assumed

selectivity pattern is mis-specified, which is consistent with prior work on age-structured models (Martell and Stewart, 2014). Moreover, there is no major disadvantage (bias, precision) in estimation performance when allowance is made for time-varying selectivity in the estimation method, even when selectivity is actually time-invariant. The performance difference between estimation methods with or without time-varying selectivity increases with the extent selectivity changes in the operating model.

Surprisingly, estimation methods using time blocking perform just as well as those using random parameters. Based on the worst MAREs, EM_{block} has the lowest value for one out of the three management quantities and is only 0.01 percent larger than a second one. In addition, EM_{block} never has the largest worst MARE. This supports that EM_{block} (even with an arbitrary choice of blocks) might be sufficiently flexible to capture the impacts of time-varying selectivity. An advantage of the EM_{block} is that it reduces the number of estimated parameters compared to EM_{RP}. Fewer estimated parameters reduce the variance of estimated quantities, particularly when there is substantial autocorrelation in the selectivity parameters. In addition, the parsimonious parameterization associated with EM_{block} reduces the computational demands associated with conducting an assessment when using EM_{RP}.

3.5.2 Likelihood functions in relation to time-varying selectivity.

No likelihood function performs best for all operating models. The three likelihoods perform similarly in terms of the relative error in spawning stock biomass/mature male biomass. Choosing between the likelihoods is difficult. Arguments can be made for any one of them. The multinomial has the most best and fewest worse MARE values. However, it never has the lowest worst MARE and it is unable to estimate the effective sample size as part of the model fitting process. The multivariate normal can estimate R_0 slightly better than other likelihoods when the generating function is Dirichlet-multinomial and estimate effective sample size regardless of the generating

functions although accuracy varies. However, the multivariate normal is the most prone to retrospective patterns. The multivariate normal is also the most computationally demanding as it requires the variance-covariance matrix to be inverted for each function call. The Dirichlet-multinomial has two of the three lowest worst MARE values but has difficulty estimating effective sample size unless the generating function is also Dirichlet-multinomial.

If we had to choose, we would recommend the multinomial because it has the most best and fewest worse MARE values and was not prone to retrospective patterns for spawning stock biomass. The multinomial is never the least worst MARE but it is the best MARE the majority of the time. Although we recommend the multinomial likelihood, we acknowledge an argument can be made for either the Dirichlet-multinomial or the multivariate normal.

3.5.3 Future work and caveats

The EM_{block} approach used in the first simulation study arbitrarily assigns each block a length of five years. Alternative block lengths, such as seven or ten years, are not explored to determine if longer blocks could adequately model time-varying selectivity. Longer blocks would further reduce the number of estimated parameters compared to random parameters and hence further reduce the variance in estimated quantities, but perhaps at the cost of increased bias. Most stock assessments that use EM_{block} try to determine the block size based on changes in fishing practices. Further comparisons between EM_{block} and EM_{RP} would help determine if basing EM_{block} on fishing practices is better than having blocks of similar lengths.

As for the second simulation study, Maunder (2011) describes likelihood functions for size composition data in addition to those considered in this paper. These include a method from Fournier et al. (1990) that is a normal approximation of the multinomial distribution whose variance is proportional to $\hat{P}_i(1-\hat{P}_i)$ and a method from Punt and Kennedy (1997) that is a lognormal likelihood with a variance proportional to $1/\hat{P}_i$. Further analyses should be conducted

examining these and other combinations of commonly used generating and likelihood functions to determine which likelihood function is the most robust when different generating functions are used.

3.5.4 Conclusion and implications for the next-gen stock assessments

The results from the simulation studies show that it is better in terms of estimation performance to include some version of time-varying selectivity in assessments than not. The blocking method appears to be a sufficient alternative to random parameters however this has only been tested for when the time period considered in the assessment is divided into equal-sized time blocks. As for the likelihood function choice, the results show that all three likelihoods perform similarly. Though we recommend using the multinomial likelihood, an argument can also be made for the Dirichlet-multinomial or multivariate normal.

3.6 Acknowledgements

This publication was [partially] funded by the Joint Institute for the Study of the Atmosphere and Ocean (JISAO) under NOAA Cooperative Agreement NA15OAR4320063, Contribution No. 2020-1130

Table 3.1. Values for the biological parameters of the operating models and whether they are estimated in the estimation methods.

Parameter	Operating Model Value	Estimation Method Treatment
Natural mortality, *M*	$0.23yr^{-1}$	Assumed known
Male Parameters		
Growth		
$\tilde{a}$	16.2 mm	Estimated
$\tilde{b}$	0.9	Estimated
CV of growth increment	0.25	Estimated
Length-weight relationship, $\alpha_{M,1}$	$1.6x10^{-4}$ gm/mm^3	Assumed known
Length-weight relationship, $\beta_{M,1}$	3.1	Assumed known
τ	0.2	Assumed known
Mat50	104.4	Assumed known
Female Parameters		
Growth		
$\tilde{a}$	17.9 mm	Estimated
$\tilde{b}$	0.9	Estimated
Standard error of growth increment	0.25	Estimated
Length-weight relationship, $\alpha_{M,2}$	$6.4x10^{-4}$ gm/mm^3	Assumed known
Length-weight relationship, $\beta_{M,2}$	2.8	Assumed known
τ	0.2	Assumed known
Mat50	74.6	Assumed known
Distribution of length-at-age 0, α	7.24	Assumed known
Distribution of length-at-age 0, β	5.04	Assumed known
κ	1.13	Assumed known
R_{50}	122.5	Assumed known
R_0	e^9	Estimated
Recruitment variation, σ_R	0.57	Assumed known
Selectivity		
Directed Fishery	Figure 3.1a	Estimated*
Snow Crab Fishery	Figure 3.1a	Estimated
Survey Fleet	Figure 3.1a	Estimated

* Multiple options considered for the functional form and hence number of parameters.

Table 3.2. Specifications for the data generated from the operating model.

Data Type	Years	CV or Effective sample Size
Catch (directed fishery, only males retained)	1968 – 2015	Known
Discard (directed fishery, males and females)	1968 – 2015	CV = 0.20
Size-Composition (directed fishery)	1968 – 2015	Sample Size = 100
Size-Composition (discard, directed fishery)	1992 – 2015	Sample Size = 25
Discard (snow crab fishery, males and females)	1977 – 2015	CV = 0.20
Size-Composition (snow crab fishery, males and females)	1992 – 2015	Sample Size = 25
Size-Composition (survey, males and females)	1976 – 2016	Sample Size = 100
Index	1975 – 2015	CV = 0.20
Tagging	2005	Sample Size = 1,000

Table 3.3. The operating model variants used in the two simulation studies

Simulation Study	Selectivity specification for the directed fishery	Parameters defining variation in selectivity ($\sigma^2_{\delta\theta}$, $\sigma^2_{\delta L50}$)	Distribution function for the size-composition data
1	Time-Invariant	N/A	Multinomial
1	Low random walk	(0.001, 1)	Multinomial
1	Medium random walk	(0.003, 2)	Multinomial
1	High random walk	(0.005, 3)	Multinomial
1	Very high random walk	(0.007, 4)	Multinomial
2	Time-Invariant	N/A	Multinomial
2	Medium random walk	(0.003, 2)	Multinomial
2	Time-Invariant	N/A	Dirichlet-multinomial
2	Medium random walk	(0.003, 2)	Dirichlet-multinomial
2	Time-Invariant	N/A	Multivariate normal
2	Medium random walk	(0.003, 2)	Multivariate normal

Table 3.4. Median relative errors (MREs) and median absolute relative errors (MAREs), expressed as percentages, for the simulation study exploring time-varying selectivity. The three bold values in each column of the MARE section represent the worst (largest) MARE value for each estimation method across operating models. The underlined bold value is the least worst MARE value out of the three bold values, thus the least worst estimation method for each management quantity.

Run		MRE			MARE		
Operating Model	**Estimation Method**	SSB_0	SSB_{2016}/SSB_0	SSB_{2016}/SSB_{1968}	SSB_0	SSB_{2016}/SSB_0	SSB_{2016}/SSB_{1968}
Time-Invariant	EM_{invar}	3.18	3.56	8.54	5.69	7.48	11.05
	EM_{block}	3.18	3.53	8.18	5.73	7.32	11.24
	EM_{RP}	3.17	3.21	10.59	6.22	7.43	12.14
Low Time-Varying	EM_{invar}	4.95	1.29	5.84	7.23	9.07	9.64
	EM_{block}	5.15	1.44	4.96	7.19	9.39	9.20
	EM_{RP}	5.40	1.50	5.93	7.40	9.09	9.95
Medium Time-Varying	EM_{invar}	5.40	3.03	7.44	6.79	7.61	14.19
	EM_{block}	5.06	2.44	9.39	6.74	7.54	**<u>12.29</u>**
	EM_{RP}	5.28	2.12	10.30	6.48	7.68	**12.82**
High Time-Varying	EM_{invar}	6.60	2.64	5.71	**<u>7.45</u>**	**10.55**	15.29
	EM_{block}	5.96	-0.09	6.01	7.19	9.12	9.52
	EM_{RP}	6.42	-0.62	7.19	7.43	9.28	10.89
Very High Time-Varying	EM_{invar}	5.13	3.00	5.74	7.29	10.42	**18.61**
	EM_{block}	5.58	0.62	6.50	**7.62**	**9.74**	9.56
	EM_{RP}	5.97	0.36	7.39	**7.89**	**<u>9.73</u>**	10.61

Table 3.5. Median relative errors (MREs) and median absolute relative errors (MAREs), expressed as percentages, for the simulation study exploring different likelihood functions in relation to time-varying selectivity. Under "Run", the first letter indicates how selectivity is modeled: time-invariant (I) or time-varying (V). The second set of letters indicates the generating and likelihood function: multinomial (M), Dirichlet-multinomial (D) or multivariate-normal (MN). The six bold values in each column of the MARE section represents the worst (largest) MARE value for each estimation method across operating models. The underlined bold value is the least worst MARE value out of the three bold values, thus the least worst estimation method for each management quantity.

Run		MRE			MARE		
Operating Model	**Estimation Method**	SSB_0	SSB_{2016}/SSB_0	SSB_{2016}/SSB_{1968}	SSB_0	SSB_{2016}/SSB_0	SSB_{2016}/SSB_{1968}
I-M	I-M	3.18	3.56	8.54	5.69	7.48	11.05
	I-D	3.39	3.42	8.81	5.91	7.50	10.77
	I-MN	2.04	5.09	10.37	6.09	9.16	11.58
	V-M	3.17	3.21	10.59	6.22	7.43	12.14
	V-D	3.34	3.04	10.70	6.31	7.46	12.58
	V-MN	2.25	4.22	9.15	5.93	8.74	11.67
I-D	I-M	2.74	4.82	9.20	6.16	8.50	12.24
	I-D	3.56	3.08	9.11	7.08	7.36	11.30
	I-MN	3.03	4.35	10.03	6.90	8.95	13.39
	V-M	2.76	3.49	10.27	6.55	7.99	13.15
	V-D	4.35	1.37	10.17	6.94	7.20	12.82
	V-MN	3.71	3.25	9.84	6.91	8.57	13.27
I-MN	I-M	5.10	3.09	8.67	6.36	7.25	9.88
	I-D	5.92	3.60	9.68	6.77	7.43	10.65
	I-MN	1.50	4.62	7.23	6.22	10.45	10.14
	V-M	4.24	2.30	10.38	6.26	7.63	10.99
	V-D	5.25	2.44	10.72	6.81	7.55	11.59
	V-MN	1.81	4.21	11.00	6.22	10.49	13.22
V-M	I-M	5.40	3.03	7.44	6.79	7.61	**14.19**
	I-D	5.54	2.72	7.73	6.73	7.46	**13.73**
	I-MN	4.68	3.05	10.17	6.56	9.16	**14.55**
	V-M	5.28	2.12	10.30	6.48	7.68	12.82
	V-D	5.55	1.89	10.04	6.64	7.45	**13.32**
	V-MN	3.84	2.03	11.27	6.69	9.82	13.42
V-D	I-M	5.23	4.39	6.26	7.45	9.51	13.84
	I-D	7.51	1.50	5.77	8.13	8.12	11.90
	I-MN	5.19	3.36	9.17	**7.42**	8.80	13.88
	V-M	5.53	1.51	9.37	**8.33**	**9.25**	**13.58**
	V-D	7.05	0.62	9.15	8.22	8.66	13.23
	V-MN	5.65	0.16	7.65	**8.41**	8.00	**14.01**
V-MN	I-M	6.56	2.40	5.68	**7.49**	**10.03**	11.61
	I-D	7.52	3.00	6.87	**8.27**	**9.48**	11.19
	I-MN	2.91	2.79	2.14	6.04	**11.81**	13.01
	V-M	6.30	0.83	3.79	8.21	8.38	10.47
	V-D	7.25	0.54	3.78	**8.34**	**8.99**	10.68
	V-MN	2.85	0.46	4.67	6.30	**10.51**	11.44

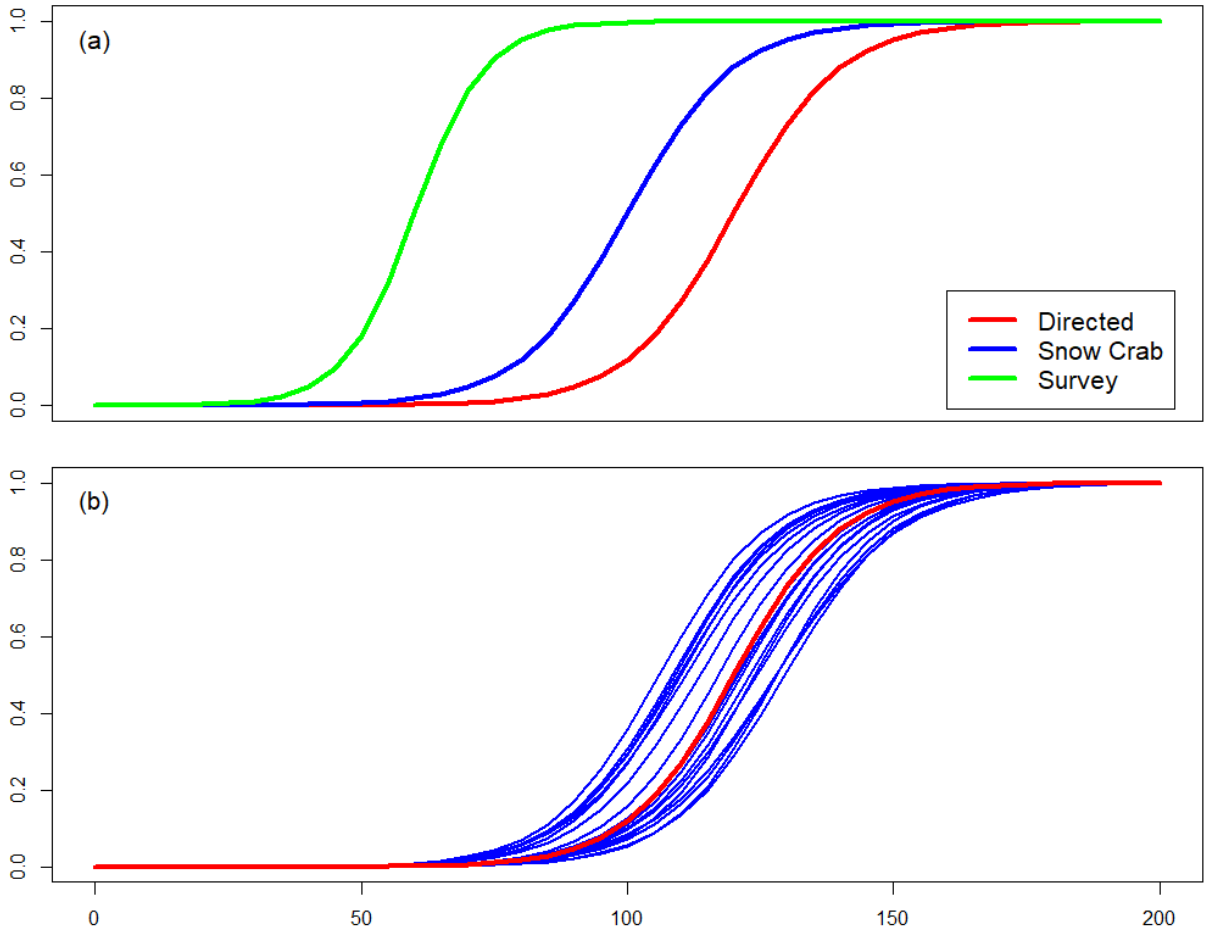

Figure 3.1. Initial selectivity curves for the directed (red), snow crab (blue) and survey (green) fleets (a), and an example of time-varying selectivity for the directed fishery (b). The red line (b) is the initial selectivity curve, while the blue lines represent how selectivity varies through time under the medium level of time-varying selectivity.

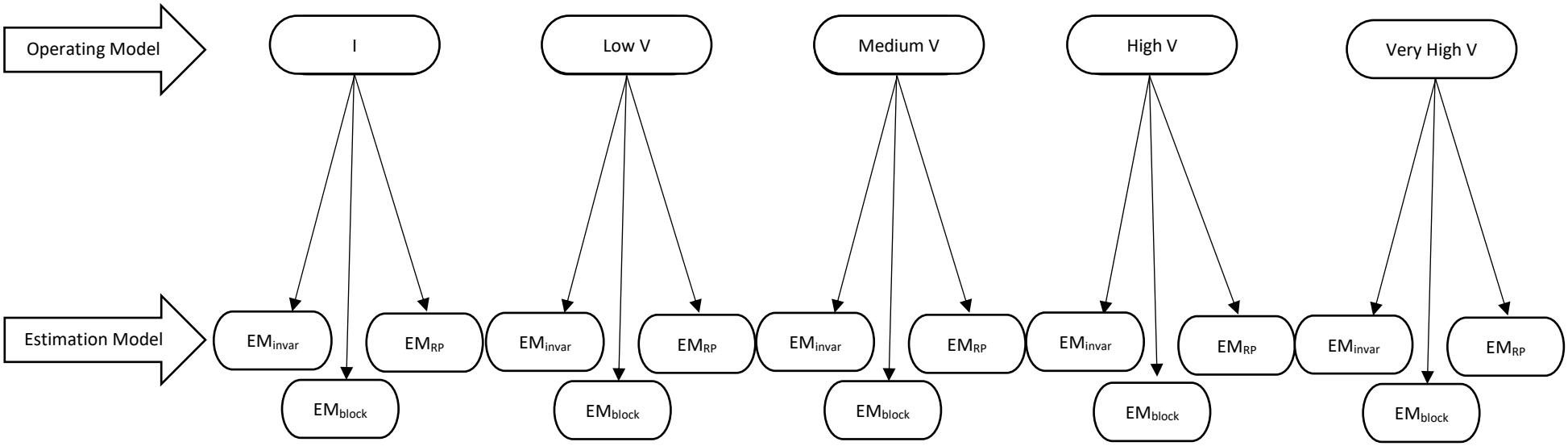

Figure 3.2. A visual representation of the simulation study exploring how best to model time-varying selectivity. Each oval specifies how selectivity is modeled in the operating model and estimation method. 'I' denotes time-invariant selectivity, 'V' denotes time-varying selectivity. The three estimation methods are: time-invariant (EM_{invar}), blocking (EM_{block}) and random parameters (EM_{RP}).

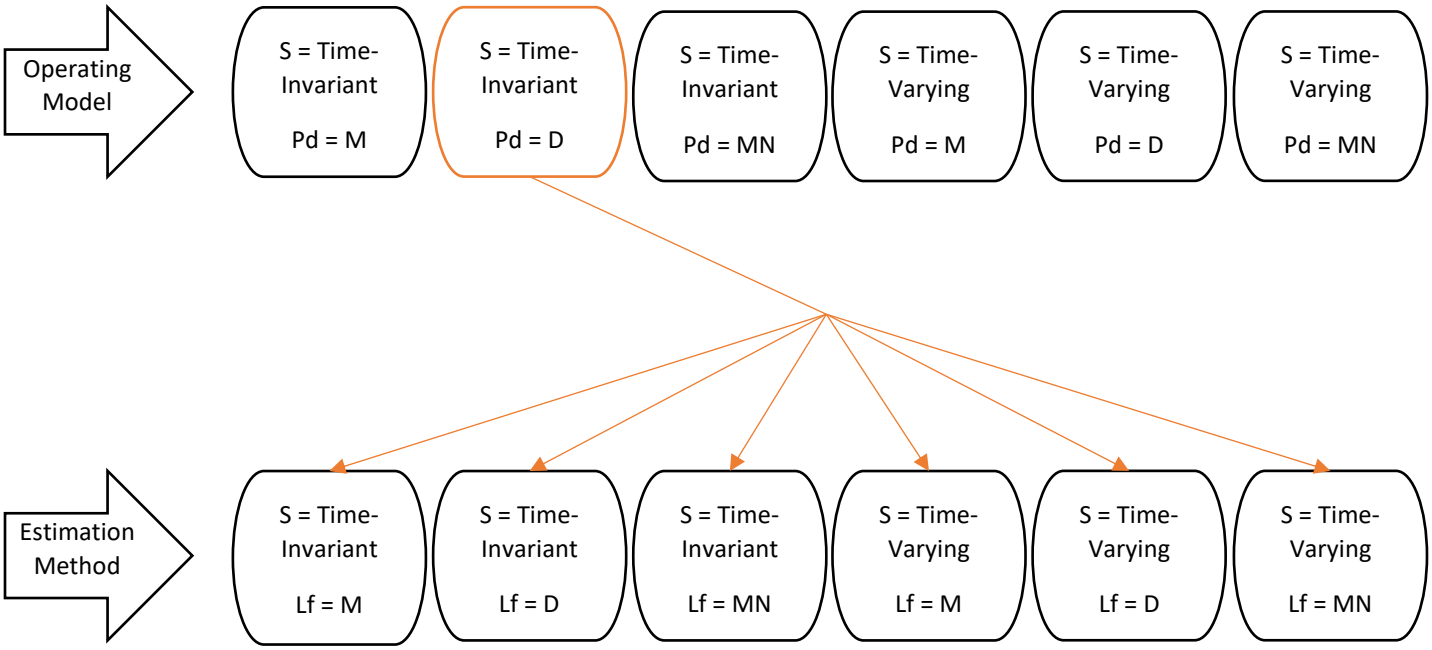

Figure 3.3. A visual representation of the simulation study examining the effects of various likelihood functions for the size-composition data given time-varying selectivity. Pd is the probability distribution used to generate the size-composition data with options M (multinomial), D (Dirichlet-multinomial) and MN (multivariate normal). S is the selectivity method with options Time-Invariant or Time-Varying. Lf refers to the likelihood function, with the same options as the probability distribution. The orange oval and lines is an example of how each generated data set is analyzed using the estimation methods.

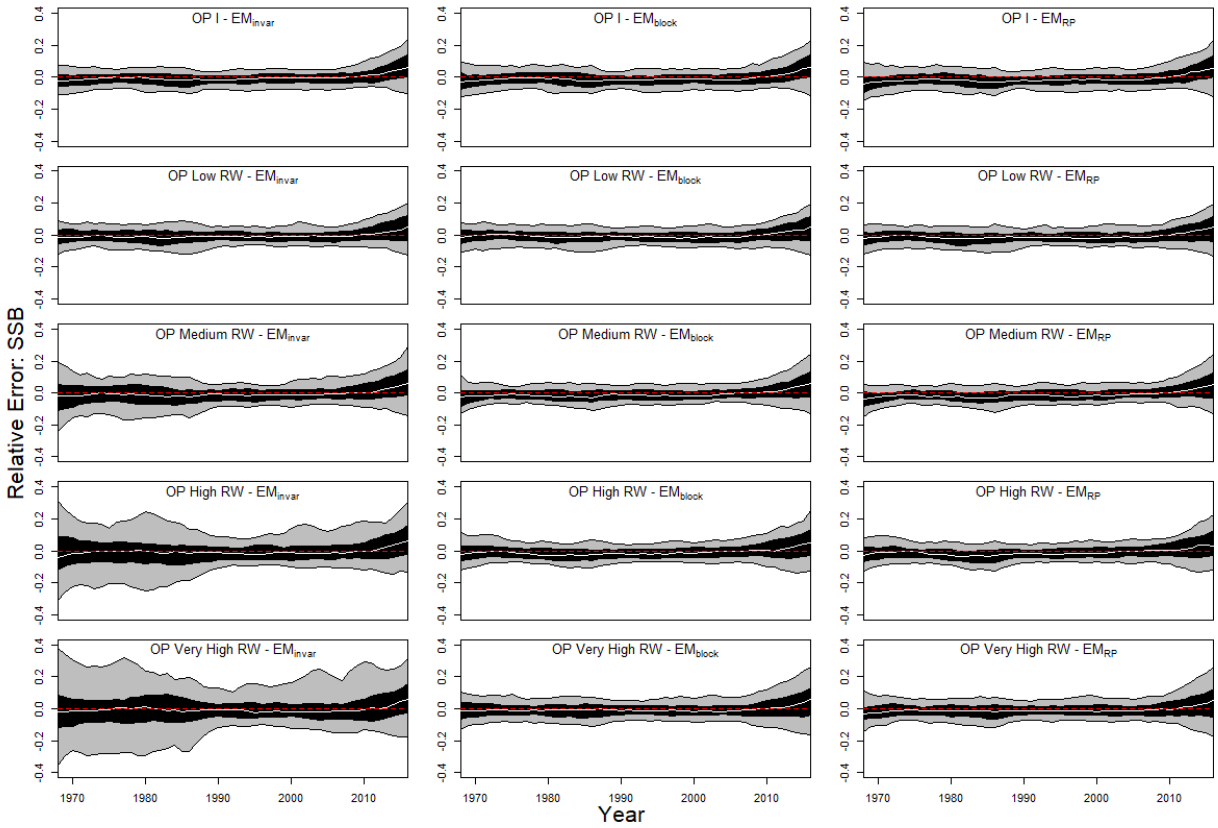

Figure 3.4. Relative error for spawning stock biomass (SSB) for the simulation study exploring how best to model time-varying selectivity. The five operating models (OP) are: time-invariant (I), low random walk (Low RW), medium random walk (Medium RW), high random walk (High RW) and very high random walk (Very High RW). The three estimation methods are: time-invariant (EM_{invar}), blocking (EM_{block}) and random parameters (EM_{RP}). The white line is the median relative error, the red dash line is the zero line, the black shaded region is the 50% quantile and the grey shaded region is the 90% quantile.

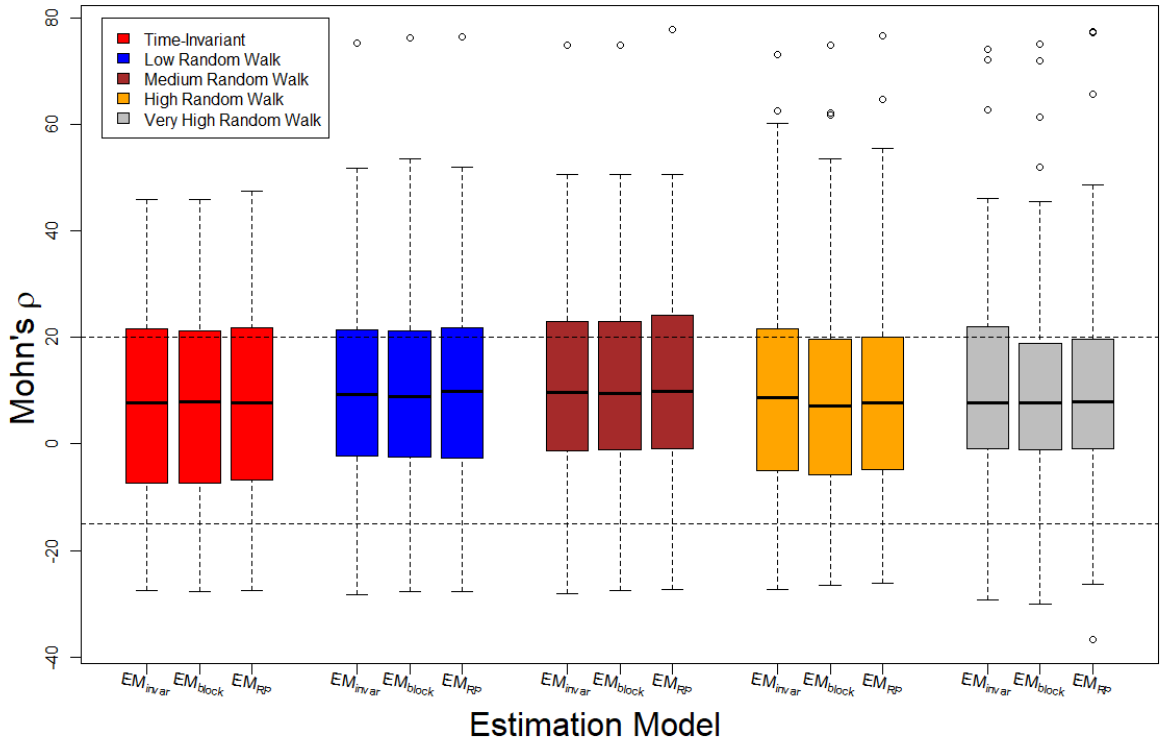

Figure 3.5. A boxplot of Mohn's ρ for recruitment times 100 from the study exploring how best to model time-varying selectivity. The colors represent the operating model scenario. The x-axis indicates the estimation method (EM_{invar} – time-invariant, EM_{block} – blocking, EM_{RP} – random parameters). The dashed lines are the bounds for Mohn's ρ that do not suggest retrospective patterns.

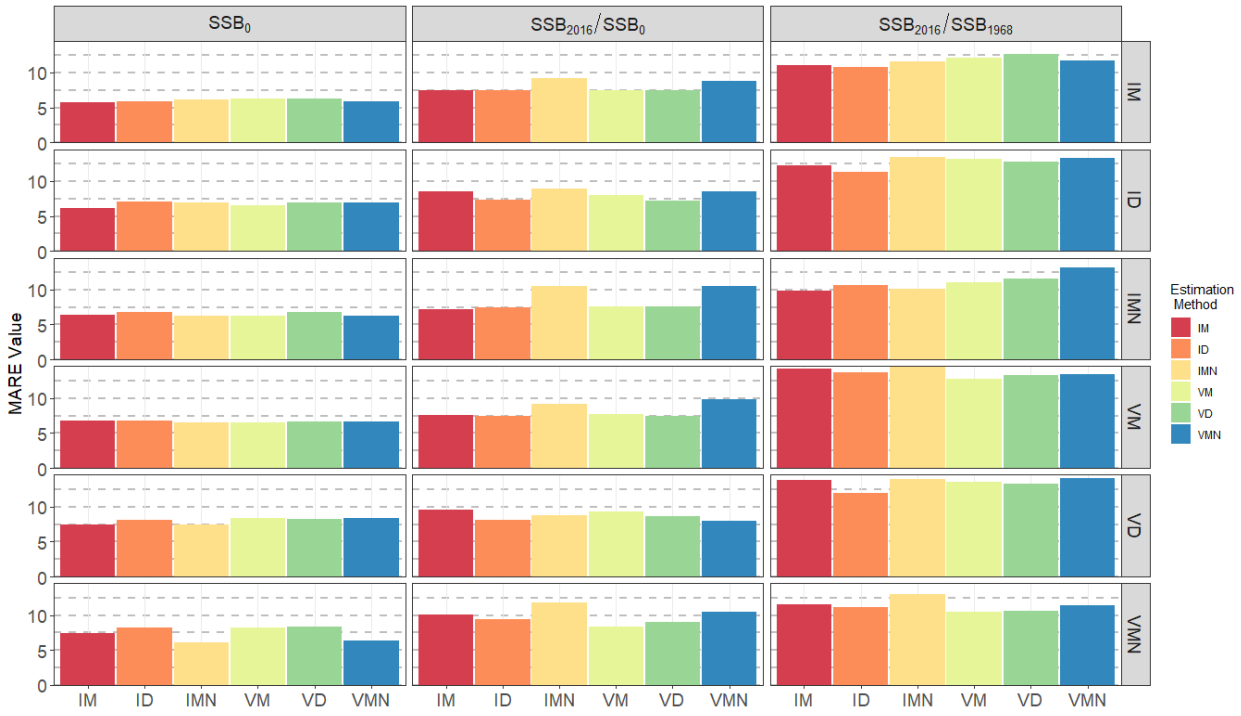

Figure 3.6. Median absolute relative errors (MAREs), from Table 3.5, for the simulation study exploring different likelihood functions in relation to time-varying selectivity. Each row is based on a different operating model. The first letter indicates how selectivity is modeled: time-invariant (I) or time-varying (V). The second set indicates the generating and likelihood function: multinomial (M), Dirichlet-multinomial (D) or multivariate-normal (MN). The columns are different performance metrics.

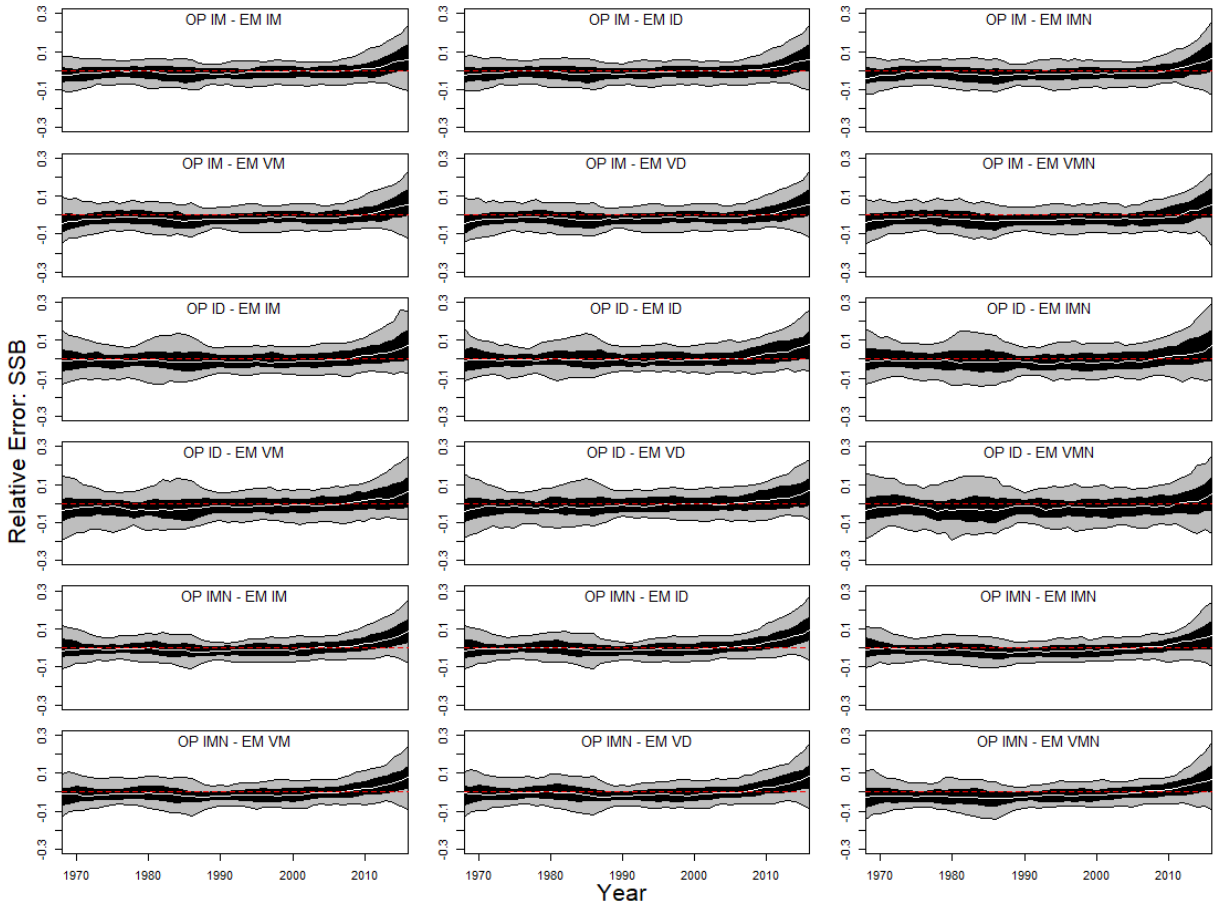

Figure 3.7. Relative errors for spawning stock biomass (SSB) for the simulation study examining likelihood functions in relation to time-varying selectivity when the operating model (OP) has time-invariant (I) selectivity. The generating functions are either multinomial (M), Dirichlet-multinomial (D), or multivariate-normal (MN). The estimation methods (EM) assume either time-invariant (I) or time-varying (V) selectivity. The likelihood functions for the estimation methods are either multinomial (M), Dirichlet-multinomial (D), or multivariate-normal (MN). The white line is the median relative error, the red dash line is the zero line, the black shaded region is the 50% quantile and the grey shaded region is the 90% quantile.

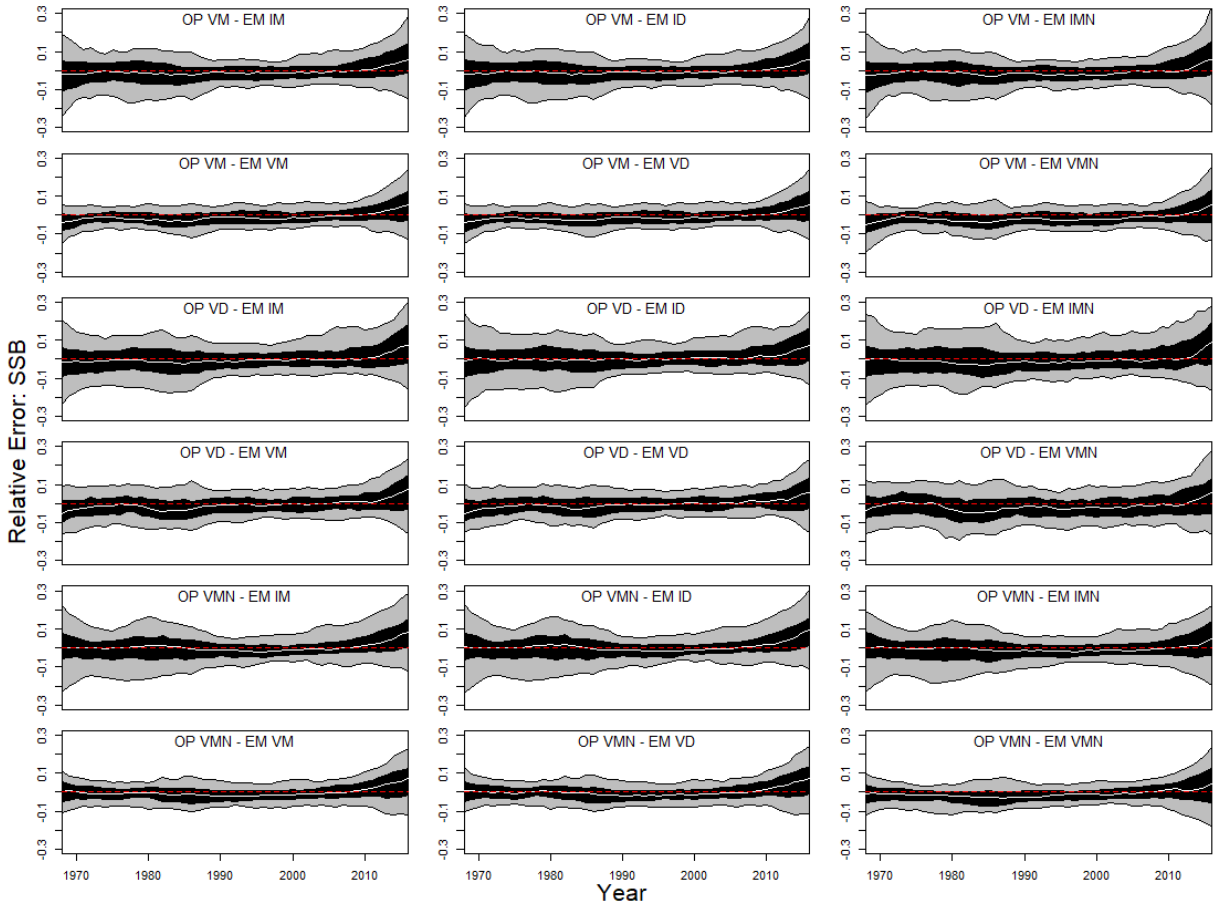

Figure 3.8. The same as Figure 3.7 except the operating model has time-varying (V) selectivity.

Chapter 4 : Factors Influencing Size-Structured Models' Ability to Estimate Natural Mortality

4.1 Abstract

Natural mortality (*M*) is crucially important for stock assessment since it strongly influences estimates of spawning stock biomass, MSY and fishing mortality. Variation in *M* can occur between sex, maturity stages and temporally. Estimates of *M* are confounded with those of catchability, recruitment, and growth. Previous studies, which focused on age-structured population dynamics models, suggest that it would be better to estimate *M* rather than pre-specify it. Size-structured population dynamics models, used for species that are hard to age such as crabs, have additional unique characteristics such as a size-transition matrix and possibility a terminal molt, that could impact the ability to estimate *M*. Therefore, this study explores the ability of size-structured stock assessments models to estimate *M* for several scenarios. These scenarios include whether the operating model and estimation method have a terminal molt, whether *M* is constant, time-varying, or sex- and maturity stage-specific, the quality/sample size of the biomass index and the size-composition data and whether growth is estimated simultaneously with *M*. The results show that size-structured models can estimate time-varying, sex- and maturity stage-specific *M* when the estimation method mirrors the operating model. Surprisingly, terminal molt does not affect the ability to estimate *M*. However, estimating growth simultaneously with *M* has a negative impact on the ability to estimate *M* but a positive effect on the quality of the estimates for spawning stock biomass.

4.2 Introduction

Commercial fisheries are a valuable industry in the USA and around the world. Many fished species are managed based on stock assessment outcomes such as estimates of spawning stock

biomass (SSB), the status of the stock relative to management reference points, and the fishing mortality corresponding to maximum sustainable yield (F_{MSY}). A crucial component of the models on which stock assessments are based is mortality, which is divided into fishing and natural. Natural mortality (*M*) accounts for all mortality that is not directly associated with fishing. *M* strongly influences the outputs from a stock assessment since it directly impacts the productivity of the stock (Clark, 1999; Kenchington, 2014; Punt et al., 2021). It also impacts target fishing mortality rates since these often scale with *M* (Andrews and Mangel, 2012; Punt et al., 2021). Thus, inaccurate *M* values can lead to over- or underestimation of important management quantities such as SSB, MSY and depletion (Thompson, 1994; Punt and Walker, 1998).

M is a difficult parameter to estimate since it is hard to separate from total mortality (Vetter, 1988; Beverton and Holt, 1957). It can also be confounded with catchability, fishing mortality and recruitment (Schnute and Richards, 1995). For this reason, many stock assessments pre-specify *M* to an externally estimated value or to an estimate of *M* for a species with similar biological characteristics (Hewitt et al., 2007; Wang and Ellis, 2005). Sensitivity analysis can be used to explore the consequences of uncertainty in the pre-specified value for *M*, although its usefulness can be limited (Punt and Hilborn, 1997). Over the years, stock assessments have integrated more data sources, potentially allowing for better estimation of *M*. For this reason, and because of *M*'s major influence on management quantities, studies (e.g., Johnson et al., 2015; Lee et al., 2011; Punt et al., 2021) are increasingly suggesting that it is better to estimate *M* within stock assessments than fix it.

The estimation of *M* can become increasingly more complicated when it varies by sex, age/size maturity or temporally. Variation in *M* by sex can occur when there is some degree of sexual dimorphism (e.g., Murphy et al., 2018; Windsland, 2015) while age/size and maturity variation

result from changes in the level of predation as individuals grow/mature (Lorenzen, 1996; Vetter 1988). As a result, many assessments, such as those for Tanner crab (*Chionoecetes bairdi*) and arrowtooth flounder (*Atheresthes stomias*) in the North Pacific, estimate separate *M*'s by sex and maturity (Spies et al., 2019; Stockhausen, 2019). Temporal variation in *M* can be caused by changes in predator/prey abundance and/or environmental conditions (Swain, 2011; Tyrrell et al., 2011). Studies exploring this variation suggest estimating time-varying *M* only when evidence supports such variation (Fu and Quinn, 2011; Kanaiwa et al., 2008).

Most studies exploring when and how well *M* can be estimated have focused on age-structured population dynamics models, because these are commonly used in stock assessments (Punt et al., 2016). Unfortunately, there are many marine species, such as crabs and lobsters, which are difficult to age. These species are often described using size-structured population dynamics models that do not use age data for parameter estimation. They also have unique characteristics such as a size-transition matrix. With a focus on growth, a size-structured model's ability to estimate *M* would likely be impaired since studies have shown that growth and *M* covary (e.g., Gislason et al., 2010; Chu et al., 2008). However, some species assessed using size-structured models, such as Tanner crab in the eastern Bering Sea, exhibit a terminal molt whereby individuals stop growing once they mature (Stockhausen, 2019). This could potentially improve estimation of *M* because there is no further growth following terminal molt, thereby reducing the confounding relationship between growth and natural mortality.

Size-structured population dynamics models have characteristics that could potentially impact the ability to estimate *M*. Therefore, this study investigates the ability of stock assessments based on size-structured models to estimate *M* using a simulation study where *M* is either constant, varying by sex and maturity stage, or time-varying. The operating models and estimation methods

are based on eastern Bering Sea (EBS) Tanner crab. We also explore the impact of terminal molt, having growth estimated or pre-specified correctly, and the quality of the available data. Model performance is determined through its ability to estimate *M*, the parameters determining growth, and quantities of interest to management, specifically spawning stock biomass.

4.3 Methods

A simulation study has two components: the operating model and the estimation method. The operating model, which generates data sets, represents the real world. There are multiple versions of the operating model given uncertainty about real world dynamics. The estimation methods estimate the parameters in the population dynamics model using the generated data sets and differ in terms of their parameterization (and perhaps population model and data). The aim of a simulation study is to determine which estimation method performs best at estimating parameters such as *M* and quantities of management interest given the set of operating models.

4.3.1 Operating model

The operating model, a two-sex size-structured population dynamics model with 32 size-classes, is based on a simplified version of the actual stock assessment for EBS Tanner crab. There are several differences between the operating model and the actual stock assessment model. For example, the operating model generates tag-recapture data while the actual assessment uses molt increment data. Each size-class is 5mm wide, with the smallest size-class beginning at 25mm and the final size-class being a plus group, i.e., all individuals larger than the lower limit of final size-class (180mm) are placed within the last size-class. The actual stock assessment for EBS Tanner crab has four fishery fleets, but only the directed fishery and the snow crab fishery take an appreciable number of Tanner crabs. Consequently, the operating model includes three ‘fleets’: two fishery fleets (directed and snow crab), and a survey. Males can be retained or discarded

(usually the smaller ones) in the directed fishery while all females and all animals captured in the snow crab fishery are discarded.

There are two versions of the operating model. One has individuals grow until they die and other has terminal molt in which individuals stop growing once mature. For the operating model with no terminal molt, the population dynamics are governed by:

$$N_{y,s,m,i,1} = \begin{cases} \sum_{j \le i} X_{s,j,i}\left(N_{y-1,s,1,j,1}e^{Z_{y-1,s,1,i}} + (N_{y-1,s,2,j,1}Pmat_{s,1,j,i} + N_{y-1,s,2,j,0}Mat_{s,j})e^{Z_{y-1,s,2,i}}\right), m = 1\ (mature) \\ \sum_{j \le i} X_{s,j,i}\left(N_{y-1,s,2,j,1}Pmat_{s,2,j,i} + N_{y-1,s,2,j,0}(1 - Mat_{s,j})\right)e^{Z_{y-1,s,2,i}}, m = 2\ (immature) \end{cases} \quad (4.1a)$$

$$N_{y,s,2,i,0} = R_{y,s,i} \quad (4.1b)$$

while the operating model with terminal molt:

$$N_{y,s,m,i,1} = \begin{cases} N_{y-1,s,1,i,1}e^{Z_{y-1,s,1,i}} + \sum_{j \le i} X_{s,j,i}\left(N_{y-1,s,2,j,1}Pmat_{s,1,j,i} + N_{y-1,s,2,j,0}Mat_{s,j}\right)e^{Z_{y-1,s,2,i}}, m = 1\ (mature) \\ \sum_{j \le i} X_{s,j,i}\left(N_{y-1,s,2,j,1}Pmat_{s,2,j,i} + N_{y-1,s,2,j,0}(1 - Mat_{s,j})\right)e^{Z_{y-1,s,2,i}}, m = 2\ (immature) \end{cases} \quad (4.1c)$$

$$N_{y,s,2,i,0} = R_{y,s,i} \quad (4.1d)$$

where $N_{y,s,m,i,1}$ is the number of animals at the start of year *y* of sex *s* in maturity state *m* (1=mature; 2=immature) and size-class *i* that have been in the fished population for at least a year, $N_{y,s,2,i,0}$ is the number of animals at the start of year *y* of sex *s* in maturity state 2 (immature) and size-class *i* that have been in the fished population for less than a year. We assume all individuals entering the fished population are immature. $X_{s,j,i}$ is the probability of individuals of sex *s* transitioning from size-class *j* to size-class *i* over a year (assumed time-invariant). $Pmat_{s,m,j,i}$ is the proportion of immature individuals of sex *s* that grow from size-class *j* to size-class *i* and transition to/stay in maturity state *m.* This variable ensures that the population follows a pre-specified maturity curve. $R_{y,s,i}$ is the number of immature recruits in size-class *i* of sex *s* added to the population at the end of year *y*, and $Z_{y,s,m,i}$ is the total mortality during year *y* for animals of sex *s* in maturity state *m* in size-class *i*. Equations 4.1b and 4.1d imply that new recruits enter the population as a separate

category ($N_{y,s,2,i,0}$) from the rest of the population. This is because all new recruits are immature. After a year, the recruits mature with probability $Mat_{s,i}$ (Equation 4.3) to mirror the proportion mature in the population.

The value for $Pmat_{s,m,j,i}$, which ensures that the population follows the maturity curve, is determined by:

$$Pmat_{s,1,j,i} = \begin{cases} 0\,, & i = j \\ \frac{Mat_{s,i} - Mat_{s,j}}{(1 - Mat_{s,i})}\,, & i \neq j \end{cases} \tag{4.2a}$$

$$Pmat_{s,2,j,i} = \begin{cases} 1\,, & i = j \\ 1 - Pmat_{s,1,j,i}\,, & i \neq j \end{cases} \tag{4.2b}$$

where $Mat_{s,i}$ is the proportion of the animals of sex *s* in size-class *i* that are mature in an unfished population. It is assumed that individuals remaining in the same size-class after one year did not molt and therefore could not mature and reach terminal molt. $Mat_{s,i}$ is a logistic function:

$$Mat_{s,i} = \frac{1}{1 + e^{-\omega_s(\bar{L}_i - M50_s)}} \tag{4.3}$$

where ω_s is the slope parameter for sex *s*, $M50_s$ is the inflection point for sex *s* and $\bar{L}_i$ is the midpoint of size-class i. Equations 4.2a and 4.2b do not ensure that the population follows the maturity curve when there is terminal molt. Details about Equations 4.2a and 4.2b and its limitations are provided in Appendix C. However, given that the estimation methods and operating models use the same forms for the population dynamics equations, this issue will not lead to bias.

The number of recruits is given by:

$$R_{y,s,i} = R_0 e^{\varepsilon_y - \sigma_R^2/2} P(i|a = 0) \tag{4.4}$$

where R_0 is the expected number of recruits produced at unfished equilibrium, ε_y is the deviation during year *y* about the expected ln-scale recruitment, σ_R is the degree of variation about expected ln-scale recruitment, and $P(i|a = 0)$ is the proportion of age-0 individuals of sex *s* that recruit to

size-class *i*. Recruitment is assumed to be independent of spawning stock biomass, which is consistent with the assumption made by the EBS Tanner crab assessment (Stockhausen, 2020). $P(i|a=0)$ is a gamma distribution integrated over each size-class range, approximated by:

$$P(i|a=0) = \frac{\bar{L}_i^{\alpha-1} e^{-\frac{\bar{L}_i}{\beta}}}{\sum_j \bar{L}_j^{\alpha-1} e^{-\frac{\bar{L}_j}{\beta}}} \tag{4.5}$$

where α and β are the parameters of the gamma distribution. New recruits are all immature and only join the first 10 size-classes i.e., 25mm - 75mm.

A size-transition matrix (*X*) is used to determine the probability of an individual transitioning from one size-class to another (only increases in growth are allowed). The expected growth increment of an individual of sex *s* follows a linear relationship between growth increment and current size: with two parameters, the asymptotic height ($L_{\infty,s}$) and the slope (k_s).

$$\tilde{L}_j = \left(L_{\infty,s} - \bar{L}_j\right)(1 - e^{-k_s}) \tag{4.6}$$

where $\tilde{L}_j$ is the expected growth increment of an individual in size-class *j*, and $L_{\infty,s}$ and k_s are the growth parameters for individuals of sex *s*. This growth curve was chosen because it is commonly used in stock assessments (Punt et al., 2016). The size-transition matrix is determined by:

$$X_{s,j,i} = \begin{cases} \int_{-\infty}^{\bar{L}_i + \Delta L/2} P\left(l \middle| \bar{L}_j, L_{\infty,s}, k_s, \tilde{\sigma}_s\right) dl & if\ i = j \\ \int_{\bar{L}_i - \Delta L/2}^{\bar{L}_i + \Delta L/2} P\left(l \middle| \bar{L}_j, L_{\infty,s}, k_s, \tilde{\sigma}_s\right) dl & if\ i \neq j\ and\ i \neq x^L \\ \int_{\bar{L}_i - \Delta L/2}^{\infty} P\left(l \middle| \bar{L}_j, L_{\infty,s}, k_s, \tilde{\sigma}_s\right) dl & if\ i = x^L \end{cases} \tag{4.7}$$

where ΔL is the width of each size-class (5mm), $\tilde{\sigma}_s$ is the coefficient of variation of the growth increment, x^L is the maximum size-class and $P\left(l \middle| \bar{L}_j, L_{\infty,s}, k_s, \tilde{\sigma}_s\right)$ is the probability density function for size after growth for an individual of size *L*, i.e.:

$$P\left(l\middle|\bar{L}_j, L_{\infty,s}, k_s, \tilde{\sigma}_s\right) = \frac{1}{\sqrt{2\pi}\tilde{\sigma}_s\tilde{L}_j} exp\left[-\frac{\left(l-\tilde{L}_j\right)^2}{2\tilde{\sigma}_s^2\tilde{L}_j{}^2}\right] \quad (4.8)$$

The catch is divided into retained and discarded:

$$C^R_{y,f,s,i} = \sum_m \frac{F_{y,f}\Omega_{f,s,i}S_{f,s,i}}{Z_{y,s,m,i}} N_{y,s,m,i}(1 - e^{-Z_{y,s,m,i}}) \quad (4.9a)$$

$$C^D_{y,f,s,i} = \sum_m \frac{F_{y,f}[1-\Omega_{f,s,i}]S_{f,s,i}}{Z_{y,s,m,i}} N_{y,s,m,i}(1 - e^{-Z_{y,s,m,i}}) \quad (4.9b)$$

where $C^R_{y,f,s,i}$ is the retained catch during year *y* by fleet *f* of sex *s* individuals in size-class *i*, $C^D_{y,f,s,i}$ is the discarded catch during year *y* by fleet *f* of sex *s* individuals in size-class *i*, $S_{f,s,i}$ is the selectivity for fleet *f* on animals of sex *s* in size-class *i*, $F_{y,f}$ is the fully-selected fishing capture rate for fleet *f* during year *y*, $\Omega_{f,s,i}$ is the proportion of animals of sex *s* in size-class *i* caught by fleet *f* that are retained. The value for $S_{f,s,i}$ is determined using a logistic function:

$$S_{f,s,i} = \frac{1}{1+ e^{-\theta_{f,s}(\bar{L}_i - L50_{f,s})}} \quad (4.10)$$

where $\theta_{f,s}$ is the slope parameter for fleet *f* for individuals of sex *s*, and $L50_{f,s}$ is the inflection point for fleet *f* for individuals of sex *s*. A logistic function is also used to determine the value for $\Omega_{f,s,i}$ for the directed fishery:

$$\Omega_{dir,s,i} = \frac{1}{1+ e^{-\kappa_{dir,s}(\bar{L}_i - R50_{dir,s})}} \quad (4.11)$$

where $\kappa_{dir,s}$ is the slope parameter for the directed fishery for sex *s*, and $R50_{dir,s}$ is the size-at-50%-retention for the directed fishery for sex *s*. The values for $\Omega_{f,s,i}$ for females in the directed fishery are zero because the fishery is male-only, and the values for $\Omega_{f,s,i}$ for both sexes in the snow crab fishery are zero because the entire catch is discarded. The weights of the landed ($\tilde{C}^R_{y,f}$) and discarded ($\tilde{C}^D_{y,f}$) catch by fleet *f* during year *y* are:

$$\tilde{C}^{R}_{y,f} = \sum_s \sum_i C^{R}_{y,f,s,i} w_{s,i} \quad (4.12a)$$

$$\tilde{C}^{D}_{y,f} = \sum_s \sum_i C^{D}_{y,f,s,i} w_{s,i} \quad (4.12b)$$

where $w_{s,i}$ is the mean weight of individuals of sex *s* in size-class *i*. The value for $w_{s,i}$ is determined by:

$$w_{s,i} = \gamma_s \bar{L}_i^{\epsilon_s} \quad (4.13)$$

where γ_s and τ_s are parameters defining the length-weight relationship by sex.

Total mortality for each year accounts for natural mortality and fishing mortality:

$$Z_{y,s,m,i} = M_{y,s,m} + \sum_f F_{y,f}(\Omega_{f,s,i} + [1 - \Omega_{f,s,i}]\lambda_f) S_{f,s,i} \quad (4.14)$$

where $M_{y,s,m}$ is the natural mortality rate during year *y* for animals of sex *s* in maturity state *m* and λ_f is the mortality rate for individuals discarded by fleet *f*. There were three *M* scenarios in the operating model (Table 4.1): (a) time-, sex-, and maturity-stage-invariant *M* (denoted "Single"), which is the default for the EBS Tanner crab assessment, (b) sex- and stage-specific time-variant *M* (denoted "Sex +Stage") where separate values are set for adult males, adult females, and immature males/females, and (c) time-varying (but not sex- or stage-specific) *M* (denoted "Time-varying"), where there is a linear increase in *M* from the start to the end of the catch time series (1968 to 2016) with *M* set to a base value before fishery starts (Figure 4.1). A similar scenario was used by Deroba and Schueller (2013) to explore the impact of mis-specified age- and time-varying *M* in stock assessment models.

Table 4.2 lists the values for the biological and fishery parameters used in the operating models and whether these values are estimated by the estimation methods. The values were taken from the 2020 EBS Tanner assessment (Stockhausen, 2020). The operating models start the simulated Tanner crab population in 1947 at unfished equilibrium. The population is then projected with

random recruitment so that it is no longer at equilibrium by the start of the first year with catch (1968).

4.3.2 Data generation

Five data sets (retained catch, discarded catch, biomass index, size-composition and tag-recapture) are generated by the operating model. Table 4.3 lists the quantities that define how precise the various data sources are. The catch data, in kilograms, are generated for the two fishery fleets. Retained catches are assumed known without error while the discarded catches are not, because these are based on extrapolated observer reports. Estimates of discarded catches are log-normally distributed with a standard error in log-space of σ_d:

$$\tilde{C}_{y,f}^{D,obs} = \tilde{C}_{y,f}^{D} e^{\delta d_y - \sigma_d^2/2} \tag{4.15}$$

where $\tilde{C}_{y,f}^{D,obs}$ is the observed discard catch for fleet *f* during year *y*, δd_y is the observation error for the estimate of the discard catch or year *y* and σ_d is the standard error of δd_y.

The survey provides an index of the number of individuals by sex and maturity state, subject to log-normal error, i.e.:

$$N_{y,s,m}^{obs} = [\sum_i N_{y,s,m,i} S_{sur,i}] e^{\delta s_{y,s,m} - \sigma_{sur}^2/2} \tag{4.16}$$

where $N_{y,s,m}^{obs}$ is the observed survey index, $\delta s_{y,s,m}$ is the observation error for sex *s*, maturity state *m* and year *y*, and σ_{sur} is the standard error of $\delta s_{y,s,m}$. The benefits of improving the amount of information provided by the survey index is explored by lowering the "base" coefficient of variance from a "base" level to a "better info" level for several simulations (Table 4.3).

Four size-composition data sets are generated using multinomial distributions. Each data set distinguishes individuals by sex and maturity stage. The directed fishery has two size-composition data sets, one for the retained catch and another for the discarded catch. The retained catch only includes males. The remaining two size-composition data sets are for the snow crab and survey

fleets. Both include males and females. We explored the benefits of increasing the amount of information in the survey fleet by increasing the effective sample size for the size-composition data from the "base" level to "better info" for several simulations (Table 4.3).

As stated earlier, tag-recapture data are used instead of molt-increment data. This is because growth is commonly estimated using tag-recapture data (Punt et al., 2016). For the tag-recapture data, individuals are tagged during a survey conducted in 2005. A total of 1,000 individuals (males and females) are randomly sampled from the survey catch. The time-at-liberty for each individual is randomly chosen between 1-5 years. The probability of recapturing a tagged individual in size-class *i* that was initially in size-class *j* when the time-at-liberty equals *t* ($Pc_{t,j,i}$) is determined by:

$$Pc_{t,j,i} = \frac{X^t_{s,j,i}\, S_{1,y,i}}{\sum_i X^t_{s,j,i}\, S_{1,y,i}} \tag{4.17}$$

The number of tagged individuals released in size-class *i* and recaptured in size-class *j* are randomly chosen using a multinomial distribution with the vector of $Pc_{t,j,i}$ from size-classes *j* as the expected proportions.

4.3.3 Estimation methods

The basic structure of the estimation methods mirrors the operating models. The estimation method variants differ in how growth and *M* are modeled (Table 4.4). There are two options for growth in the estimation methods: estimated or pre-specified (to the true values). There are six options for how *M* is treated in the estimation method. Three involve pre-specifying *M* to either the base-*M* (P-1), half of the base-*M* (P-0.5) or 1.5 times the base-*M* (P-1.5). When the operating model is Sex + Stage, base-*M* is replaced with the average between *male-M* and *female-M*. The remaining three estimation methods estimate *M* as either a single value (E-M), values by sex and stage (E-SS) or as time-varying using a blocking design (E-B). E-SS estimates three values for *M* (for mature males, mature females and immature males/females). The blocking design for *M* divides 1968 -

2016 into ten five-year time blocks (the last is four years only), with *M* estimated for each block but constant within blocks. This is accomplished by estimating *M* for the first block then estimating deviations to this value for the remaining blocks:

$$M_{b,s,m} = M_{1,s,m} e^{\tau_b} \tag{4.18}$$

where τ_b is the estimated deviation for *M* for block *b*. A weak penalty is placed on the *M* deviations to limit how much *M* can vary over time:

$$\frac{\sum_y {\tau_b}^2}{2\sigma_\tau^2} \tag{4.19}$$

where σ_τ is the standard error for τ_b (set to 1). Table 4.5 lists the abbreviations for the six options.

4.3.4 Experimental design

Table 4.4 lists the operating model and estimation method variants explored in this paper. Two key simulation experiments are undertaken. To evaluate how model misspecification affect the ability to estimate *M* and spawning stock biomass, we set data quality to a base level (effective sample size 100, biomass index CV 0.2). All other factors in Table 4.4 are crossed except Time-varying with E-SS and Sex + Stage with E-B (i.e., data from an operating model with time-varying *M* are not analyzed using an estimation method with sex- and stage-specific time-invariant *M*, and data from an operating model with time-invariant sex- and stage-specific *M* are not analyzed using an estimation method with time-varying *M*). To evaluate how data quality affects the ability to estimate *M* and spawning stock biomass, the experiment crossed the "Single" and "Sex + Stage" variants of the operating model, the P-1, E-M, and E-SS estimation methods, and three levels of data quality ("Base", which matches that for Sim 1, "H-SS" with higher effective sample sizes, and "S-CV" with a lower biomass index CV).

The performances of the estimation methods are summarized in two ways. First, we examined the ability to estimate *M* and the three growth parameters L_∞, k, and $\tilde{\sigma}$ (when they are estimated),

and used median relative error (MRE) and median absolute relative error (MARE) to quantify estimation performance:

$$MRE = 100 * median_i\left(\frac{E_i - T_i}{T_i}\right) \quad (4.20a)$$

$$MARE = 100 * median_i\left(\frac{|E_i - T_i|}{T_i}\right) \quad (4.20b)$$

where E_i is the estimated value and T_i is the true value. We focus on the median instead of the mean because there are occasional outlying estimates. When the operating model is Time-varying, we compared the mean true *M* within each time-block to the E-B estimated *M* within the same time-block. MRE and MARE values are not calculated for E-M, P-1, P-0.5, and P-1.5 for the Time-varying operating model since there no simple basis for comparison.

The second evaluation method involves the estimation method's ability to approximate five quantities of interest to management based on spawning stock biomass (*SSB*). *SSB* is determined yearly by:

$$SSB_y = \sum_s \sum_i N_{y,s,1,i,1}\, w_{s,i} \quad (4.21)$$

These five management quantities are the unfished spawning stock biomass (SSB_0), the ratio between the spawning stock biomass in 2016 and SSB_0 (SSB_{2016}/SSB_0), the ratio SSB_{2016}/SSB_{1968} (1968 is the year when fishing started), SSB_{2016} and the average *SSB* over the entire time series (SSB_{avg}). These values are relevant to fisheries management since they determine the status of the stock and whether it is overfished. We again use the MRE and MARE to quantify performance. Each operating model data set is analyzed twice for each *M* option (growth estimated and pre-specified). Tallying the number of times either estimating or pre-specifying growth led to lower MREs (in absolute terms) and MAREs can illuminate whether having terminal molt impacts performance. We also averaged the percent difference from the lowest MARE value per operating model scenario to quantify the difference between estimated or pre-specified growth.

4.4 Results

4.4.1 Model Misspecification: Parameter estimation

For the Single operating model, no estimation method performed worse than P-1.5 and P-0.5 in which *M* was fixed to the wrong value (Figure 4.2). The methods that estimate *M*, E-M, E-SS and E-B, produce unbiased[2] estimates (see Appendix Table C.1 for the MREs and MAREs). The exceptions are the E-SS estimates for immature *M* (Figure 4.2). MARE values for E-M are consistently lower than those for E-SS and E-B. As expected, pre-specifying growth generally leads to lower MAREs for *M*. There are four exceptions; the E-SS estimate of immature *M* (E-SS(I)), the E-B estimate of *M* for block 2 with no terminal molt, and the E-B estimates of *M* for blocks 7 and 10 with terminal molt.

When the operating model has *M* vary by sex and maturity stage, P-1.5 and P-0.5 again performed the worst (Figure 4.3). The E-M and E-SS estimates for male and female *M* are unbiased with one exception, female *M* from E-SS with no terminal molt and pre-specified growth (Appendix Table C.2). All estimates of immature *M* are biased except for E-SS with terminal molt. E-SS does not consistently outperform E-M or even P-1 with regards to male and female *M*, but E-SS always outperforms E-M and P-1 for immature *M* (Figure 4.3). However, all methods had difficulty estimating immature *M* (MAREs > 10% except for E-SS with terminal molt and pre-specified growth). Again, pre-specifying growth leads to lower MAREs than estimating growth, with two exceptions (E-SS male *M* and E-M female *M* with no terminal molt).

The estimates of *M* from the blocking design follow the true time-trajectory of *M* well when the operating model is Time-varying (Figure 4.4). The estimated *M*'s for E-B are unbiased except for the first block when there is no terminal molt (Appendix Table C.3). The larger MAREs occur

[2] MRE value is between -10% and 10%

at the ends of the time series. Again, pre-specifying growth leads to better estimates of *M* than having growth estimated although there are exemptions to this conclusion.

There are several consistent trends regarding the estimation of the growth parameters, albeit with some exceptions (Appendix Tables C.4, C.5, C.6). Specifically, P-1.5 and P-0.5 lead to the poorest estimates, but terminal molt improves the estimates of L_∞ and *k* for P-1.5 and P-0.5. As for the other estimation methods, the differences in MAREs among estimation methods and whether there is terminal molt are less than 5%, with several exceptions when estimating L_∞.

4.4.2Model Misspecification: Estimation of SSB Trajectories

The largest MAREs (poorest performance) for any management quantity for the Single operating model occur for P-0.5 (Table 4.6). Estimation methods P-1, E-M and E-SS all produce unbiased estimates regardless of whether there is terminal molt or how growth is modeled (Appendix Table C.7). P-0.5, P-1.5 and E-B provide biased estimates for some of the management quantities, with P-0.5 leading to largest number of biased estimates. The lowest MAREs vary among estimation methods with P-1 having the most (6 out of 10) (Appendix Table C.8). Estimating or pre-specifying growth (regardless of whether there is terminal molt) does not impact estimation performance except for P-0.5 and P-1.5 (Table 4.7) for which estimating growth outperforms pre-specifying growth. This impact is also apparent in the Average statistic (Table 4.7). Estimating growth always leads to lower MAREs than pre-specifying growth for P-0.5 and P-1.5, with a large difference between the two growth modeling methods (> 100%), except for P-1.5 when there is terminal molt. All other estimation methods had small differences between estimating and pre-specifying growth (< 10 %), with estimating growth not always outperforming pre-specifying growth. This is visually apparent for the relative error in SSB (Figure 4.5).

P-0.5 tends to have the largest MARE values when the operating model is Sex + Stage (Table 4.8). Estimation methods P-0.5, P-1.5, P-1 and E-M all produce some biased estimates (Appendix

Table C.9). However, P-1 and E-M never lead to biased estimates when growth is estimated. E-SS never produces a biased estimate. The lowest MAREs vary among estimation methods, with the majority occurring with pre-specified growth (6 out of 10) (Appendix Table C.10). However, the Count statistic favors estimating growth, especially when the operating model does not have terminal molt (Table 4.7). This is further supported by the Average statistic, with estimating growth leading to smaller MAREs than pre-specifying growth for all *M* estimation methods except E-SS with terminal molt. This is visually apparent when looking at the relative error in SSB (Appendix Figure C.1).

All estimation methods perform poorly when the operating model is Time-varying (Figure 4.6). Visually, E-B appears to perform the best. However, all estimation methods produce biased management quantities (Appendix Table C.11). The MAREs for the estimated management quantities do not mirror Figure 4.6. E-B produces the second most largest MAREs (Table 4.9) and only has the lowest MARE 3 out of 10 times (Appendix Table C.12). As for estimating versus pre-specifying growth, the Count and Average statistics favor estimating over pre-specifying growth. However, this is primarily because P-0.5 performs better with estimated versus pre-specified growth (Table 4.7).

4.4.3 Data quality

The general trends from the simulation study exploring model misspecification are robust to changes in the biomass index CV and the effective sample size for the composition data. Table 4.10 lists the MAREs for *M* when the operating model is Single and Sex + Stage. The ability to estimate *M* varies between data quality scenarios with little consistency. The base level only performed best three times but is not consistently the worst.

Appendix Tables C.13 and C.14 lists the MAREs for the management quantities when the operating model is Single and Sex + Stage. Again, no data quality scenarios consistently

outperformed the others. However, the Average statistics suggest that having a smaller biomass index CV leads to a better performance (smaller MAREs) regardless of the operating model (Table 4.11).

4.4.4 Impact from choosing performance metrics

For the Time-varying operating model, there is a contradiction in the interpretation of the estimated management quantities and SSB relative error. Figure 4.6 visually shows that E-B is the best estimation method for the Time-varying operating model and that all other estimation methods have a U-shaped relative error pattern except for one of the P-0.5 estimation methods. It also shows that E-B has difficulty estimating SSB at the start of the time-series. Out of the five management quantities examined, two are ratios between the ends of the time-series, one is SSB_0 and the other is the average SSB over the entire time-series. These metrics could favor a U-shaped plot over a flat plot with larger errors at the start of the time-series. The final management quantity is the SSB at the end of the time-series where E-B consistently performs the best regardless of whether there is terminal molt and how growth is modeled (Table 4.9).

4.5 Discussion

4.5.1 Estimation of M

The simulation experiments focused on five key questions regarding the ability to estimate *M* in size-structured stock assessment models.

- Can size-structured stock assessment models estimate *M* or should it be pre-specified? Overall, the size-structured assessment models considered here can estimate *M*. Results suggest it is better to estimate *M* than to pre-specify it since pre-specifying *M* to the wrong value can lead to inaccurate and imprecise estimates for growth parameters and spawning stock biomass. This mirrors results found in previous studies for age-structured assessment models (e.g., Johnson et al., 2015; Lee et al., 2011; Punt et al., 2021).

- Should effort be made to estimate growth outside of the assessment? Estimation of *M* improved when growth was known compared to when growth was estimated simultaneously with *M*. This occurs irrespective of how *M* is modeled in the operating model. This result is not unsurprising given the confounding between estimated growth and *M* (e.g., Gislason et al., 2010; Chu et al., 2008).
- Do stocks with terminal molt lead to more accurate and precise estimates of *M* given the lesser confounding between estimated growth and *M*? Unexpectedly, there was little evidence that terminal molt improves or impairs the ability to estimate *M* (Appendix Tables C.1, C.2 and C.3). This could be due to the shape of the growth curve, with the size of the growth increment becoming increasing smaller as individuals approach the maximum asymptotic length. However, the maturity curve has all individuals maturing (reaching terminal molt) before reaching the maximum asymptotic length. It could be that terminal molt does not reduce the confounding between estimated growth and *M* because the growth parameters are still estimated when there is terminal molt.
- Can estimation of *M* be improved with more precise data? Somewhat surprisingly increasing the amount of size-composition data or reducing the CV of the biomass index did not impact the ability to estimate *M* substantially. Increasing the amount of the information in the biomass index did improve the estimation of spawning stock biomass, although this improvement appears to be independent of the ability to estimate *M* (Tables 4.10 and 4.11). It could be that other factors impact estimation of *M* more (e.g., different levels of contrast in the composition data) or that the increases in information considered here were not large enough to lead to obviously detectable improvements in the estimates for *M*.

- Can size-structured assessment models estimate stage-specificity and time-variation in *M*? Size-structured models can estimate stage-specific and time-varying *M* if the estimation method and operating model match. However, the degree of accuracy is variable (Figures 4.3 and 4.4). A limiting factor could be the amount of information available considering the highest variability occurs for immature *M* for E-SS and at the tail ends of the time-series for E-B.

An unexpected result was E-SS's difficulty estimating immature *M*. This happens for both the Single and Sex + Stage operating models. A possible explanation is the amount of information available for immature individuals. There are three fleets in each operating model: the directed fishery, the snow crab (bycatch) and the survey. Each fleet has its own selectivity curve. When comparing the population's maturity curve to the fleets' selectivity curves, it is apparent that information on immature individuals comes predominately from the survey with none coming from the directed fishery (Figure 4.7). The relative lack of information on immature animals could explain the difficulty E-SS has in estimating immature *M*. This hypothesis is further supported by E-SS leading to better estimates for mature male and mature female *M* than immature *M* when the operating model is either Single or Sex + Stage.

4.5.2 Estimation of spawning stock biomass

The factors that led to improved estimation of spawning stock biomass and the associated management quantities differed somewhat from those that led to improved estimation of *M*. *M* estimation improved if growth was known, but spawning stock biomass estimates are better when growth is estimated. This happens particularly when the estimation method is mis-specified relative to the operating model. For example, there is no evidence of improvement in SSB estimation when growth is estimated versus known for E-SS when the operating model is Single even though the Single operating model has a single *M* and E-SS estimates three separate *M* values

because there is nothing preventing those three values from being the same. Even though the E-SS assumption for *M* is different than that underlying the Single operating model, it can still mirror the Single operating model. Alternatively, P-0.5 and P-1.5 both assume there is a single *M* value, just like the Single operating model, but have markedly improved SSB estimation when growth is estimated versus known. *M* is pre-specified to the wrong value for P-0.5 and P-1.5 and can never match the Single operating model. When the operating model and estimation method match, it is hard to distinguish a difference in performance between estimating and knowing growth.

Estimating growth allows for more flexibility in the estimation method. If *M* in the estimation method is mis-specified, estimating growth can compensate to some extent given the confounding between *M* and growth, albeit at the cost of inaccurate estimates for the growth parameters. This is most pronounced when *M* is pre-specified to the wrong value; when *M* is pre-specified, growth is adjusted to a larger extent to compensate for model misspecification. Estimating growth does not, however, always lead to better estimates of spawning stock biomass (Appendix Table C.15).

An exception occurs with the Time-varying operating model. As expected, P-0.5 and P-1.5 lead to better estimates of spawning stock biomass when growth is estimated, but this is not the case for P-1 and E-M. This is surprising since P-1 and E-M cannot mirror time-varying natural mortality. A possible reason for this is that the average *M* value over the entire time series for the Time-varying operating model is *M*-base. *M* is pre-specified to *M*-base in P-1 while the E-M estimate of *M* is very close to *M*-base. Hence *M* is as close to the 'best value' as it can be, considering the limitations of their assumptions. There might not be much estimating growth can do to improve estimation of spawning stock biomass.

4.5.3 Estimation of growth parameters

Estimation methods that do not pre-specify *M* produce good estimates for the growth parameters. This is most likely due to *M* being estimated since it allows for more flexibility in the model. P-

0.5 and P-1.5 produce more biased estimates of growth since the model has less flexibility with a fixed *M*. However, the growth parameter estimates from P-1 are quite well estimated regardless of the operating model. For the Single and Time-varying operating models, *M* is specified to the 'best value' which may be the reason that growth parameter estimation is not impacted. As for the Sex + Stage operating model, P-1 has *M* set to the average between *male-M* and *female-M* and not to base-*M*. *Male-M* and *female-M* are very close to each other so the average between them is a pretty good estimate. There is an immature *M* in the Sex + Stage operating model but there is not much information on immature *M* so the consequences of this misspecification seem largely inconsequential.

4.5.4 Future work and conclusions

The results from this study show that *M* can be estimated within a size-structured assessment and that model output is better when *M* is estimated instead of being pre-specifed, especially when there is model misspecification. However, research goals dictate whether growth should be estimated or pre-specified. *M* estimates are better when growth is pre-specified. However, this study pre-specified growth to the correct value. It would be beneficial to explore how pre-specifying growth to incorrect values impact *M* estimation. Additional avenues to explore are more contrast in the time-series of fishing mortalities given how closely related fishing mortality is to *M*, and changes to data quality over time. Fisheries rarely have long consistent data series. There could be gaps in data as well as important data sources starting later in the fishery timeline. Exploring these impacts on *M* estimation would be beneficial.

Estimation of spawning stock biomass improved if growth was estimated, but mainly when the operating model was mis-specified. All estimation methods had difficulty estimating spawning stock biomass when *M* was time-varying (even though time-varying *M* could be estimated well by E-B). Operating models with more complex versions of time-varying *M*, such as abrupt block

changes (Johnson, 2015), annual variation around the linear increase/decrease in *M* (Deroba, 2013) and a random walk (Fu and Quinn, 2000) should be explored for size-structured models. This study only explored a linear increase in *M*. Additional methods for estimating time-varying *M* such as annual deviations (Deroba, 2013; Murphy, 2018) or a random walk (Fu and Quinn 2000; Shibata et al., 2021) should be considered as well. Even though E-B did the best at estimating time-varying *M*, other estimation methods that allow for time-varying *M* might do better.

This study does not conduct a complete analysis of which data sources impact the ability to estimate *M* in size-structured models. More levels of information should be explored as well as other data sources. A better understanding of which data source impacts *M* will be beneficial in terms of specifying monitoring schemes. Finally, the impact of terminal molt in size-structured models should continue to be explored. Even though there was not much influence of terminal molt in this study, it is a unique characteristic that could affect multiple aspects of a size-structured model.

The overall recommendation on how to estimate natural mortality in size-structured models depends on the research goals. If the goal is to obtain estimates for *M*, growth should be determined externally. If the goal is to obtain estimates of spawning stock biomass, growth should be estimated within the model. If there is evidence of *M* varying over time or by sex and maturity stage, the model should include those sources of variability which mirror result from previous studies (Fu and Quinn, 2011; Kanaiwa et al., 2008). These recommendations help reduce the risk of the assessment producing biased results.

Table 4.1. Values for natural mortality in the operating model. 'Single' has *M* set to a base value. 'Sex + Stage' has sex- and stage-specific time-invariant *M* and 'Time-varying' has *M* varying through time with a linear increase between *M*-1968 to *M*-2016. Source: 2020 stock assessment of EBS Tanner crab (Stockhausen, 2020).

Operating model	Natural mortality parameters	Value
Single	*M-Base*	0.23 yr^{-1}
Sex + Stage	*M-Male*	0.29 yr^{-1}
	M-Female	0.32 yr^{-1}
	M-Immature	0.24 yr^{-1}
Time-varying	*M-1968*	0.1725 yr^{-1}
	M-2016	0.2875 yr^{-1}

Table 4.2. Values for the biological parameters of the operating models and whether they are estimated in the estimation methods. Assumed known/Estimated applies to the growth parameters that are set to the true value when growth is pre-specified and estimated when growth is estimated.

Parameter	Operating model value	Estimation method treatment
Male parameters		
Growth		
L_∞	230 mm	Assumed known/Estimated
k	0.06948	Assumed known/Estimated
$\tilde{\sigma}$	0.25	Assumed known/Estimated
Length-weight relationship, γ_1	1.6×10^{-4} gm/mm^3	Assumed known
Length-weight relationship, ϵ_1	3.1	Assumed known
Maturity slope, ω_1	0.2	Assumed known
Maturity inflection Point, $M50_1$	104.4mm	Assumed known
Female parameters		
Growth		
L_∞	150 mm	Assumed known/Estimated
k	0.1273	Assumed known/Estimated
$\tilde{\sigma}$	0.25	Assumed known/Estimated
Length-weight relationship, γ_2	6.4×10^{-4} gm/mm^3	Assumed known
Length-weight relationship, ϵ_2	2.8	Assumed known
Maturity slope, ω_2	0.2	Assumed known
Maturity inflection Point, $M50_2$	74.6mm	Assumed known
Distribution of length-at-age 0, α	7.24	Assumed known
Distribution of length-at-age 0, β	5.04	Assumed known
$\kappa_{dir,s}$	1.13	Estimated
$R50_{dir,s}$	122.5mm	Estimated
R_0	e^9	Estimated
σ_R	0.57	Assumed known
λ_f	0.32	Assumed known
Selectivity		
Directed Fishery		
Slope, $\theta_{1,1}$	0.1	Estimated
Inflection point, $L50_{1,1}$	120mm	Estimated
Slope, $\theta_{1,2}$	0.1	Estimated
Inflection point, $L50_{1,2}$	120mm	Estimated
Snow Crab Fishery		
Slope, $\theta_{2,1}$	0.1	Estimated
Inflection point, $L50_{2,1}$	100mm	Estimated
Slope, $\theta_{2,2}$	0.1	Estimated
Inflection point, $L50_{2,2}$	100mm	Estimated
Survey		
Slope, $\theta_{3,1}$	0.15	Estimated
Inflection point, $L50_{3,2}$	60mm	Estimated
Slope, $\theta_{3,2}$	0.15	Estimated
Inflection point, $L50_{3,2}$	60mm	Estimated

Table 4.3. Specifications for the data generated from the operating model.

Data type	Years	CV or effective sample Size
Catch (directed fishery, only males retained)	1968 – 2015	Known
Discard (directed fishery, males and females)	1968 – 2015	CV = 0.20
Size-composition (directed fishery)	1968 – 2015	Effective sample size = 100
Size-composition (discard, directed fishery)	1992 – 2015	Effective sample size = 25
Discard (snow crab fishery, males and females)	1977 – 2015	CV = 0.20
Size-composition (snow crab fishery, males and females)	1992 – 2015	Effective sample size = 25
Size-composition (survey, males and females)	1976 – 2016	Effective sample size = 100 (base) Effective sample size = 200 (better info)
Index	1975 – 2015	CV = 0.20 (base) CV = 0.10 (better info)
Tagging	2005	Sample Size = 1,000

Table 4.4. Outline for the simulation study.

Operating model					
Terminal molt					
Yes				No	
Natural mortality					
Pre-specified		Time-varying (linear increase)		Sex + Stage	
Data quality					
Size-composition effective sample size			*Biomass index coefficient of variance*		
100	200		0.2	0.1	
Estimation method					
Natural mortality					
Pre-specified (Base Value)	Pre-specified (0.5 x Base Value)	Pre-specified (1.5 x Base Value)	Single estimated value	Time-varying (blocking design)	Sex + Maturity stage
Growth					
Pre-specified (True Value)				Estimated	

Table 4.5. Abbreviations for the six options for how natural mortality (*M*) is treated in the estimation methods.

Estimation method for *M*	**Abbreviation**
Pre-specified (Base Value)	P-1
Pre-specified (0.5 x Base Value)	P-0.5
Pre-specified (1.5 x Base Value)	P-1.5
Single estimated value	E-M
Sex + Maturity stage	E-SS
Time-varying (blocking design)	E-B

Table 4.6. Median absolute relative errors (MAREs) (expressed as percentages) for the management quantities from the Single operating model for the simulation study exploring model misspecification.

Terminal molt	Growth estimated?	Estimation method	SSB_0	$SSB_{16/0}$	$SSB_{16/68}$	SSB_{16}	SSB_{avg}
No	Yes	P-1	5.31	7.81	5.29	4.54	2.15
		P-0.5	49.71	32.51	41.01	7.34	9.68
		P-1.5	9.12	6.52	22.56	5.80	3.82
		E-M	4.88	7.49	5.79	4.76	1.93
		E-SS	9.72	10.69	7.19	4.72	2.33
		E-B	11.34	19.51	11.98	12.41	2.22
	Pre-specified	P-1	4.67	7.78	5.45	4.91	1.5
		P-0.5	64.07	25.64	86.97	26.09	15.46
		P-1.5	6.95	4.95	30.3	5.03	12.84
		E-M	5.41	8.06	5.08	5.16	1.92
		E-SS	8.10	9.38	6.01	5.53	2.72
		E-B	11.92	19.47	11.94	12.85	2.07
Yes	Yes	P-1	5.46	9.55	7.84	4.79	1.98
		P-0.5	46.92	24.97	16.21	11.5	2.29
		P-1.5	12.21	12.63	9.31	5.08	1.96
		E-M	5.88	8.87	7.99	5.15	1.91
		E-SS	7.31	7.61	7.65	5.36	1.93
		E-B	11.22	18.85	11.44	10.42	1.92
	Pre-specified	P-1	5.95	8.27	7.37	4.40	2.24
		P-0.5	52.37	17.02	20.1	32.54	8.15
		P-1.5	11.88	13.49	9.60	5.26	2.27
		E-M	6.13	8.33	7.74	4.61	2.20
		E-SS	5.97	8.10	7.48	4.47	2.12
		E-B	10.61	18.92	10.39	9.12	2.64

Table 4.7. Summary of the median absolute relative error (MAREs) (expressed as percentages) for the management quantities for the simulation study exploring model misspecification. Count is the number of times an estimated management quantity was better for estimated versus pre-specified growth for a given estimation method. The maximum value of Count is five since there are five management quantities. Average is the average (over management quantities) of the percent difference from the lowest MARE.

Estimation method	Count				Average			
	No terminal molt		Terminal molt		No terminal molt		Terminal molt	
	Growth estimated	Growth pre-specified	Growth estimated	Growth pre-specified	Growth estimated	Growth pre-specified	Growth estimated	Growth pre-specified
Operating model = Single								
P-1	2	3	2	3	24	15	9	7
P-0.5	4	1	4	1	567	941	258	424
P-1.5	2	3	4	1	131	262	47	53
E-M	3	2	2	3	21	24	10	9
E-SS	2	3	2	3	65	57	12	6
E-B	2	3	2	3	159	161	89	86
Total/Mean	15	15	16	14	161	243	71	97
Operating model = Sex + Stage								
P-1	4	1	3	2	16	145	27	135
P-0.5	4	1	4	1	421	603	291	498
P-1.5	4	1	3	2	104	450	55	189
Est-M	5	0	3	2	15	125	24	129
E-SS	3	2	0	5	36	38	32	14
Total/Mean	20	5	13	12	118	272	86	193
Operating model = Time-varying								
P-1	2	3	4	1	68	65	41	42
P-0.5	5	0	5	0	238	1277	198	354
P-1.5	3	2	2	3	122	241	35	32
E-M	3	2	3	2	65	65	38	39
E-B	3	2	3	2	173	172	192	190
Total/Mean	16	9	17	8	133	364	100	131

Table 4.8. Median absolute relative errors (MAREs) (expressed as percentages) for the management quantities from the Sex + Stage operating model for the simulation study exploring model misspecification.

Terminal molt	Growth estimated?	Estimation method	SSB_0	$SSB_{16/0}$	$SSB_{16/68}$	SSB_{16}	SSB_{avg}
No	Yes	P-1	8.62	8.08	7.63	6.15	1.59
		P-0.5	58.28	34.57	36.36	8.19	8.64
		P-1.5	6.51	7.15	22.43	4.96	6.40
		E-M	8.90	8.12	7.10	6.19	1.60
		E-SS	10.68	10.69	8.23	5.53	2.22
	Pre-specified	P-1	15.49	8.80	7.28	7.30	9.81
		P-0.5	73.52	29.17	63.94	24.73	9.14
		P-1.5	7.36	7.05	28.17	13.51	29.77
		E-M	18.46	10.20	7.31	7.80	6.96
		E-SS	9.72	10.38	8.69	5.79	2.47
Yes	Yes	P-1	8.62	8.16	8.68	6.81	2.27
		P-0.5	56.27	30.65	12.81	10.13	2.47
		P-1.5	8.22	15.07	8.59	6.56	3.09
		E-M	8.29	7.70	8.78	6.70	2.28
		E-SS	8.71	10.55	8.76	5.73	2.36
	Pre-specified	P-1	12.03	6.46	7.79	12.12	11.81
		P-0.5	64.49	19.85	18.1	35.55	14.30
		P-1.5	5.91	16.86	7.20	14.61	15.34
		E-M	12.29	5.90	7.55	12.17	11.29
		E-SS	6.93	8.65	7.52	5.27	2.25

Table 4.9. Median absolute relative errors (MAREs) (expressed as percentages) for the management quantities from the Time-varying operating model for the simulation study exploring model misspecification.

Terminal molt	Growth estimated?	Estimation method	SSB_0	$SSB_{16/0}$	$SSB_{16/68}$	SSB_{16}	SSB_{avg}
No	Yes	P-1	9.92	23.55	6.30	35.96	2.72
		P-0.5	26.28	17.00	33.73	41.10	5.38
		P-1.5	5.59	23.97	22.82	26.22	6.42
		E-M	9.16	21.89	7.02	34.79	2.83
		E-B	39.76	32.23	16.26	11.83	2.40
	Pre-specified	P-1	9.96	22.37	6.99	34.96	2.43
		P-0.5	59.92	49.51	239.89	134.46	13.08
		P-1.5	6.26	17.33	41.79	18.11	15.27
		E-M	9.78	22.22	6.64	34.30	2.84
		E-B	40.55	32.52	15.24	12.27	2.27
Yes	Yes	P-1	7.31	18.11	12.71	25.28	2.23
		P-0.5	37.00	18.28	23.74	37.67	2.41
		P-1.5	5.38	27.01	10.27	22.76	2.24
		E-M	7.09	16.60	13.09	25.35	2.14
		E-B	45.19	31.40	18.13	11.54	2.93
	Pre-specified	P-1	6.94	18.59	12.96	25.45	2.34
		P-0.5	43.37	19.45	40.68	70.67	6.49
		P-1.5	5.43	25.86	9.46	21.40	2.50
		E-M	7.05	16.96	13.03	25.45	2.25
		E-B	43.77	30.54	19.21	11.80	3.15

Table 4.10. Median absolute relative errors (MAREs) (expressed as percentages) for *M* from the Single and Sex + Stage operating models for the simulation study exploring data quality. The Single operating model only has one *M* value, therefor the MAREs for each sex and stage is the same for P-1 and E-M. The bold values are when the base data level performs best.

Run				Single			Sex + Stage		
Terminal bolt	**Growth estimated?**	**Estimation method**	**Data**	**Male *M***	**Female *M***	**Immature *M***	**Male *M***	**Female *M***	**Immature *M***
Both	Both	P-1	Base		0		5.17	4.69	27.08
			H-SS		0		5.17	4.69	27.08
			S-CV		0		5.17	4.69	27.08
No	Yes	E-M	Base		**4.40**		5.30	4.67	**27.15**
			H-SS		4.43		5.29	4.62	27.17
			S-CV		5.18		5.93	4.10	28.00
		E-SS	Base	5.92	6.18	24.21	5.96	5.83	25.32
			H-SS	5.51	6.17	22.77	4.93	6.06	19.37
			S-CV	3.91	8.03	22.69	4.13	6.71	24.73
	Pre-specified	E-M	Base		2.19		2.65	11.49	18.01
			H-SS		2.01		2.13	11.08	18.56
			S-CV		1.81		3.33	12.39	16.81
		E-SS	Base	5.56	2.89	11.41	6.17	2.38	11.29
			H-SS	5.89	2.02	9.07	6.15	1.83	8.51
			S-CV	3.41	4.04	13.84	3.68	2.62	12.44
Yes	Yes	E-M	Base		4.42		5.93	6.43	**27.96**
			H-SS		3.84		7.90	5.50	30.22
			S-CV		4.19		8.28	5.30	30.43
		E-SS	Base	5.77	5.72	10.29	5.97	5.70	14.82
			H-SS	5.68	4.58	9.57	5.52	4.97	12.95
			S-CV	4.84	5.54	10.05	5.33	4.99	12.36
	Pre-specified	E-M	Base		3.40		5.68	4.37	27.68
			H-SS		2.39		7.51	3.29	29.91
			S-CV		3.24		9.18	2.78	31.93
		E-SS	Base	4.81	4.71	11.92	4.68	4.52	9.77
			H-SS	4.46	3.03	11.12	3.57	2.45	8.25
			S-CV	3.72	3.48	6.91	4.56	3.60	7.78

Table 4.11. Summary of the median absolute relative error (MAREs) (expressed as percentages) for the management quantities for the simulation study exploring data quality. Count is the number of times an estimated management quantity was better for estimated versus pre-specified growth for a given estimation method. The maximum value of Count is five since there are five management quantities. Average is the average (over management quantities) of the percent difference from the lowest MARE.

Method estimation	Data	Count				Average			
		No terminal molt		Terminal molt		No terminal molt		Terminal molt	
		Growth estimated	Growth pre-specified	Growth estimated	Growth pre-specified	Growth estimated	Growth pre-specified	Growth estimated	Growth pre-specified
Operating model = Single									
P-1	Base	2	3	2	3	39	30	38	36
	H-SS	3	2	3	2	36	29	36	36
	S-CV	2	3	2	3	15	6	13	14
E-M	Base	3	2	2	3	37	40	39	392
	H-SS	2	3	2	3	33	35	32	31
	S-CV	2	3	2	3	15	10	11	15
E-SS	Base	2	3	2	3	80	76	42	34
	H-SS	2	3	2	3	78	72	39	37
	S-CV	2	3	2	3	39	21	22	14
Total/Mean		20	25	19	26	41	35	30	28
Operating model = Sex + Stage									
P-1	Base	4	1	3	2	38	192	52	196
	H-SS	5	0	4	1	35	171	34	179
	S-CV	5	0	3	2	7	127	31	131
E-M	Base	5	0	3	2	37	162	49	188
	H-SS	4	1	4	1	40	133	34	171
	S-CV	4	1	3	2	12	115	29	135
E-SS	Base	3	2	0	5	59	63	57	36
	H-SS	2	3	1	4	55	53	31	26
	S-CV	2	3	0	5	12	10	24	14
Total/Average		34	11	21	24	33	114	38	119

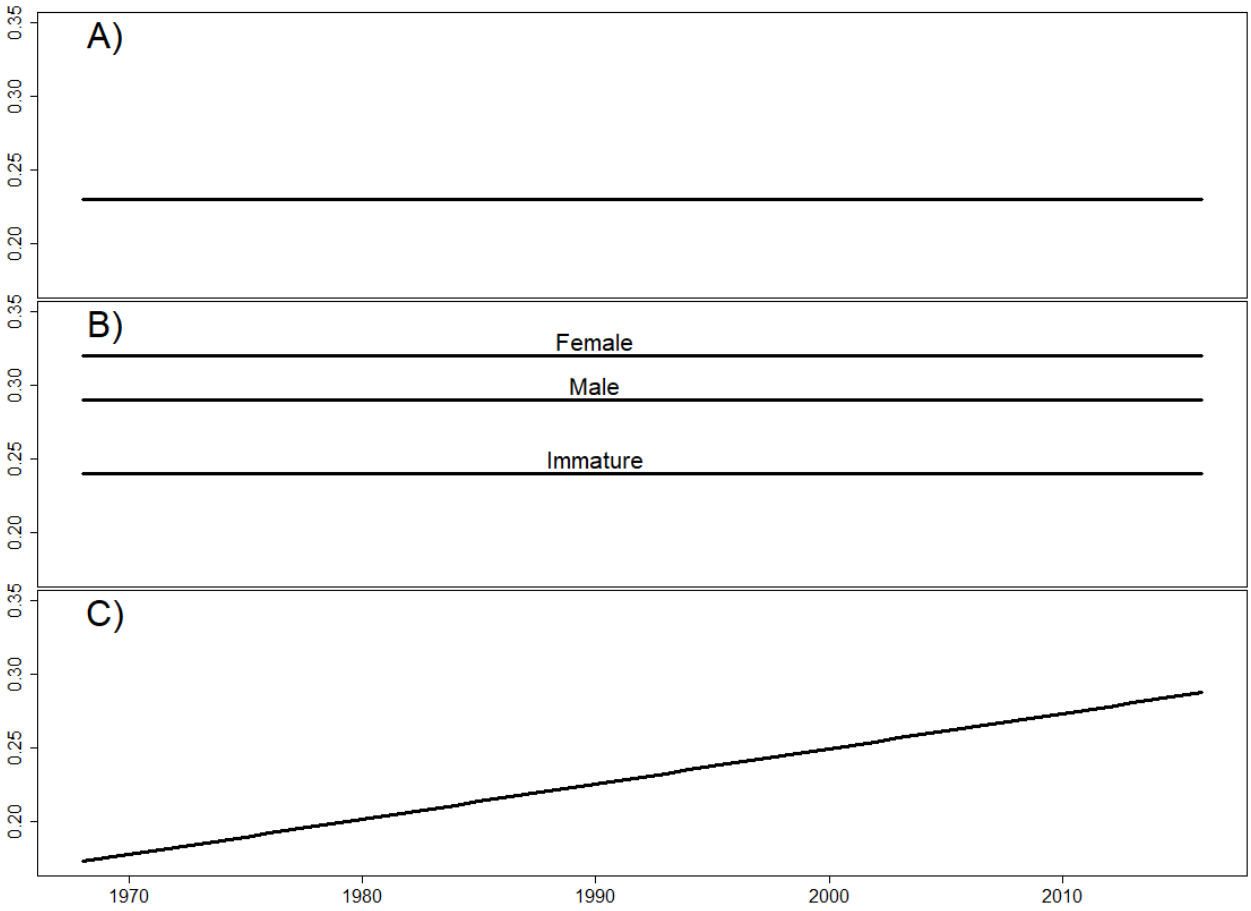

Figure 4.1. *M* value timeline by operating model: Single (A), Sex + Stage (B), and Time-varying (C).

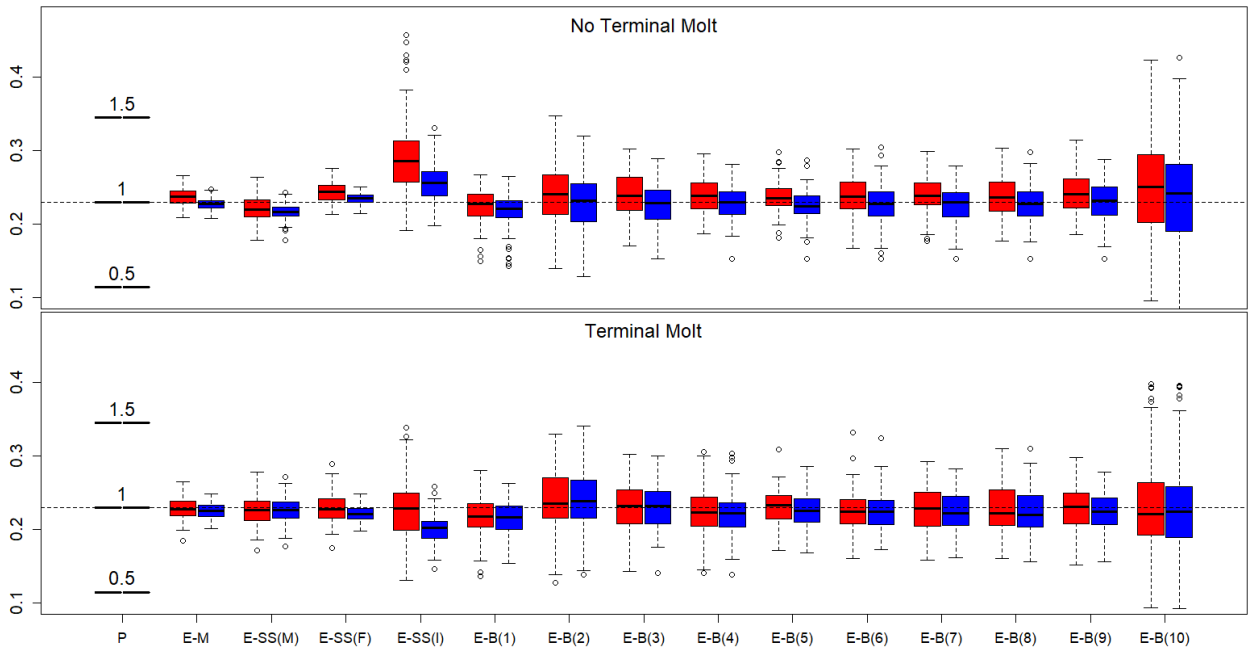

Figure 4.2. Boxplots of estimated/pre-specified *M* for each estimation method for the Single operating model for the simulation study exploring model misspecification. The x-axis is the estimation method. P has three fixed options based on what *M*-base is multiplied by (0.5, 1 or 1.5). E-SS estimates *M* for three stages: mature males (M), mature females (F) and immature males/females (I). E-B estimates 10 values, one for each time-block with 1 representing the start of the time series. The color indicates whether the growth parameters are estimated (red – estimated, blue – pre-specified). The dashed line is the true value in the operating model.

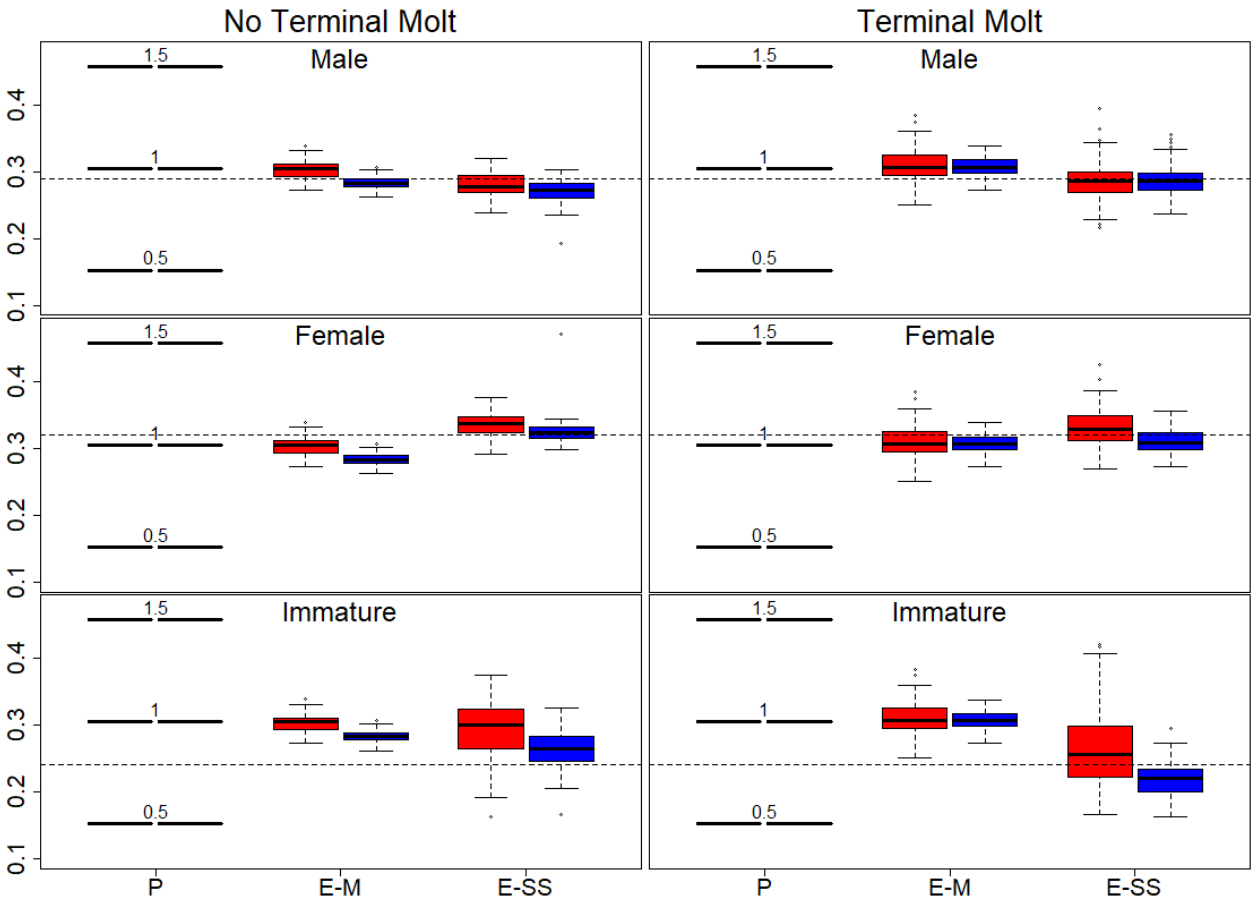

Figure 4.3. Boxplots of estimated/pre-specified *M* for each estimation method for the Sex + Stage operating model for the simulation study exploring model misspecification. The titles indicate sex or stage (male, female, and immature). The x-axis is the estimation method. P has three options based on what the *M*-base is multiplied by (0.5, 1 or 1.5). The color indicates whether the growth parameters are estimated (red – estimated, blue – pre-specified). The dashed line is the true value in the operating model.

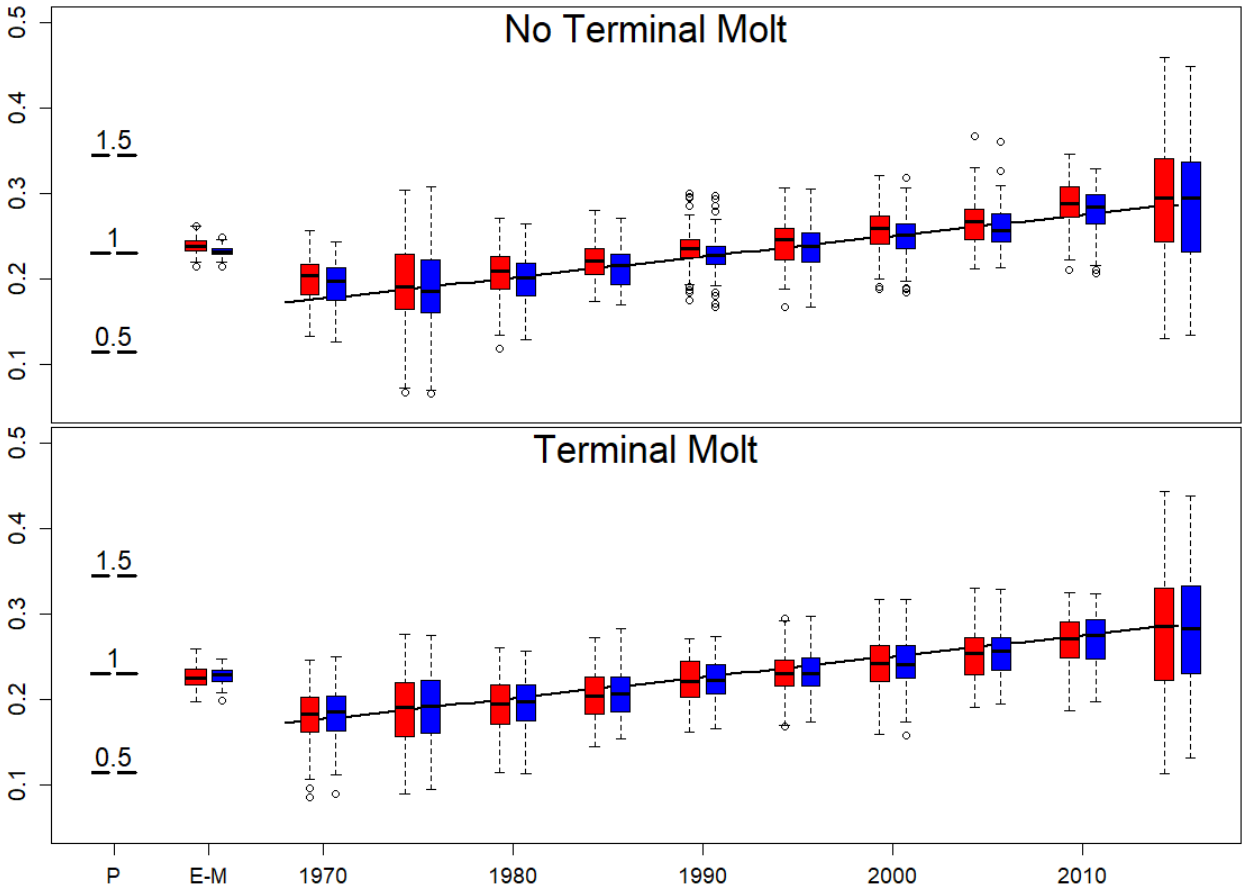

Figure 4.4. Boxplots of estimated/pre-specified *M* for each estimation method for the Time-varying operating model for the simulation study exploring model misspecification. The x-axis is the estimation method. P has three options based on what the *M*-base is multiplied by (0.5, 1 or 1.5). The timeline on the x-axis indicates blocking design. There are ten time-blocks, with each boxplot plotted at the midpoint of their associated time-block. The color indicates whether the growth parameters are estimated (red – estimated, blue – pre-specified). The dashed line is the true value in the operating model.

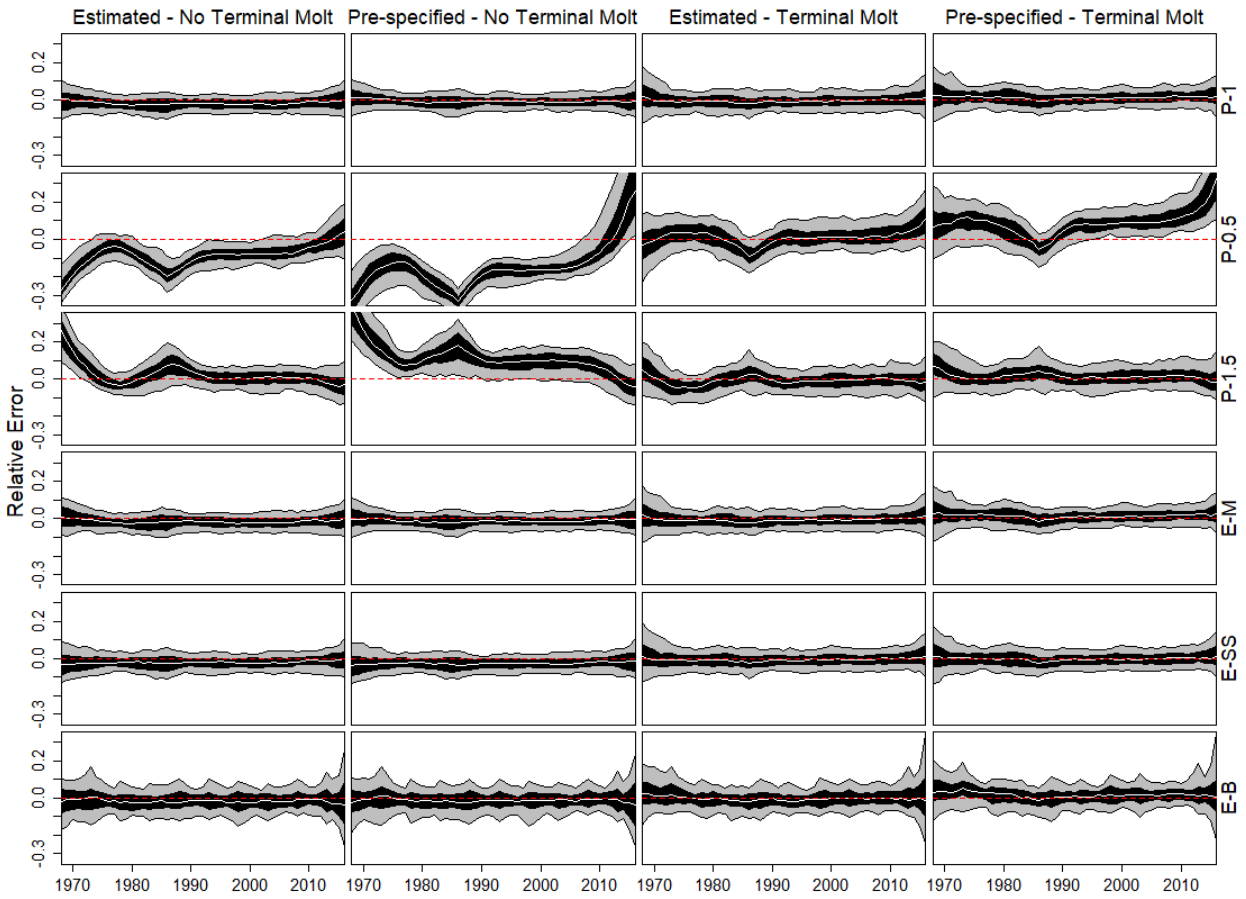

Figure 4.5. Relative error distributions for spawning stock biomass (SSB) for the Single operating model for the simulation study exploring model misspecification. The column titles indicate whether growth is estimated or pre-specified and if there is terminal molt. The row titles are the estimation method for *M*. The white line is the median relative error, the red dash line is the zero line, the black shaded region is the 50% quantile, and the grey shaded region is the 90% quantile.

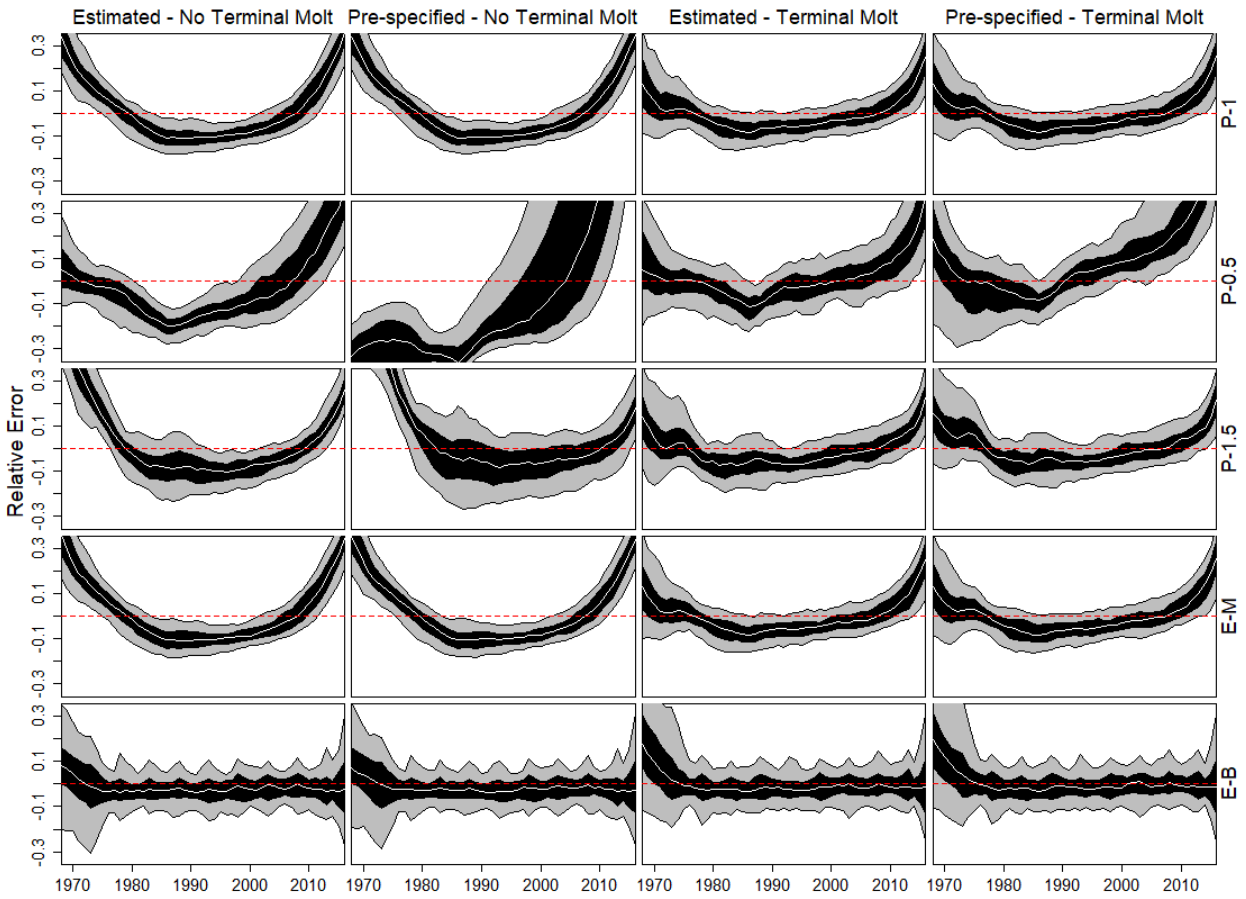

Figure 4.6. As for Figure 4.5 except for the Time-varying operating model.

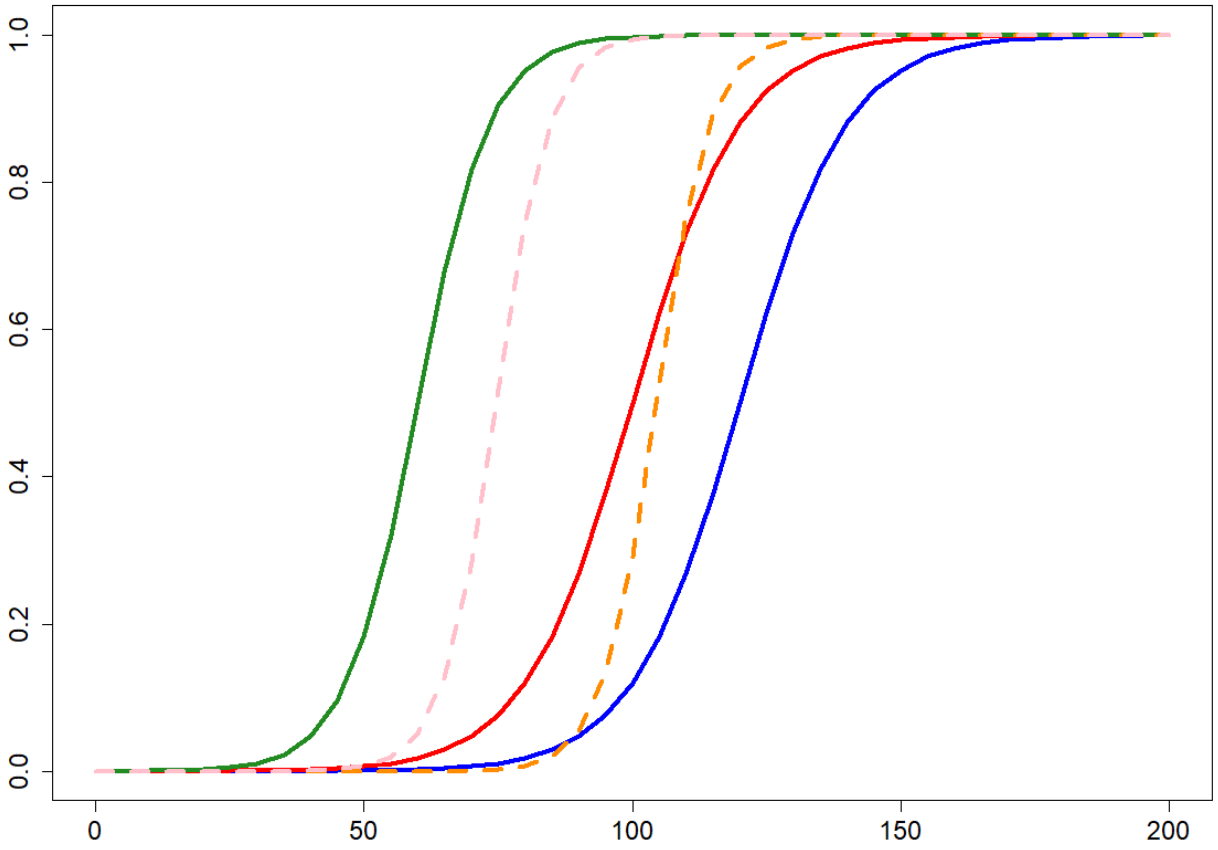

Figure 4.7. The maturity and selectivity curves. The solid lines are the selectivity curves (blue – directed, red – Snow crab and green – survey) and the dashed lines are the maturity curves (orange – male, pink – female)

Discussion

Size-structured population dynamics models are an important tool for assessing stocks of hard to age fished species. These models, like all models, involve assumptions that limit the ability of the model to replicate the real world. The focus of this dissertation was to investigate and improve upon the assumptions within size-structured models to allow stock assessments to estimate quantities important to management with greater accuracy and precision. Three aspects of size-structured models were explored: growth, selectivity and natural mortality. The results from this work can provide guidance and suggestions for how to model these characteristics and what issues a modeler needs to be aware of when conducting stock assessments based on size-structured population dynamics model.

The first two chapters focused on modeling variation in growth within size-structured models. Previous methods incorporated growth variation by either process error around a single curve, allowing the population to have multiple growth curves through the platoon method or allowing one of the growth parameters to vary among individuals (Methot and Wetzel, 2013; Punt et al., 2009). Realistically, each individual follows its own growth curve, characterized by individual values for the growth parameters. The first chapter developed a new method that allows both of the parameters of the von Bertalaffy growth curve (L_∞ and k) to vary among individuals through numerical integration. This can be accomplished in multiple ways. However, the goal was to identify a method that is most accurate and robust. Several factors were considered when developing this new method. These included which numerical integration technique to use, how many evaluation points to use when conducting the necessary integrals, which equation (Equation-L_∞ or Equation-k (Equations 1.7a and 1.7b from Chapter 1)) to use to reduce the dimensionality of the four dimensional integral and how to conduct a numerical integration over the range $[0,\infty]$.

These factors resulted in six versions of the numerical integration technique, with the 32-point Gaussian quadrature using either Equation-L_∞ or -k performing best.

Chapter 2, as with all the remaining chapters, involved a simulation experiment with an operating model and estimation methods. The benefits of testing assessment methods using simulation are that we know the true dynamics (and hence size) of the population, and we can specify the quality and quantity of the data used by the estimation methods. This allows one to determine how accurate and precise an estimation method is. However, the operating model, like all models, makes assumptions about the real world and thus deviates from it. We try to compensate for this by having multiple versions of the operating model in each simulation study. It is important to acknowledge this limitation of simulation studies since the real world might be more complex than what any one operating model can simulate.

The second chapter of this dissertation compared the newly developed numerical integration technique with two earlier methods for accounting for individual variation in growth. The first earlier method assumed that individuals follow a single growth curve with process error and the other earlier method was the platoon method, which allows individuals to follow one of three growth curves, each with process error. The numerical integration technique can be simplified to allow only one of the two growth parameters to vary among individuals while the platoon method varies based on the degree of overlap between the three growth curves. Nine methods for constructing size-transition matrices using tag-recapture data were compared with the intention of determining which method was the most robust. The results showed that the performance of a method for estimating the size-transition matrix is determined by the width of the size-classes, the time-at-liberty of the tagged individuals, and the distribution of the size-at-releases for the tagged animals. Large size-class widths mask the source of growth variability leading to an overall

reduction in performance. Longer time-at-liberties result in greater violation to the assumed uniform size distribution within each size-class that underlie the numerical methods for integrating over individual variation in growth. As for size-at-release, all methods for constructing a size-transition matrix violate the assumed uniform size-at-release distribution since actual size-at-release data have a bell-shaped distribution. This results in no one method consistently performing the best. When constructing a size-transition matrix, the modeler must be aware of the quality of their tagging data and what width to use for the size-classes. However, the goal of this work was to provide recommendations on how to model growth variation. Therefore, if we assume individuals in the real world each follow their own growth curve, then my recommendation is to use the simplified numerical integration method that allows L_∞ to vary among individuals (Vary Linf) with smaller size-class widths and only one year of tagging data.

The third chapter of the dissertation explored two aspects of time-varying selectivity within size-structured models. The first was how to model time-varying selectivity. Three methods were considered: time-invariant, blocking design and random parameters. Time-invariant has a single selectivity curve for the entire fishing history. Blocking design divides the fishing history into equal-sized time-blocks where selectivity remains constant within a time-block but varies between time-blocks. Random parameters has a separate selectivity curve for each year. The results showed that it is better to have some version of time-varying selectivity regardless of what is occurring in the real world (i.e., whether selectivity is time-varying or not). This finding mirrors results from selectivity studies that were based on age-structured models (e.g., Martell and Stewart, 2014). A surprising result was that blocking design performed just as well as random parameters even when the operating model used to generate the pseudo data sets more closely mimicked random parameters. Blocking design is likely a better option than random parameters since it involves

fewer parameters, which reduces the variance of estimated quantities with little increase in bias. Stock assessments that use a blocking design for time-varying selectivity typically have selectivity change based on known significant changes in fishing practices (Stockhausen, 2015; Thompson, 2015). Chapter 3 arbitrarily divided the timeline into equal-sized time-blocks. Overall, I would recommend a blocking design with equal-sized time-blocks. However, further research should be conducted to determine the optimal length for the time-blocks.

The second part of Chapter 3 explored likelihoods for the size-composition data. There are many likelihood functions available for size-composition data (Maunder, 2011), but few studies have evaluated which is more robust when the true distribution of the size-composition data is unknown (Punt et al., 2014). Three likelihoods were examined: the multinomial, the Dirichlet-multinomial and the multivariate normal. The results showed that no one likelihood function consistently outperformed the others, with each having their pros and cons. The multinomial was never the least worst option. The Dirichlet-multinomial had difficulty estimating the effective sample size and the multivariate normal was the most prone to retrospective patterns. With no clear best likelihood function, it is important to understand the strengths and weakness of each likelihood function so that a modeler can make an informed decision on which likelihood function best fits their goals. Again, the intention of this work was to provide guidance on how to construct a stock assessment based on a size-structured population model. Therefore, I would recommend the multinomial likelihood function for size-composition data, although I acknowledge that arguments can be made for the both Dirichlet-multinomial and multivariate normal.

The final chapter of this dissertation explored natural mortality (M). M is an important parameter that directly impacts stock productivity (Clark, 1999; Kenchington, 2014; Punt et al., 2021). Overall, the results showed that stock assessments based on size-structured models can

estimate *M* and that it is better to estimate *M* than pre-specify it. This mirrors results from previous studies exploring *M* in stock assessments based on age-structured models (Johnson et al., 2015; Lee et al., 2011; Punt et al., 2021). It was expected that terminal molt would have a positive impact on the ability to estimate *M* since individuals stop growing after maturing, which should reduce the confounding relationship between growth and *M*. The results showed that terminal molt did not impact *M*. However, performance was influenced by whether growth was estimated. *M* estimates were better when growth was known while estimates of spawning stock biomass were more accurate and precise when growth was estimated. These conflicting trends are based on the relationship between growth and *M*. Having pre-specified growth allows for more information to be used to estimate *M*, but reduces the overall flexibility of the estimation method to estimate spawning stock biomass especially when the model underlying the estimation method does not match that underlying the operating model. Therefore, how best to construct a size-structured population dynamics model depends on the modeler's goals. If the focus is on natural mortality, it is better to estimate growth externally then pre-specify it within the size-structured model. However, it is important to note that in this study the size-transition matrix was pre-specified to values used in the operating model. If the focus is on spawning stock biomass, it is better to estimate growth internally with the understanding that the accuracy and precision of the parameter estimates for natural mortality and growth could be poorer.

Though extensive work was done exploring growth, selectivity and natural mortality in size-structured models there is still additional work that can be done. For growth, the number of years of tagging data had a large impact on the ability to estimate size-transition matrices. Analyses examining the trade-off between using fewer tag-recaptures (perhaps only those tags at liberty for a year or two) and reducing bias versus using more tags and increasing bias but reducing variance

should be conducted. Also, many stock assessment for crustaceans estimate the probability of animals molting. Tests should be conducted to examine how including molt probability influences the performance of the estimation methods.

For selectivity, we arbitrarily choose five-year block lengths for the blocking design. Alternative block length such as, seven or ten years, should be explored to see if they perform just as well as random parameters. This would further reduce the number of estimated parameters which would further reduce the variance of estimated quantities. As for the size-composition likelihood function, additional likelihood functions such as the Fournier et al. (1990) normal approximation of the multinomial distribution, should be examined to determine the pros and cons of other commonly used likelihood functions.

Whether growth was estimated or not had a large impact on the ability to estimate *M*. Chapter 4 pre-specified growth to the correct value. Studies should be conducted that pre-specify the size-transition matrix incorrectly to see how robust the benefit from pre-specifying growth is for the estimation of *M*. In addition, all estimation methods had difficulty estimating spawning stock biomass when *M* was time-varying. The only estimation method with time-varying *M* in Chapter 4 used the blocking design method. Other methods such as annual deviations and random walk should also be explored.

The goal of this dissertation was to evaluate and improve upon the assumptions within size-structure population dynamics models to more accurate represent reality. This dissertation focused on three aspects of a size-structure model: growth, selectivity and natural mortality. The results from each section reveal that there is no universally best method. There were always several additional factors that needed to be considered when choosing how best to construct the model. It is important to understand the influence and limitations associated with the choices made when

conducting stock assessments based on a size-structured model to make informed decisions and improve the overall performance of stock assessments.

Bibliography

Abramowitz, M. and Stegunm I. A., 1972. Handbook of mathematical functions with formulas, graphs, and mathematical tables. Washington, D.C.: U.S. Dept. of Commerce: U.S. G.P.O.

Andrew, K. I. and Mangel, M., 2012. Asymptotic size and natural mortality of long-lived fish for data poor stock assessments. Fisheries Research 127 – 128, 45 – 48.

Atlantic States Marine Fisheries Commission (ASMFC), 2015. American Lobster Benchmark Stock Assessment and Peer-review Report. Atlantic States marine Fisheries Commission, Washington, DC (493 pp.).

Berger, A. M., Edwards, A. M., Grandin, C. J. and Johnson, K. F., 2019. Status of the Pacific Hake (whiting) stock in the U.S. and Canadian waters in 2019. Prepared by the Joint Technical Committee of the U.S. and Canadian Pacific Hake/Whiting Agreement, National Marine Fisheries Service and Fisheries and Oceans Canada. 249 p.

Beverton, R. J. H. and Holt, S. J., 1957/1993. On the Dynamics of Exploited Fish Populations. Chapman and Hall, London.

Buckworth, R. C., Deng, R. A., Plagányi, E. E., Punt, A. E., Upston, J., Pascoe, S., Miller, M. T. Hutton, T., Lawrence, E., and Venables, W., 2015. Northern Prawn Fishery RAG Assessments 2013-15. Final Report to the Australian Fisheries Management Authority, Research Project 2013/0005, June 2015. CSIRO. Brisbane. 177 p.

Chu, C. T., Chien, H. K., and Lee, R. D., 2008. Explaining the optimality of U-shaped age-specific mortality. Theoretical Population Biology 73, 171 – 180.

Clark, W., 1999. Effects of an erroneous natural mortality rate on a simple age-structured stock assessment. Canadian Journal of Fisheries and Aquatic Sciences 56, 1721 – 1731.

Deroba, J. J. and Schueller A. M., 2013. Performance of stock assessments with misspecified age- and time-varying natural mortality. Fisheries Research 146, 27 – 50.

Federal Register. 2009. Magnuson-Stevens act provisions; annual catch limits; national standard guidelines; final rule., 74, 3178–3213.

Francis, R. I. C. C. 1988. Maximum likelihood estimation of growth and growth variability from tagging data. New Zealand Journal of Marine and Freshwater Research 22, 43 – 51.

Fournier, D. A. and Archibald, C. P. 1982. A general theory for analyzing catch at age data. Canadian Journal of Fisheries and Aquatic Science 39, 1195 – 1207.

Fournier, D. A., Sibert, J. R., Majkowski, J. and Hampton, J., 1990. MULTIFAN a likelihood-based method for estimating growth parameters and age composition from multiple length frequency data sets illustrated using data from southern Bluefin tuna (*Thunnus maccoyii*). Can. J. Fish. Aquat. Sci. 47, 301 – 317.

Fu, C., Quinn, T. J. II., 2000. Estimability of natural mortality and other population parameters in a length-based model: *Pandalus borealis* in Kachemak Bay, Alaska. Canadian Journal of Fisheries and Aquatic Sciences 57, 2420 – 2432.

Garber-Yonts, B. and Lee, J., 2014. Stock Assessment and Fishery Evaluation Report for King and Tanner Crab Fisheries of the Bering Sea and Aleutian Islands Regions: Economic Status of the BSAI Crab Fisheries, 2015.

Garber-Yonts, B. and Lee, J., 2018. Stock Assessment and Fishery Evaluation Report for King and Tanner Crab Fisheries of the Gulf of Alaska and Bering Sea/Aleutian Islands Area: Economic Status of the BSAI King and Tanner Crab Fisheries off Alaska, 2017.

Garber-Yonts, B. and Lee, J., 2019. Stock Assessment and Fishery Evaluation Report for King and Tanner Crab Fisheries of the Gulf of Alaska and Bering Sea/Aleutian Islands Area: Economic Status of the BSAI King and Tanner Crab Fisheries off Alaska, 2018.

Garber-Yonts, B. and Lee, J., 2020. Stock Assessment and Fishery Evaluation Report for King and Tanner Crab Fisheries of the Gulf of Alaska and Bering Sea/Aleutian Islands Area: Economic Status of the BSAI King and Tanner Crab Fisheries off Alaska, 2019.

Gislason, H., Daan, N., Rice, J. C. and Pope, J. G., 2010. Size, growth, temperature and the natural mortality of marine fish. Fish and Fisheries 11, 149 – 158.

Gudmundsson, G., 1994. Time series analysis of catch-at-age observations. Appl. Statistics 43, 117 – 126.

Gudmundsson, G. and Gunnlaugsson, T. 2012. Selection and estimation of sequential catch-at-age models. Canadian Journal of Fisheries and Aquatic Science 69, 1760 – 1772.

Hewitt, D. A., Lambert, D. M., Hoeing, J. M., Lipcius, R. N., 2007. Direct and Indirect estimates of natural mortality for Chesapeake bay blue crab. American Fisheries Society 136, 1030 – 1040.

Hulson, P-J. F. and Hanselman, D. H. 2014. Tradeoffs between bias, robustness, and common sense when choosing selectivity forms. Fisheries Research 158, 63 – 73.

Hurtado-Ferro, F., Szuwalski, D. H., Valero, J. L., Anderson, S. C., Cunningham, C. J., Johnson, K. F., Licandeo, R., McGilliard, C. R., Monnahan, C. C., Muradian, M. L., Ono, K., Vert-Pre, K. A., Whitten, A. R. and Punt, A. E., 2015. Looking in the rear-view mirror: bias and retrospective patterns in integrated, age-structured stock assessments models. ICES J. Mar. Sci. 72, 99 – 11.

Huxley, T. H., 1884. Inaugural Address. Fisheries Exhibition Literature 4, 1 – 22.

Ianelli, J. N., Honkalehto, T., Barbeaux, S. and Kotwicki, S., 2015. Assessment of the walleye pollock stock in the Eastern Bering Sea. Pg. 53-152. In: Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Bering Sea/Aleutian Islands Regions. North Pacific Fishery Management Council, Anchorage, AK.

Johnson, K. F., Monnahan, C. C., McGillaird, C., Vert-pre, K., Anderson, S. C., Cunningham, C., Hurtado-Ferro, F., Licandeo, R., Muradian, M., Ono, K., Szuwalski, C., Valero, J., Whitten A., Punt, A. E., 2015. Time-varying natural mortality in fisheries stock assessment models: identifying a default approach. ICES Journal of Marine Science 72, 137–150.

Jones, C., 1986. Determining age of larval fish with the otolith increment technique. Fish. Bull. 84, 91 – 103.

Kanaiwa, M., Chen, Y., and Wilson, C., 2005. Evaluating a seasonal, sex-specific size-structured stock assessment model for the American lobster, *Homarus americanus.* Marine and Freshwater Research 59, 41 – 56.

Kenchington, T. J., 2014. Natural mortality estimators for information-limited fisheries. Fish and Fisheries 15, 533 – 562.

Lee, H-H., Maunder, M. N., Piner, K. R., and Methot, R. D., 2011. Estimating natural mortality within a fisheries stock assessment model: An evaluation using simulation analysis based on twelve stock assessment. Fisheries Research 109, 89 – 94.

Linton, B. C. and Bence, J. R. 2011. Catch-at-age assessment in the face of time-varying selectivity. ICES Journal of Marine Science 68, 611 – 624.

Lorenzen, K., 1996. The relationship between body weight and natural mortality in juvenile and adult fish: a comparison of natural ecosystems and aquaculture. Journal of Fish Biology 49, 627 – 647.

Marsh, C. and Fu, D., 2017. The 2016 stock assessment of paua (*Haliotis iris*) for PAU 5D. New Zealand Fisheries Assessment Report 2017/33.

Martell, S. and Stewart, I., 2014. Towards defining good practices for modeling time-varying selectivity. Fisheries Research 158, 84 – 95.

Maunder, M. N., 2011. Review and evaluation of likelihood functions for composition data in stock-assessment models: Estimating the effective sample size. Fisheries Review 109, 311 – 319.

Maunder M. N. and Piner K. R., 2015. Contemporary fisheries stock assessment: many issues still remain. ICES Journal of Marine Science 72, 7 – 18.

McAllister, M.K., Ianelli, J.N., 1997. Bayesian stock assessment using catch-age data and the sampling/importance resampling algorithm. Can. J. Fish. Aquat. Sci. 54, 284 – 300.

Methot, R. D. and Wetzel, C. R. 2013. Stock synthesis: A biological and statistical framework for fish stock assessment and fishery management. Fisheries Research 142, 86–99.

Mohn, R., 1999. The retrospective problem in sequential population analysis: An investigation using cod fishery and simulated data. ICES J. Mar. Sci. 56, 473-488.

Moser, S. M., Macintosch, D. J., Pripanapong, S., and Tongdee, N., 2002. Estimated growth of the mud crab *Sylla olivacea* in the Ranong mangrove ecosystem, Thailand, based on a tagging and recapture study. Marine and Freshwater Research, 53, 1083 – 1089.

Murphy, J. T., Rugolo, L. J., and Turnock, B. J., 2018. Estimation of annual, time-varying natural mortality and survival for Eastern Bering Sea snow crab (*Chionoecetes opilio*) with state-space population models. Fisheries Research 205, 122 – 131.

National Marine Fisheries Service. 2018. Fisheries Economics of the United States, 2016. U.S. Dept. of Commerce, NOAA Tech. Memo. NMFS-F/SPO-187, 243 p.

Pengilly, D., 2016. 2016 Stock Assessment and Fishery Evaluation Report for the King and Tanner Crab Fisheries of the Bering Sea and Aleutian Islands Regions. Pg. 723-772. In: Aleutian Islands Golden King Crab – 2016 Tier 5 Stock Assessment. North Pacific Fishery Management Council, Anchorage, AK, USA.

Pilling, G. M., Kirkwood, G. P., and Walker, S. G., 2002. An improved method for estimating individual growth variability in fish, and the correlation between von Bertalanffy growth parameters, Canadian Jouurnal of Fisheries and Aquatic Sciences 59, 424 – 431.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flanery, B. P., 2007. Numerical Recipes: The Art of Scientific Computing (3rd ed.), New York: Cambridge University Press.

Punt, A. E., Akselrud, C. A. and Cronin-Fine, L., 2017. The effects of applying mis-specified age- and size-structured models. Fisheries Research 188, 58 – 73.

Punt, A.E., 2017. Some insights into data weighting in integrated stock assessments. Fish. Res. 192, 52 – 65.

Punt, A. E., Buckworth, R. C., Dichmont, C. M. and Ye, Y. 2009. Performance of methods for estimating size-transition matrices using tag-recapture data. Marine and Freshwater Research 60, 168–182.

Punt, A. E., Campbell, R. A., and Smith, A. D. M. 2001. Evaluating empirical indicators and reference points for fisheries management: application to the broadbill sword fishery off eastern Australia. Marine Freshwater Research 52, 819-832

Punt, A. E., Castillo-Jordan, C., Hamel, O. S., Cope, J. M., Maunder, M. N. and Ianelli, J. N., 2021. Consequences of error in natural mortality and its estimation in stock assessment models. Fisheries Research 233, 105759. DOI: 10.1016/j.fishres.2020.105759.

Punt, A. E., Haddon, M. and McGarvey, R. 2016. Estimating growth within size-structured fishery stock assessments: What is the state of the art and what does the future look like? Fisheries Research 180, 147 – 160.

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

Punt, A. E., Huang, T-C. and Maunder, M. N. 2013. Review of integrated size-structured models for stock assessment of hard-to-age crustacean and mollusk species. ICES Journal of Marine Science 70, 16-33.

Punt, A. E., Hurtado-Ferro, F. and Whitten, A.R., 2014. Model selection for selectivity in fisheries stock assessment. Fisheries Research 158, 124 – 134.

Punt, A. E. and Kennedy, R. B., 1997. Population modelling of Tasmanian rock lobster, *Jasus edwardsii*, resources. Marine and Freshwater Research 48, 967-980.

Punt, A. E. and Szuwalski, C. 2012. How well can F_{MSY} and B_{MSY} be estimated using empirical measures of surplus production? Fisheries Research 134-136, 113-124.

Radomski, P. J., Bence, J. and Quinn, T., 2005. Comparison of virtual population analysis and statistical kill-at-age analysis for a recreational, kill-dominated fishery. Can. J. Fish. Aquat. Sci. 62, 436 – 452.

Sainsbury, K. J., 1980. Effect of individual variability on the von Bertalanffy growth equation. Canadian Journal of Fisheries and Aquatic Sciences 37, 241 – 247.

Sampson, D. B. and Scot, R. D. 2012. An exploration of the shape and stability of population – selectivity curves. Fish and Fisheries 13, 89 – 104.

Sampson, D. B., 2014. Fishery selection and its relevance to stock assessment and fishery management. Fisheries Research 158, 5 – 14.

Schnute, J. T. and Richards L. J., 1995. The influence of error on population estimates from catch-age models. Canadian Journal of Fisheries and Aquatic Sciences 52, 2063 – 2077.

Shibata, Y., Nagao, J., Narimatsu, Y. Morikawa, E., Suzuki, Y., Tokioka, S., Yamada, M., Kakehi, S. and Okamura, H. 2021. Estimating the maximum sustainable yield of snow crab (Chionoecetes opilio) off Tohoku, Japan via a state-space stock assessment model with time-varying natural mortaliy. Population Ecology 63, 41 – 60.

Siddeek, M. S. M., Zheng, J., Punt, A. E. and Vanek, V., 2016. Estimation of size-transition matrices with and without probability for Alaska golden king crab using tag-recapture data. Fisheries Research 180, 161 – 168.

Siddeek, M. S. M., Watson, L. J., Barnard, D. R., and Gish, R. K. 2008. Aleutian Islands golden king crab (*Lithodes aequispinus*) stock assessment. In:Stock Assessment and Fishery Evaluation Report for the King and Tanner Crab Fisheries of the Bering Sea and Aleutian Islands Regions. North Pacific Fishery Management Council, Anchorage, AK, pp. 443 – 495.

Smith, M. T. and Addison, J. T. 2003. Methods for stock assessment of crustacean fisheries. Fisheries Research 65, 231 – 256.

Spies, I., Aydin, K., Ianelli, J. N., Palsson, W., 2010. Assessment of the Arrowtooth Flounder Stock in the Gulf of Alaska. In: The Gulf of Alaska, NPFMC Gulf of Alaska SAFE. North Pacific Fishery Management Council, Seattle, WA.

Stockhausen, W.T., 2015. 2015 Stock Assessment and Fishery Evaluation Report for the Tanner Crab Fisheries of the Bering Sea and Aleutian Islands Regions. Pg. 295-440. In: Stock Assessment and Fishery Evaluation Report for the King and Tanner Crab Fisheries of the Bering Sea and Aleutian Islands Regions. North Pacific Fishery Management Council, Anchorage, AK, USA.

Stockhausen, W. T., 2019. 2019 Stock Assessment and Fishery Evaluation Report for the Tanner Crab Fisheries of the Bering Sea and Aleutian Islands Regions. In: Stock Assessment and Fishery Evaluation Report for the King and Tanner Crab Fisheries of the Bering Sea and Aleutian Islands Regions. North Pacific Fishery Management Council, Anchorage, AK, USA.

Stockhausen, W. T., 2020. 2020 Stock Assessment and Fishery Evaluation Report for the Tanner Crab Fisheries of the Bering Sea and Aleutian Islands Regions. In: Stock Assessment and Fishery Evaluation Report for the King and Tanner Crab Fisheries of the Bering Sea and Aleutian Islands Regions. North Pacific Fishery Management Council, Anchorage, AK, USA.

Swain, D. P., 2011. Life-history evolution and elevated natural mortality in a population of Atlantic cod (*Gadus morhua*). Evolutionary Applications 4, 18 – 29.

Szuwalski, C. D., Foy, R. J., and Turnock, B. J. 2014. 2014 Stock assessment and fishery evaluation report for the Pribilof Island red king crab fishery of the Bering Sea and Aleutian Islands regions. Pg 546-605. In: Stock Assessment and Fishery Evaluation Report for the King and Tanner Crab Fisheries of the Bering Sea and Aleutian Islands Regions. North Pacific Fishery Management Council 605 W. 4th Avenue, #306, Anchorage, AK 99501.

Szuwalski, C. and Turnock, B. J. 2016. Stock assessment of eastern Bering Sea snow crab. Pp 167-251. In: Stock Assessment and Fishery Evaluation Report for the King and Tanner Crab Fisheries of the Bering Sea and Aleutian Islands Regions. North Pacific Fishery Management Council 605 W. 4th Avenue, #306, Anchorage, AK 99501.

Taylor, I. G., and Methot, R. D., 2013. Hiding or dead? A computationally efficient model of selective fisheries mortality. Fisheries Research 142, 75 – 85.

Thompson, G. G., 1994. Confounding of gear selectivity and the natural mortality rate in cases where the former is a nonmonotone function of age. Canadian Journal of Fisheries and Aquatic Sciences 51, 2654 – 2664.

Thompson, G.G., 2015. 2015 Assessment of the Pacific Cod Stock in the Eastern Bering Sea. Pg. 251-470. In: 2015 North Pacific Groundfish Stock Assessment and Fishery Evaluation Reports for 2016. North Pacific Fishery Management Council, Anchorage AK, USA.

Troynikov, V., 1998. Probability density functions useful for parametrization of heterogeneity in growth and allometry data. Bulletin of Mathematical Biology 60, 1099 – 1122.

Turnock, B. J. and Rugolo, L. J. 2013. Stock assessment of eastern Bering Sea snow crab. In: Stock Assessment and Fishery Evaluation Report for the King and Tanner Crab Fisheries of the Bering Sea and Aleutian Islands Regions. North Pacific Fishery Management Council, Anchorage, AK, pp. 39–167.

Tyrrell, M. C., Link, J. S., Moustahfid, H., 2011. The importance of including predation in fish population models: Implications for biological reference points. Fisheries Research 108, 1 – 8.

Vetter, E. F., 1988. Estimation of natural mortality in fish stocks: a review. Fishery Bulletin 86, 25 – 43.

Walters, C. and Maguire, J. 1996. Lessons for stock assessment from the northern cod collapse. Reviews in Fish Biology and Fisheries 6, 125–137.

Wang, Y.-G. and Ellis, N., 2005. Maximum likelihood estimation of mortality and growth with individual variability from multiple length frequency data. U.S. National Marine Fisheries Service Fishery Bulletin 103, 380 – 391.

Wang, Y-G., Thomas, M. R., and Somers, I. F., 1995. A maximum likelihood approach for estimating growth from tag-recapture data. Canadian Journal of Fisheries and Aquatic Sciences 52, 252 – 259.

Wilberg, M. J., Bence, J.R., 2006. Performance of time-varying catchability estimators in statistical catch-at-age analysis. Can. J. Fish. Aquat. Sci. 63, 2275 – 2285.

Windsland, K., 2015. Total and natural mortality of red king crab (*Paralithodes camtschaticus*) in Norwegian waters: catch-curve analysis and indirect estimation methods. ICES Journal of Marine Science 72(2), 642 – 650.

Appendix A

Appendix A1: *How to obtain σ_{within_i} for the platoon method*

The ratio between $\sigma_{within \cdot i}$ versus $\sigma_{between \cdot i}$ ($\rho = {}^{\sigma_{within \cdot i}}/_{\sigma_{between \cdot i}}$) and $\sigma_{total \cdot i}$: is used to determine the values for $\sigma^2_{within \cdot i}$ and $\sigma^2_{between \cdot i}$,

$$\sigma_{within \cdot i} = \frac{\rho \sigma_{total \cdot i}}{(\rho^2+1)^{1/2}} \quad \text{(A1.1a)}$$

$$\sigma_{between \cdot i} = \frac{\sigma_{total \cdot i}}{(\rho^2+1)^{1/2}} \quad \text{(A1.1b)}$$

The value for $\sigma^2_{between \cdot i}$ is computed as the variance of a discrete distribution since there is a fixed number of platoons an individual can be in:

$$\sigma^2_{between \cdot i} = \sum_p \tau_p (\mu_{i_p} - \mu_{i_m})^2 \quad \text{(A1.2)}$$

where τ_p is the proportion of recruits that settle to platoon *p*, μ_{i_p} is the expected growth increment for platoon *p* when the initial size is $\bar{l}_i$, and μ_{i_m} is the expected growth increment for the "medium" platoon when the initial size is $\bar{l}_i$. For the case of three platoons, Equations A1.1b and A1.2 can be combined and simplified to determine the value for $\sigma^2_{total \cdot i}$:

$$\sigma^2_{total \cdot i} = [\sigma^2_L (1 - e^{-k})^2 (\tau_s + \tau_L)](\rho^2 + 1) \quad \text{(A1.3)}$$

where τ_s is the proportion of recruits to the "small" platoon and τ_L is the proportion of recruits to the "large" platoon. The value for σ_{within_i} is then derived by combining Equations A1.1a and A1.3. σ_{within_i} is therefore a function of σ_L, τ_s, τ_L and ρ. The parameter σ_L is estimated while τ_s and τ_L are fixed at 0.16, justified in main text. The value for ρ controls the extent to which the platoons overlap. This parameter was pre-specified because it is difficult to estimate (*sensu* Taylor and Methot, 2013).

Appendix A2: *Details on numerical integration method*

Method 3a: Variation in k

The number of integrals in Equation 2.4 reduces to three when only *k* varies among individuals, since L_∞ is fixed. However, since the variation in the ending size-class strictly depends on the initial size-class and *k*, the integral over growth variation (ending size-class) is also removed. The range over the ending size-class is incorporated into the range over which *k* is integrated by making it depend on L_∞. This altered range is defined by rearranging Equation 2.2 (with t = 1) to find the value for *k* given L_∞. The modified version of Equation 2.4 is:

$$f_k\left(l_j^+,\ l_{1i}\right) = -\ln\left(1 - \frac{l_j^+ - l_{1i}}{L_\infty - l_{1i}}\right)$$

$$X_{i,j} = \int_{l_i^-}^{l_i^+} \int_{f_k\left(l_j^-,\, l_{1i}\right)}^{f_k\left(l_j^+,\, l_{1i}\right)} F(l_{1i}, k)\mathrm{d}k\mathrm{d}l_{1i} \qquad \text{(A2.1)}$$

where the first integral is over the range for the initial size-class and the second integral is over the range for k. There are situations in which $f_k\left(l_j^+\right)$ is undefined and require special alterations (Appendix Table A.1). The function $F(l_{1i}, k)$ in Equation A2.1 is the product of the pdfs for l_i and k. It is assumed that $l_{1i} \sim U[l_i^-, l_i^+]$, and $k \sim lognormal\left(\bar{k}, \sigma_k\right)$. Therefore $F(l_{1i}, k)$ is:

$$F(l_{1i}, k) = \frac{1}{l_i^+ - l_i^-} \frac{1}{\sqrt{2\pi\sigma_k^2}} e^{\frac{\left(\ln(k) - \bar{k}\right)^2}{2\sigma_k^2}} \qquad \text{(A2.2)}$$

There are three estimable parameters (L_∞, $\bar{k}$ and σ_k) in Equation A2.1.

Method 3b: Variation in L_∞

The number of integrals in Equation 2.4 reduces by two when only L_∞ varies among individuals. However, now the range for the ending size-class is incorporated into the range over which L_∞ is integrated by making it depend on *k* by rearranging Equation 2.2 (with t = 1), i.e.:

$$f_{L\infty}\left(l_j^+,\ l_{1i}\right) = l_{1i} + \frac{l_j^+ - l_{1i}}{1 - e^{-k}}$$

$$X_{i,j} = \int_{l_i^-}^{l_i^+} \int_{f_{L\infty}\left(l_j^-,\, l_{1i}\right)}^{f_{L\infty}\left(l_j^+,\, l_{1i}\right)} F(l_{1i}, L_\infty)\mathrm{d}L_\infty\mathrm{d}l_{1i} \qquad \text{(A2.3)}$$

where the first integral is over the range for the initial size-class and the second integral is over the range for L_∞. There are situations in which $f_{L\infty}\left(l_j^+\right)$ is undefined and require special alterations (Appendix Table A.2). The function $F(l_{1i}, L_\infty)$ in Equation A2.3 is the product of the pdfs for l_{1i} and L_∞. The distribution for l_{1i} is the same as Method 3a while we assume $L_\infty \sim lognormal\left(\overline{L_\infty}, \sigma_{L_\infty}\right)$. Therefore $F(l_{1i}, L_\infty)$ is:

$$F(l_{1i}, L_\infty) = \frac{1}{l_i^+ - l_i^-} \frac{1}{\sqrt{2\pi\sigma_{L_\infty}^2}} e^{\frac{(\ln(L_\infty) - \overline{L_\infty})^2}{2\sigma_{L_\infty}^2}} \quad \text{(A2.4)}$$

There are three estimable parameters (k, $\overline{L_\infty}$ and σ_{L_∞}) in Equation A2.3.

Method 3c: Variation in both k and L_∞

The number of integrals in Equation 2.4 reduces by one when both growth parameters vary among individuals because the variation in growth comes directly from the initial size and growth parameters. The range for the ending size-class needs to be incorporated into one of the growth parameter ranges, resulting in that parameter range depending on the ending size-class and the other growth parameter. This study focuses on $f_{L\infty}(l_j^+)$ because initial experiments revealed that this leads to a better numerical approximation of the size-transition matrix. The modified Equation 2.4 is:

$$X_{i,j} = \int_{l_i^-}^{l_i^+} \int_0^\infty \int_{f_{L\infty}(l_j^-)}^{f_{L\infty}(l_j^+)} F(l_{1i}, L_\infty, k) \mathrm{d}L_\infty \mathrm{d}k \mathrm{d}l_{1i} \quad \text{(A2.5)}$$

where the first integral is over the range for the initial size-class, the second integral is over the range for L_∞ and the third integral is over the range for k. The function $F(l_{1i}, L_\infty, k)$ in Equation A2.5 is the product of the pdfs for l_{1i}, L_∞, and k. The same pdfs from Method 3a and 3b are used. Therefore $F(l_{1i}, L_\infty, k)$ is:

$$F(l_{1i}, L_\infty, k) = \frac{1}{l_i^+ - l_i^-} \frac{1}{\sqrt{2\pi\sigma_{L_\infty}^2}} e^{\frac{(\ln(L_\infty) - \overline{L_\infty})^2}{2\sigma_{L_\infty}^2}} \frac{1}{\sqrt{2\pi\sigma_k^2}} e^{\frac{(\ln(k) - \bar{k})^2}{2\sigma_k^2}} \quad \text{(A2.6)}$$

There are four estimable parameters ($\overline{L_\infty}$, σ_{L_∞}, $\bar{k}$ and σ_k) in Equation A2.5.

Numerical Integration

Equations A2.1, A2.3 and A2.5 were approximated using Gaussian quadrature. The equation for Gaussian quadrature is:

$$\int_{-1}^{1} f(x) \mathrm{d}x \approx \sum_{i=1}^{m} w_i f(x_i) \quad \text{(A2.7)}$$

where x_i are the evaluation points, w_i are the weights associated with the evaluation points and m is the number of evaluation points. The number of evaluation points is user-specified. This work uses 32 evaluation points. The evaluation points and associated weights are taken from Abramowitz and Stegun, (1972), where the evaluation points are determined based on the roots of a polynomial belonging to a class of orthogonal polynomials (Press *et al.*, 2007). Gaussian quadrature uses evaluation points that are bounded between -1 to 1. The evaluation points need to be transformed to cover the desired range when integrating over a range other than -1 to 1, which was accomplished using:

$$y_i = \frac{x_i+1}{2}(y^+ - y^-) + y^- \quad \text{(A2.8)}$$

where y_i is the evaluation point within the desired range, x_i is the original evaluation point between -1 and 1, y^+ is the upper limit of y, and y^- is the lower limit of y. The largest ending size-class has a problem because it is a plus group. This means that its upper bound is infinity. To address this problem, we numerically integrated additional ending size-classes that would directly follow the largest size-class in the size-transition matrix with their values being added to the largest size-class in the size-transition matrix. Prior work (unpublished) revealed that adding 10 additional size-classes when only *k* varies and 20 when L_∞ or both parameters vary is sufficient. Each row of size-transition matrix is normalized to ensure that the rows sum to one.

Appendix A3: *Equations for matrix and equilibrium diagnostic*

The Matrix Diagnostic

The matrix diagnostic is the sum of the absolute differences between the estimated and "true" values of the size-transition matrix divided by the number of positions in the upper triangular portion of the size-transition matrix times 100:

$$\frac{200}{m(m+1)} \sum_i \sum_{j \geq i} \left| X_{i,j}^T - X_{i,j}^{E,S} \right| \tag{A3.1}$$

where m is the number of size-classes, $X_{i,j}^T$ is the "true" probability of an individual transitioning from size-class i to size-class j in one time step and $X_{i,j}^{E,S}$ is the estimated value of $X_{i,j}^T$ produced by fitting to simulated data set S. The fraction in front of the sums equals 100 divided by the number of positions in the upper triangular portion of the size-transition matrix since no shrinking is allowed.

The Equilibrium diagnostic

The equilibrium size-structure is determined by taking the right eigenvector of the product of the size-transition matrix and a matrix with the exponentiated mortality rate on the diagonals:

$$N_i = \begin{cases} \frac{1}{\left(1 - X_{i,i} e^{-M}\right)} & if\ i = 1 \\ \frac{\sum_{j=1}^{i-1} X_{j,i} N_j e^{-M}}{\left(1 - X_{i,i} e^{-M}\right)} & if\ i \neq 1 \end{cases} \tag{A3.2}$$

where N_i is the relative number of animals in size-class i. The equilibrium diagnostic is the average of the sum of the absolute error between the "true" and estimated equilibrium size structures times 100 for each simulation:

$$\frac{100}{m} \sum_i \left| N_i^T - N_i^{E,S} \right| \tag{A3.3}$$

where N_i^T is the true proportion of individuals in size-class i at equilibrium and $N_i^{E,S}$ is the estimated proportion of individuals in size-class i at equilibrium for simulated data set S.

Table A.1. Cases in which $f_k(l_j^+,\ l_{1i})$ are undefined and how to address them.

Scenario	Solution
$l_i^- \geq L_\infty$	$X_{i,i} = 1\ \&\ X_{i,j} = 0$
$l_j^- > L_\infty$	$X_{i,j} = 0$
$l_j^+ > L_\infty$	$l_j^+ = (L_\infty - 0.001)$

Table A.2. Case in which $f_{L\infty}(l_j^+,\ l_{1i})$ are undefined and how to address them.

Scenario	Solution
$l_j^- < l_{1i}$	$f_{L\infty}(l_j^-) = l_{1i}$
$k \leq 0$	*k* is bounded to always be greater than 0
$l_j^+ < l_{1i}$	This can never happen since we only look at the upper triangle portion of the size-transition matrix meaning that $l_j^+ \geq l_{1i}$ is always true

Table A.3. Variance equations for the growth increment when t = 1 for each estimation method. The variance for Vary k, Vary Linf and Both were determined using the delta method.

Estimation Method	Variance Equation
Fixed	$CV^2 \cdot \left[\left(L_\infty - \bar{l_i}\right)(1 - e^{-k})\right]^2$
Platoon	$[\sigma_L^2(1 - e^{-k})^2(pr_s + pr_L)](\rho^2 + 1)$
Vary k	$\left[e^{-\bar{k}}\left(L_\infty - \bar{l_i}\right)\right]^2 \cdot Var(k) + \left(1 - e^{-\bar{k}}\right)^2 \cdot Var(l_i)$
Vary Linf	$\left(1 - e^{-\bar{k}}\right)^2 \cdot \left(Var(L_\infty) + Var(l_i)\right)$
Both	$\left[e^{-\bar{k}}\left(L_\infty - \bar{l_i}\right)\right]^2 \cdot Var(k) + \left(1 - e^{-\bar{k}}\right)^2 \cdot \left(Var(L_\infty) + Var(l_i)\right)$

Table A.4. Outline of the simulation study. The column on the right lists the number of options for each setting. The total number of scenarios is the product of the number of options for each setting.

Operating Model

Options	Possible Settings				#
Growth Parameters	Fixed		Individual growth curves		2
Maximum Time-at-Liberty	1 year		7 years		2
Initial Size-Class Distribution	From the Actual tagging data		Uniform Distribution		2
Sample Size	500	1500	3000	4000	4
				Total Number of Operating Model options =	32

Estimation Method

Options	Possible Settings								#
Method	Fixed	Platoon 0.75	Platoon 1.5	Platoon 2	Platoon 2.5	Platoon 4	Vary k	Vary L_∞	Both	9
Number of Size-Classes	18			9			6			3
								Total Number of Estimation Method options =		27

Total Number of Scenarios = 27·32 = 864

Table A.5. The parameter values used in the operating and estimation models. The sigma values are in log space. The growth parameters for OP1 (L_∞, k, and CV) were determined by minimizing the summed differences between the expected size-transition matrix created using the three growth parameters and actual size-transition matrix from the 2016 golden king crab assessment. OP2 requires four parameters ($\overline{L_\infty}$, σ_{L_∞}, $\bar{k}$ and σ_k) to define the lognormal distributions for L_∞ and k. These parameters are determined in the same fashion as for OP1.

Parameter	Value	Source
Natural Mortality (M)	0.18 yr^{-1}	2016 Stock assessment
Fishing effort (q)	0.65	Tuned to fit the times-at-liberty
OP1 Growth		
L_∞	421.238 mm	Tune to match
K	0.05778 yr^{-1}	the actual
CV	0.23338	size-transition matrix
OP2 Growth		
$\overline{L_\infty}$	365.8858 mm	Tune to match
σ_{L_∞}	0.0572	the actual
$\bar{k}$	0.05858 yr^{-1}	size-transition
σ_k	0.2285	matrix
Selectivity		
θ_{50}	126.47 mm	2016 stock assessment
θ_{95}	155.26 mm	

Table A.6. Summary of the performances of the estimation methods in terms of the matrix and equilibrium diagnostic. Methods were considered not appreciably different from the best/worst method for the matrix diagnostic if the median difference in matrix values was less than 0.1 from the best/worst method. For equilibrium diagnostic, the discrepancy threshold was 0.01. Bold values indicate that best/worst performance occurred for all sample sizes. Note that it is possible for a method to be in both the 'best' and 'worst' categories depending on the ranges for the values for the metrics.

(a) All individuals follow the same growth curve (OP1)

Metrics	Size-transition matrix entries			Equilibrium size- distribution		
Number of size-classes	18	9	6	18	9	6
Best performance						
Tagging / 7	**Fixed**, **Plat 1.5**, **2**, **2.5**, **4**, **Vary k**, **Vary Linf**, **Both**	**Vary k**, **Vary Linf**, Both	Fixed, **Vary k**, Both	**Fixed**	Fixed, **Vary k**	**Fixed**
Tagging / 1	Fixed, **Plat 0.75**, **1.5**, **2**, 2.5	**Plat 0.75**, **1.5**, **Vary Linf**	**Vary k**	Fixed, **Plat 2**, **2.5**, **4**, Vary k, Both	Vary k, **Vary Linf**, Both	**Fixed**, Vary k
Uniform / 7	**Fixed**, Plat 4, **Vary k**, Vary Linf, **Both**	Plat 0.75, **Vary k**, **Vary Linf**, Both	Plat 0.75, Vary k, Vary Linf	**Fixed**	Fixed, **Vary k**	**Fixed**, **Vary k**, Vary Linf
Uniform / 1	**Fixed**, Plat 2, 2.5, 4, **Vary k**, Vary Linf, Both	**Plat 1.5**, **2**, Vary Linf	**Plat 0.75**, 1.5, Vary k, Vary Linf	**Fixed**, Vary k	Vary k, Vary Linf, **Both**	**Vary k**, Vary Linf, Both
Poorest performance						
Tagging / 7	**Plat 0.75**	**Fixed**, Plat 0.75, 1.5, 2, 2.5, 4	**Plat 0.75**, 2.5, 4	**Plat 0.75**	**Plat 0.75**	**Plat 0.75**
Tagging / 1	**Fixed**, **Plat 2.5**, **4**, **Vary k**, **Vary Linf, Both**	**Fixed**	Plat 2.5, **4**	**Plat 0.75**	**Plat 0.75**	**Plat 0.75**
Uniform / 7	**Plat 0.75**	**Fixed**, Plat 1.5, 2, 2.5, 4, Both	Fixed, Plat 1.5, 2, **2.5**, **4**	**Plat 0.75**	**Plat 0.75**	**Plat 0.75**
Uniform / 1	**Plat 0.75**	**Fixed**	**Plat 4**, Both	**Plat 0.75**	**Plat 0.75**	**Plat 0.75**

(b) Individuals follow their own growth curves (OP2)

Metrics	Size-transition matrix entries			Equilibrium size- distribution		
Number of size-classes	18	9	6	18	9	6
Best performance						
Tagging / 7	**Plat 0.75**, **1.5**	**Plat 0.75**	Fixed, Vary k, **Vary Linf**, **Both**	**Plat 2**, **2.5**	Plat 4, Vary k, **Vary Linf**, Both	**Vary k**, Vary Linf
Tagging / 1	**Plat 1.5**, **2**, **2.5**, 4, **Vary Linf**, Both	**Plat 0.75**, 1.5, Vary Linf	**Vary k**, Vary Linf, Both	Plat 2.5, **4**, Vary k, **Vary Linf**, **Both**	**Vary k**, Vary Linf, Both	**Vary k**
Uniform / 7	**Plat 1.5**, 2	**Plat 1.5**	**Plat 1.5**, **2**	Plat 1.5, **2**, 2.5,	**Plat 2**, **2.5**, 4	Vary k, Vary Linf, Both
Uniform / 1	Fixed, **Vary k**, **Both**	**Plat 1.5**, **2**, 2.5, Vary k	**Plat 0.75**, 1.5	**Vary k**, Both	Plat 4, **Vary k**, Vary Linf, Both	Fixed, Vary k, Vary Linf
Poorest performance						
Tagging / 7	**Fixed**, **Vary k**, **Vary Linf**, **Both**	**Fixed**	**Plat 0.75**	**Plat 0.75**	**Plat 0.75**	**Plat 0.75**
Tagging / 1	**Fixed**, **Plat 0.75**, Vary k,	**Fixed**	**Plat 2.5**, **4**	**Plat 0.75**	**Plat 0.75**	**Plat 0.75**
Uniform / 7	Vary k, **Vary Linf**, Both	**Fixed**	Fixed, **Plat 0.75**, Vary k, Vary Linf, Both	**Plat 0.75**	**Plat 0.75**	**Plat 0.75**
Uniform / 1	**Plat 0.75**	**Fixed**	Fixed, Plat 2.5, 4, Vary k, **Both**	**Plat 0.75**	**Plat 0.75**	**Plat 0.75**

Table A.7. The matrix diagnostic for OP1. The bold, italicized and underlined values are the lowest values for the scenario concerned. The remaining values are differences from the lowest value.

Initial sizes/ Max time-at-liberty	Method	Sample Size: 500 / Number of Size-Classes: 18	9	6	Sample Size: 1500 / Number of Size-Classes: 18	9	6	Sample Size: 3000 / Number of Size-Classes: 18	9	6	Sample Size: 4000 / Number of Size-Classes: 18	9	6
Tagging / 7 years	*Fixed*	0.002	0.46	0.1	0.021	0.403	0.094	0.036	0.444	0.104	0.017	0.481	0.109
	Plat 0.75	0.336	0.299	0.657	0.307	0.324	0.519	0.336	0.379	0.681	0.31	0.386	0.674
	Plat 1.5	***0.911***	0.342	0.375	0.022	0.443	0.364	0.023	0.46	0.421	0.013	0.461	0.455
	Plat 2	0.004	0.291	0.422	***0.838***	0.352	0.447	***0.805***	0.394	0.52	***0.811***	0.384	0.558
	Plat 2.5	0.021	0.284	0.438	0.018	0.351	0.506	0.018	0.357	0.569	0.012	0.358	0.614
	Plat 4	0.054	0.367	0.501	0.061	0.367	0.587	0.079	0.372	0.643	0.062	0.366	0.7
	Vary k	0.029	0.095	***1.534***	0.061	0.037	***1.437***	0.074	0.017	***1.357***	0.058	0.043	***1.350***
	Vary Linf	0.005	***1.375***	0.129	0.01	***1.291***	0.16	0.031	***1.286***	0.187	0.024	***1.289***	0.234
	Both	0.093	0.114	0.311	0.045	0.042	0.108	0.074	0.06	0.07	0.058	0.036	0.14
Tagging / 1 year	*Fixed*	0.093	0.767	0.152	0.127	0.809	0.133	0.145	0.863	0.122	0.145	0.913	0.103
	Plat 0.75	0.056	***1.801***	0.252	0.048	***1.779***	0.268	0.046	***1.796***	0.287	0.037	***1.769***	0.237
	Plat 1.5	***0.970***	0.013	0.384	***0.895***	0.066	0.393	***0.905***	0.011	0.363	***0.917***	0.065	0.338
	Plat 2	0.065	0.11	0.454	0.074	0.158	0.542	0.066	0.127	0.518	0.067	0.16	0.511
	Plat 2.5	0.099	0.221	0.556	0.118	0.251	0.636	0.111	0.215	0.642	0.127	0.236	0.639
	Plat 4	0.168	0.361	0.631	0.198	0.388	0.771	0.198	0.37	0.8	0.206	0.387	0.806
	Vary k	0.132	0.236	***2.097***	0.163	0.231	***1.897***	0.186	0.248	***1.915***	0.19	0.281	***1.908***
	Vary Linf	0.131	0.025	0.278	0.157	0.028	0.259	0.149	0.011	0.247	0.15	0.044	0.244
	Both	0.148	0.257	0.24	0.182	0.116	0.25	0.185	0.162	0.2	0.185	0.172	0.187
Uniform / 7 years	*Fixed*	***0.642***	0.131	0.194	***0.431***	0.266	0.382	***0.407***	0.341	0.456	***0.397***	0.393	0.511
	Plat 0.75	0.621	0.083	0.13	0.768	0.191	***1.052***	0.745	0.129	***0.960***	0.726	0.14	***0.913***
	Plat 1.5	0.25	0.2	0.156	0.322	0.298	0.332	0.328	0.232	0.371	0.321	0.215	0.436
	Plat 2	0.166	0.208	0.206	0.204	0.247	0.393	0.209	0.178	0.449	0.201	0.17	0.514
	Plat 2.5	0.105	0.208	0.247	0.154	0.237	0.403	0.151	0.18	0.482	0.145	0.169	0.535
	Plat 4	0.059	0.216	0.315	0.131	0.267	0.421	0.106	0.214	0.514	0.099	0.231	0.546
	Vary k	0.013	***0.979***	***1.318***	0.081	***0.625***	0.129	0.063	0.017	0.204	0.064	0.021	0.268
	Vary Linf	0.031	0.038	0.048	0.104	0.011	0.159	0.095	***0.557***	0.228	0.092	***0.519***	0.289
	Both	0.095	0.15	0.173	0.098	0.014	0.181	0.077	0.004	0.253	0.063	0.01	0.274
Uniform / 1 year	*Fixed*	***0.579***	0.708	0.144	***0.371***	0.796	0.221	***0.318***	0.966	0.294	***0.335***	1	0.326
	Plat 0.75	0.452	0.21	0.077	0.625	0.416	***0.727***	0.659	0.478	***0.599***	0.621	0.497	***0.546***
	Plat 1.5	0.102	***0.908***	0.039	0.207	***0.573***	0.061	0.233	***0.474***	0.107	0.213	***0.441***	0.157
	Plat 2	0.042	0.023	0.159	0.144	0.075	0.216	0.154	0.074	0.286	0.131	0.094	0.337
	Plat 2.5	0.043	0.132	0.239	0.117	0.162	0.309	0.134	0.207	0.374	0.104	0.231	0.433
	Plat 4	0.065	0.315	0.339	0.112	0.376	0.422	0.138	0.409	0.491	0.109	0.462	0.554
	Vary k	0.031	0.117	***1.084***	0.03	0.138	0.072	0.036	0.196	0.114	0.027	0.251	0.143
	Vary Linf	0.059	0.079	0.094	0.125	0.045	0.189	0.14	0.08	0.231	0.117	0.12	0.288
	Both	0.118	0.132	0.266	0.073	0.14	0.186	0.062	0.176	0.176	0.046	0.151	0.202

Table A.8. The matrix diagnostic for OP2. The bold, italicized and underlined values are the lowest values for the scenario concerned. The remaining values are differences from the lowest value.

Initial sizes/ Max time-at-liberty	Method	Sample Size: 500			1500			3000			4000		
		Number of Size-Classes: 18	9	6	18	9	6	18	9	6	18	9	6
Tagging / 7 years	*Fixed*	0.629	1.342	0.062	0.664	1.427	0.137	0.667	1.5	0.221	0.671	1.507	0.217
	Plat 0.75	0.033	***1.235***	1.485	0.06	***1.191***	1.607	0.049	***1.112***	1.535	0.063	***1.094***	1.502
	Plat 1.5	***1.145***	0.181	0.527	***1.069***	0.254	0.603	***1.040***	0.26	0.537	***1.034***	0.245	0.519
	Plat 2	0.171	0.438	0.305	0.169	0.44	0.445	0.201	0.494	0.366	0.2	0.484	0.323
	Plat 2.5	0.307	0.648	0.247	0.315	0.607	0.377	0.35	0.698	0.319	0.346	0.688	0.255
	Plat 4	0.506	0.912	0.204	0.547	0.924	0.321	0.576	1.013	0.286	0.572	1.005	0.266
	Vary k	0.646	0.636	***2.773***	0.703	0.661	0.057	0.701	0.736	0.119	0.705	0.745	0.134
	Vary Linf	0.667	0.551	0.016	0.697	0.572	***2.659***	0.724	0.622	***2.643***	0.721	0.623	***2.619***
	Both	0.658	0.659	0.018	0.699	0.609	0.019	0.716	0.689	0.067	0.715	0.698	0.085
Tagging / 1 year	*Fixed*	0.629	1.342	0.062	0.664	1.427	0.137	0.667	1.5	0.221	0.671	1.507	0.217
	Plat 0.75	0.259	0.058	0.399	0.292	***1.552***	0.197	0.304	***1.482***	0.173	0.31	***1.400***	0.142
	Plat 1.5	***1.118***	0.092	0.462	***0.989***	0.142	0.388	***0.937***	0.119	0.378	***0.921***	0.172	0.345
	Plat 2	0.018	0.197	0.508	0.047	0.255	0.469	0.044	0.239	0.476	0.047	0.338	0.451
	Plat 2.5	0.028	0.278	0.571	0.073	0.343	0.533	0.084	0.376	0.547	0.092	0.461	0.506
	Plat 4	0.073	0.469	0.662	0.135	0.525	0.602	0.15	0.571	0.618	0.162	0.663	0.568
	Vary k	0.146	0.184	***3.464***	0.208	0.265	***3.277***	0.222	0.324	***3.204***	0.248	0.421	***3.144***
	Vary Linf	0.018	***1.705***	0.204	0.066	0.085	0.091	0.098	0.133	0.077	0.11	0.216	0.091
	Both	0.093	0.227	0.154	0.135	0.258	0.096	0.175	0.289	0.084	0.194	0.369	0.057
Uniform / 7 years	*Fixed*	0.529	1.105	0.326	0.591	1.297	0.312	0.604	1.303	0.368	0.629	1.315	0.396
	Plat 0.75	0.362	0.247	0.755	0.392	0.158	0.564	0.395	0.156	0.452	0.418	0.226	0.421
	Plat 1.5	***0.872***	***0.915***	***1.607***	***0.783***	***0.782***	***1.509***	***0.762***	***0.770***	***1.447***	***0.745***	***0.718***	***1.444***
	Plat 2	0.09	0.15	0.053	0.101	0.262	0.035	0.137	0.299	0.035	0.159	0.288	0.034
	Plat 2.5	0.178	0.356	0.116	0.26	0.471	0.089	0.283	0.497	0.125	0.305	0.504	0.129
	Plat 4	0.387	0.644	0.302	0.497	0.792	0.25	0.513	0.811	0.296	0.535	0.819	0.311
	Vary k	0.541	0.471	0.266	0.592	0.61	0.306	0.605	0.588	0.351	0.626	0.598	0.371
	Vary Linf	0.638	0.617	0.266	0.712	0.764	0.324	0.732	0.746	0.366	0.762	0.819	0.371
	Both	0.62	0.564	0.569	0.597	0.675	0.379	0.621	0.607	0.424	0.646	0.607	0.429
Uniform / 1 year	*Fixed*	0.096	0.872	0.212	0.085	1.073	0.272	0.13	1.173	0.43	0.138	1.255	0.475
	Plat 0.75	0.987	0.401	***1.137***	1.067	0.399	***0.724***	1.159	0.51	***0.525***	1.188	0.551	***0.419***
	Plat 1.5	0.402	0.044	0.061	0.509	0.005	0.187	0.587	***0.440***	0.284	0.609	***0.397***	0.268
	Plat 2	0.269	***0.868***	0.149	0.332	0.04	0.289	0.4	0.039	0.377	0.423	0.078	0.402
	Plat 2.5	0.198	0.052	0.205	0.247	0.107	0.327	0.311	0.148	0.432	0.327	0.208	0.471
	Plat 4	0.139	0.219	0.252	0.187	0.299	0.424	0.225	0.367	0.506	0.241	0.445	0.575
	Vary k	***0.464***	0.018	0.228	***0.307***	***0.590***	0.261	***0.226***	0.089	0.433	***0.202***	0.114	0.491
	Vary Linf	0.229	0.206	0.194	0.304	0.193	0.243	0.376	0.237	0.368	0.395	0.307	0.423
	Both	0.083	0.207	0.402	0.016	0.12	0.514	0.008	0.123	0.503	0.004	0.151	0.589

Table A.9. Equilibrium diagnostic for OP1. The bold, italicized and underlined values are the lowest values for the scenario concerned. The remaining values are differences from the lowest value

Initial sizes/ Max time-at-liberty	Method	Sample Size: 500 / Number of Size-Classes: 18	500 / 9	500 / 6	1500 / 18	1500 / 9	1500 / 6	3000 / 18	3000 / 9	3000 / 6	4000 / 18	4000 / 9	4000 / 6
Tagging / 7 years	*Fixed*	***0.169***	0.01	***0.441***	***0.156***	0.01	***0.408***	***0.155***	0.011	***0.384***	***0.153***	0.014	***0.391***
	Plat 0.75	0.36	0.469	1.026	0.349	0.489	0.967	0.347	0.496	1.073	0.347	0.5	1.063
	Plat 1.5	0.086	0.308	0.32	0.087	0.339	0.332	0.082	0.352	0.337	0.085	0.345	0.341
	Plat 2	0.049	0.203	0.213	0.046	0.238	0.234	0.041	0.25	0.245	0.044	0.241	0.251
	Plat 2.5	0.036	0.14	0.172	0.034	0.172	0.19	0.028	0.183	0.201	0.029	0.178	0.204
	Plat 4	0.03	0.066	0.117	0.03	0.084	0.139	0.021	0.097	0.148	0.021	0.093	0.149
	Vary k	0.012	***0.292***	0.026	0.012	***0.260***	0.023	0.011	***0.257***	0.022	0.014	***0.257***	0.025
	Vary Linf	0.034	0.016	0.047	0.037	0.034	0.058	0.028	0.023	0.059	0.028	0.03	0.062
	Both	0.037	0.029	0.056	0.021	0.013	0.065	0.017	0.011	0.053	0.018	0.013	0.052
Tagging / 1 year	*Fixed*	***0.197***	0.054	***0.704***	0.008	0.047	***0.665***	0.015	0.079	0.006	0.017	0.082	0.007
	Plat 0.75	0.266	0.345	0.865	0.268	0.376	0.878	0.28	0.379	0.907	0.279	0.373	0.889
	Plat 1.5	0.032	0.147	0.394	0.025	0.152	0.409	0.034	0.158	0.397	0.033	0.15	0.407
	Plat 2	0.007	0.09	0.279	0.002	0.082	0.3	0.006	0.087	0.298	0.006	0.081	0.308
	Plat 2.5	0.001	0.061	0.215	***0.181***	0.05	0.238	**0.174**	0.053	0.244	***0.175***	0.051	0.257
	Plat 4	0.01	0.023	0.151	0.001	0.026	0.167	0.001	0.021	0.182	0.002	0.022	0.19
	Vary k	0.004	***0.383***	0.025	0.009	***0.380***	0.008	0.016	0.012	***0.666***	0.017	0.009	***0.659***
	Vary Linf	0.017	0.002	0.071	0.015	0.004	0.056	0.013	***0.366***	0.058	0.014	0.001	0.063
	Both	0.014	0.023	0.072	0.01	0.011	0.054	0.011	0.007	0.045	0.012	***0.369***	0.052
Uniform / 7 years	*Fixed*	***0.113***	***0.214***	0.002	***0.074***	0.014	***0.235***	***0.074***	0.018	***0.206***	***0.074***	0.033	***0.210***
	Plat 0.75	0.382	0.416	0.327	0.407	0.481	0.295	0.397	0.46	0.303	0.396	0.462	0.296
	Plat 1.5	0.108	0.282	0.091	0.13	0.33	0.115	0.122	0.327	0.126	0.117	0.324	0.12
	Plat 2	0.072	0.186	0.065	0.092	0.233	0.091	0.083	0.226	0.102	0.078	0.222	0.096
	Plat 2.5	0.063	0.127	0.06	0.076	0.172	0.08	0.069	0.158	0.092	0.064	0.153	0.087
	Plat 4	0.06	0.065	0.045	0.069	0.091	0.073	0.063	0.064	0.076	0.059	0.065	0.07
	Vary k	0.012	0.001	***0.261***	0.022	***0.132***	0.003	0.02	***0.125***	0.007	0.021	***0.114***	0.004
	Vary Linf	0.068	0.02	0.009	0.078	0.04	0.026	0.072	0.027	0.03	0.069	0.036	0.029
	Both	0.046	0.033	0.051	0.05	0.025	0.019	0.041	0.013	0.027	0.043	0.015	0.012
Uniform / 1 year	*Fixed*	***0.108***	0.08	0.01	***0.068***	0.097	0.013	***0.055***	0.114	0.018	***0.055***	0.134	0.021
	Plat 0.75	0.326	0.351	0.461	0.354	0.391	0.442	0.367	0.398	0.438	0.363	0.397	0.43
	Plat 1.5	0.047	0.151	0.131	0.076	0.18	0.132	0.084	0.184	0.134	0.085	0.182	0.127
	Plat 2	0.018	0.088	0.086	0.048	0.105	0.089	0.055	0.11	0.076	0.055	0.109	0.082
	Plat 2.5	0.017	0.05	0.075	0.042	0.065	0.069	0.048	0.063	0.057	0.047	0.062	0.064
	Plat 4	0.03	0.022	0.061	0.04	0.016	0.035	0.045	0.012	0.034	0.042	0.011	0.043
	Vary k	0.004	***0.212***	***0.278***	0.006	***0.142***	***0.232***	0.008	0.005	***0.205***	0.01	0.014	0.008
	Vary Linf	0.046	0.006	0.006	0.057	0.013	0.013	0.063	0.004	0.006	0.063	0.006	***0.210***
	Both	0.04	0.007	0.025	0.03	0.008	0.017	0.032	***0.125***	0.005	0.028	***0.120***	0.016

Table A.10. Equilibrium diagnostic for OP2. The bold, italicized and underlined values are the lowest values for the scenario concerned. The remaining values are differences from the lowest value.

Initial sizes/ Max time-at-liberty	Method	Sample Size: 500 / Number of Size-Classes: 18	9	6	Sample Size: 1500 / Number of Size-Classes: 18	9	6	Sample Size: 3000 / Number of Size-Classes: 18	9	6	Sample Size: 4000 / Number of Size-Classes: 18	9	6
Tagging / 7 years	*Fixed*	0.13	0.064	0.051	0.128	0.071	0.052	0.134	0.076	0.05	0.137	0.073	0.05
	Plat 0.75	0.293	0.259	1.335	0.32	0.279	1.344	0.32	0.29	1.331	0.322	0.292	1.311
	Plat 1.5	0.012	0.074	0.625	0.012	0.082	0.638	0.013	0.08	0.624	0.012	0.069	0.599
	Plat 2	***0.219***	0.038	0.409	***0.202***	0.046	0.42	***0.197***	0.046	0.408	***0.196***	0.046	0.379
	Plat 2.5	0.007	0.025	0.294	0.001	0.029	0.305	0.002	0.035	0.297	0.004	0.034	0.268
	Plat 4	0.031	0.007	0.143	0.027	0.014	0.162	0.032	0.019	0.156	0.036	0.017	0.135
	Vary k	0.109	***0.434***	***0.891***	0.108	0.005	***0.869***	0.116	0.01	***0.853***	0.118	0.011	***0.856***
	Vary Linf	0.049	0.006	0.026	0.056	***0.420***	0.022	0.062	0.001	0.016	0.069	***0.406***	0.01
	Both	0.088	0.018	0.041	0.093	0.005	0.02	0.097	***0.410***	0.016	0.103	0.011	0.024
Tagging / 1 year	*Fixed*	0.034	0.032	0.161	0.038	0.024	0.14	0.038	0.044	0.124	0.045	0.05	0.137
	Plat 0.75	0.343	0.373	0.76	0.371	0.351	0.696	0.375	0.351	0.711	0.37	0.337	0.709
	Plat 1.5	0.052	0.176	0.627	0.059	0.157	0.572	0.045	0.138	0.577	0.044	0.124	0.574
	Plat 2	0.03	0.107	0.522	0.027	0.098	0.468	0.013	0.087	0.483	0.013	0.082	0.482
	Plat 2.5	0.013	0.088	0.445	0.013	0.069	0.401	0.003	0.061	0.419	0.004	0.061	0.416
	Plat 4	0.006	0.056	0.342	0.006	0.048	0.301	0.001	0.036	0.32	0.002	0.039	0.317
	Vary k	0.01	***0.456***	***1.154***	0.009	***0.443***	***1.112***	0.01	***0.425***	***1.066***	0.015	***0.423***	***1.045***
	Vary Linf	0.005	0.013	0.137	0.006	0.01	0.104	0.004	0.005	0.108	0.007	0.005	0.116
	Both	***0.214***	0.035	0.142	***0.187***	0.016	0.101	***0.182***	0.006	0.094	***0.177***	0.007	0.11
Uniform / 7 years	*Fixed*	0.119	0.106	0.025	0.134	0.121	0.006	0.135	0.11	0.003	0.139	0.114	0.008
	Plat 0.75	0.341	0.289	0.663	0.376	0.254	0.521	0.37	0.232	0.479	0.371	0.235	0.467
	Plat 1.5	0.009	0.055	0.344	0.01	0.035	0.233	0.007	0.03	0.209	0.009	0.033	0.212
	Plat 2	***0.156***	0.005	0.2	***0.139***	***0.301***	0.126	***0.137***	0.003	0.126	***0.135***	0.001	0.127
	Plat 2.5	0.004	***0.293***	0.156	0.007	0.002	0.094	0.009	***0.307***	0.091	0.013	***0.301***	0.093
	Plat 4	0.028	0.005	0.105	0.041	0.029	0.062	0.047	0.017	0.07	0.054	0.014	0.07
	Vary k	0.099	0.025	0.013	0.114	0.041	***0.662***	0.118	0.029	***0.660***	0.122	0.03	***0.654***
	Vary Linf	0.072	0.028	***0.619***	0.079	0.049	0.007	0.09	0.036	0.012	0.095	0.04	0.014
	Both	0.121	0.047	0.103	0.116	0.051	0.005	0.124	0.034	0.021	0.129	0.034	0.021
Uniform / 1 year	*Fixed*	0.021	0.1	***0.366***	0.028	0.115	0.011	0.035	0.123	0.01	0.038	0.129	0.023
	Plat 0.75	0.425	0.417	0.508	0.455	0.441	0.502	0.478	0.441	0.504	0.481	0.434	0.515
	Plat 1.5	0.077	0.143	0.194	0.104	0.154	0.191	0.118	0.151	0.192	0.122	0.147	0.198
	Plat 2	0.048	0.069	0.137	0.066	0.069	0.12	0.079	0.06	0.131	0.083	0.061	0.14
	Plat 2.5	0.035	0.033	0.103	0.049	0.036	0.081	0.062	0.026	0.103	0.065	0.027	0.115
	Plat 4	0.031	0.01	0.071	0.039	0.014	0.057	0.049	0.007	0.073	0.052	0.013	0.084
	Vary k	***0.094***	***0.200***	0.007	***0.062***	***0.141***	0.005	***0.044***	***0.125***	0.015	0.003	***0.127***	0.022
	Vary Linf	0.039	0.009	0.012	0.05	0.02	***0.325***	0.061	0.011	***0.330***	0.063	0.014	***0.327***
	Both	0.014	0.022	0.099	0.004	0.015	0.05	0.002	0.003	0.021	***0.040***	0.007	0.036

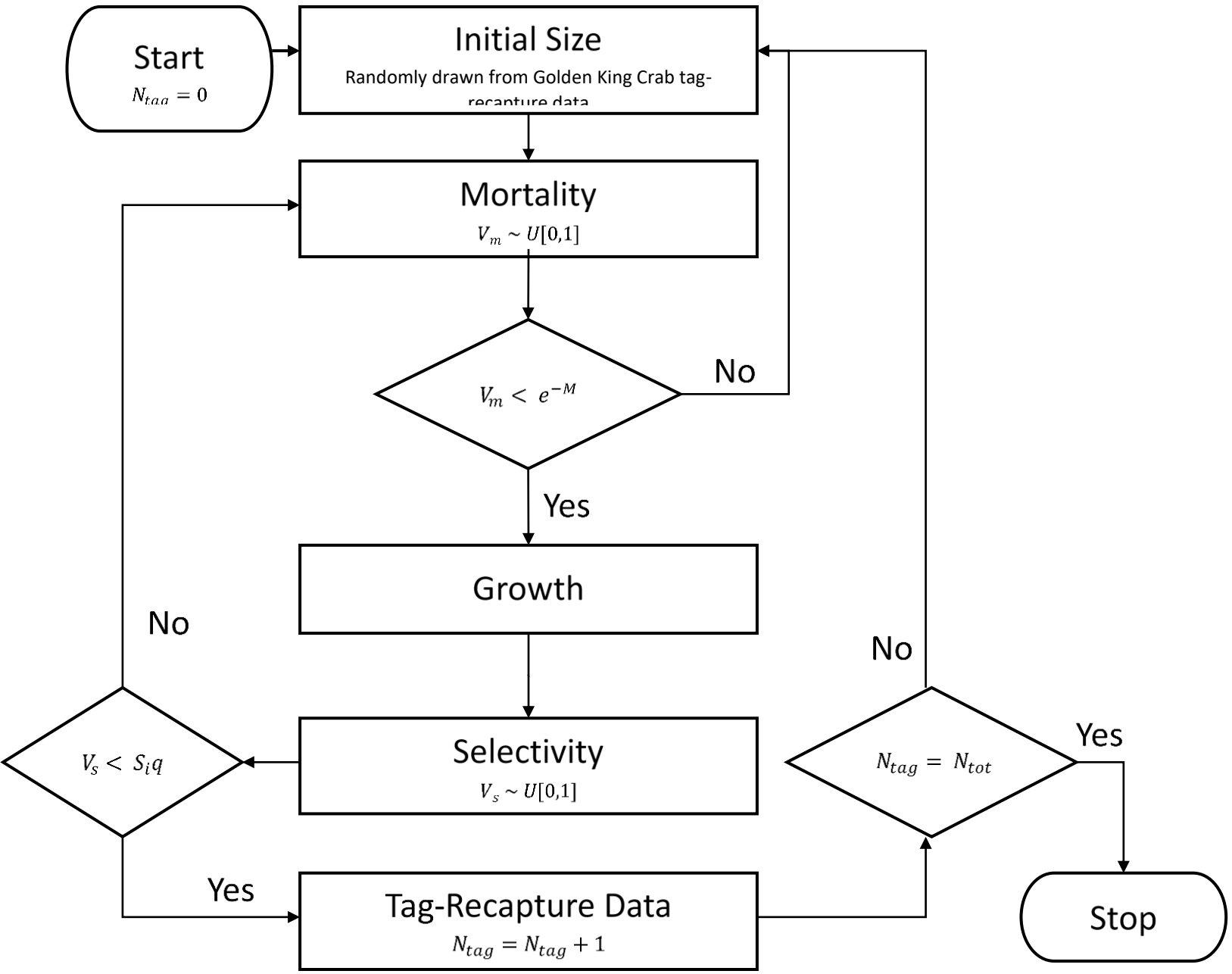

Figure A.1. Flow chart of the individual-based operating model used to generate the tag-recapture data. N_{tag} is the number of tag-recapture data points generated. Vm is a value used to determine if an individual survives. M is natural mortality. Vs is a value used to determine if an individual is caught by the fishery. S_i is the selectivity for an individual of size i. q is fishing effort. N_{tot} is the desired number of simulated tag-recapture data points.

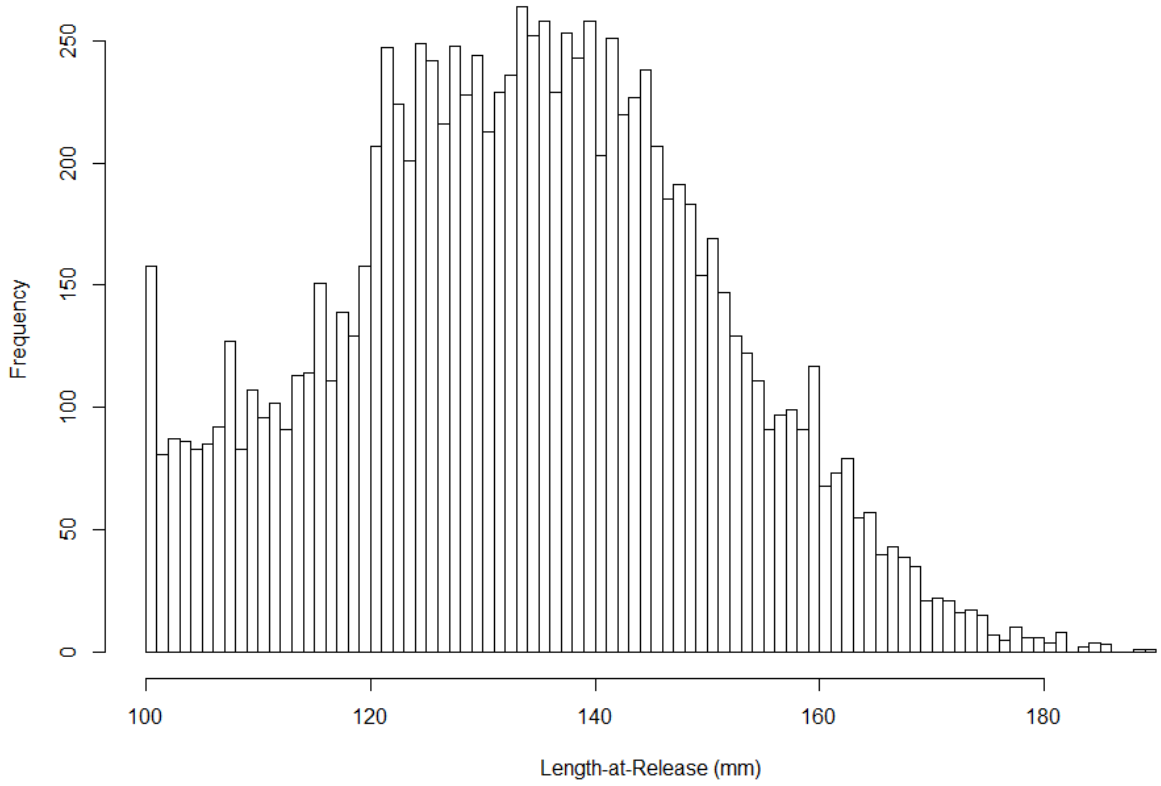

Figure A.2. Distribution of the sizes-at-release from the actual tag-recapture data set for golden king crab.

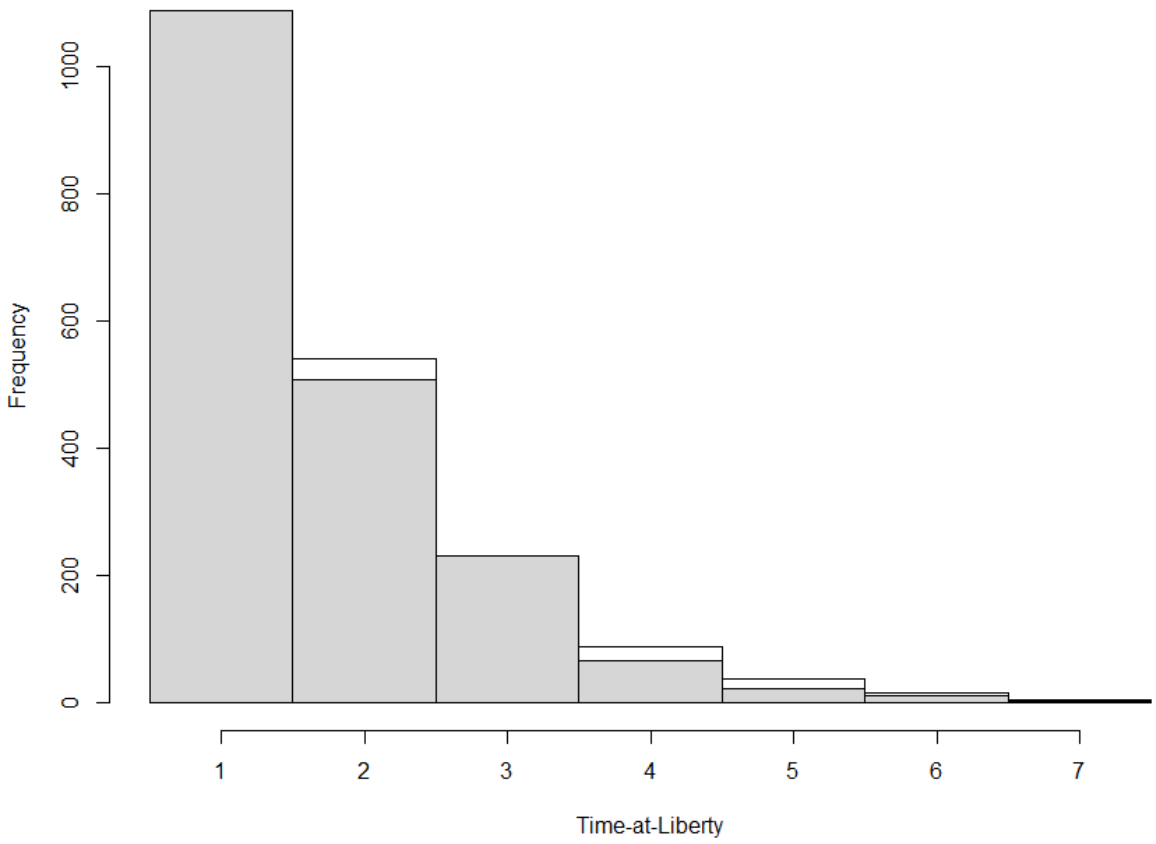

Figure A.3. Time-at-liberty from actual tag-recapture data for golden king crabs in the Aleutian Island regions of Alaska (dark grey) and the time-at-liberty from an example generated data set (white). The light grey is where the two histograms overlap.

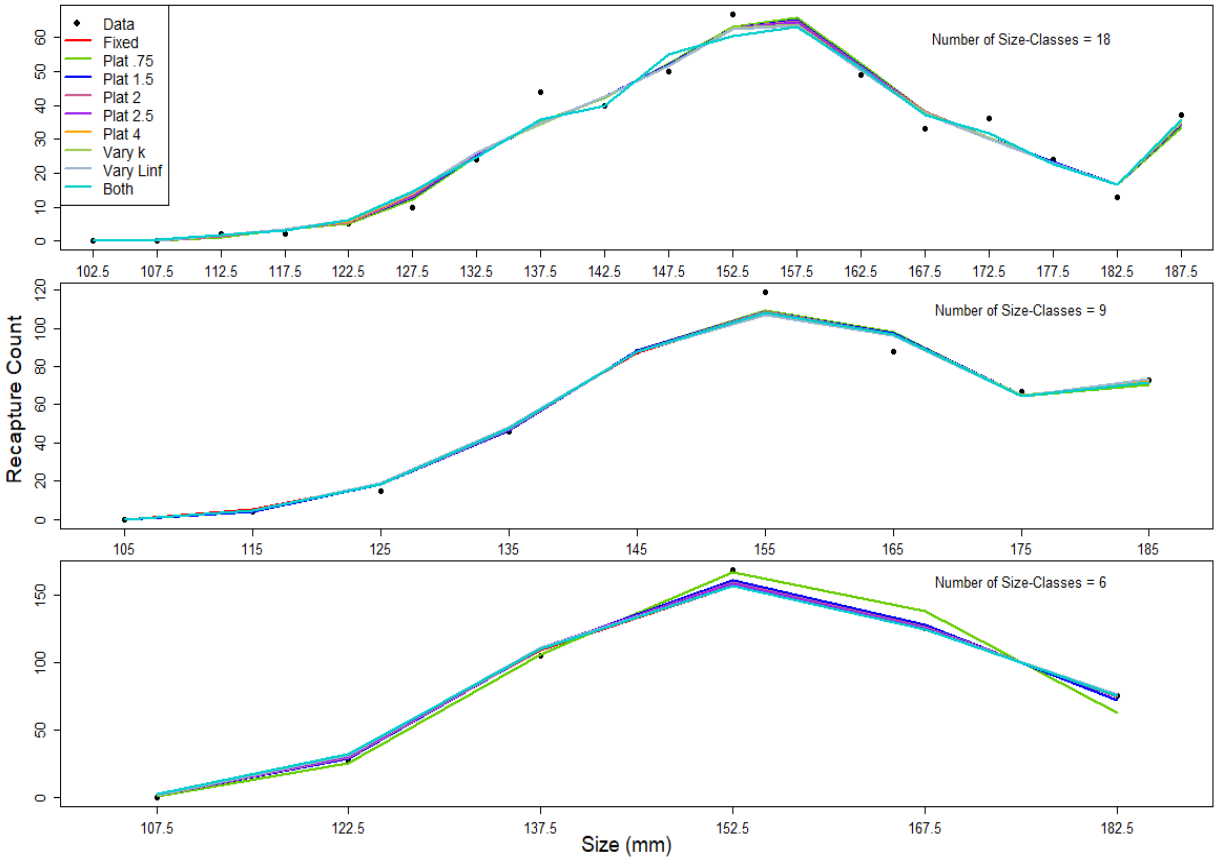

Figure A.4. An example of how well each estimation method fits the tag-recapture data for operating model Tag-7. The black dots represent the number of sizes-at-recapture within each size-class. The lines represent the predicted number of sizes-at-recapture within each size-class. The top and bottom panels had data generated when individuals followed their own growth curve. The middle panel had data generated when all individuals followed the same growth curve. All data sets have 500 data points.

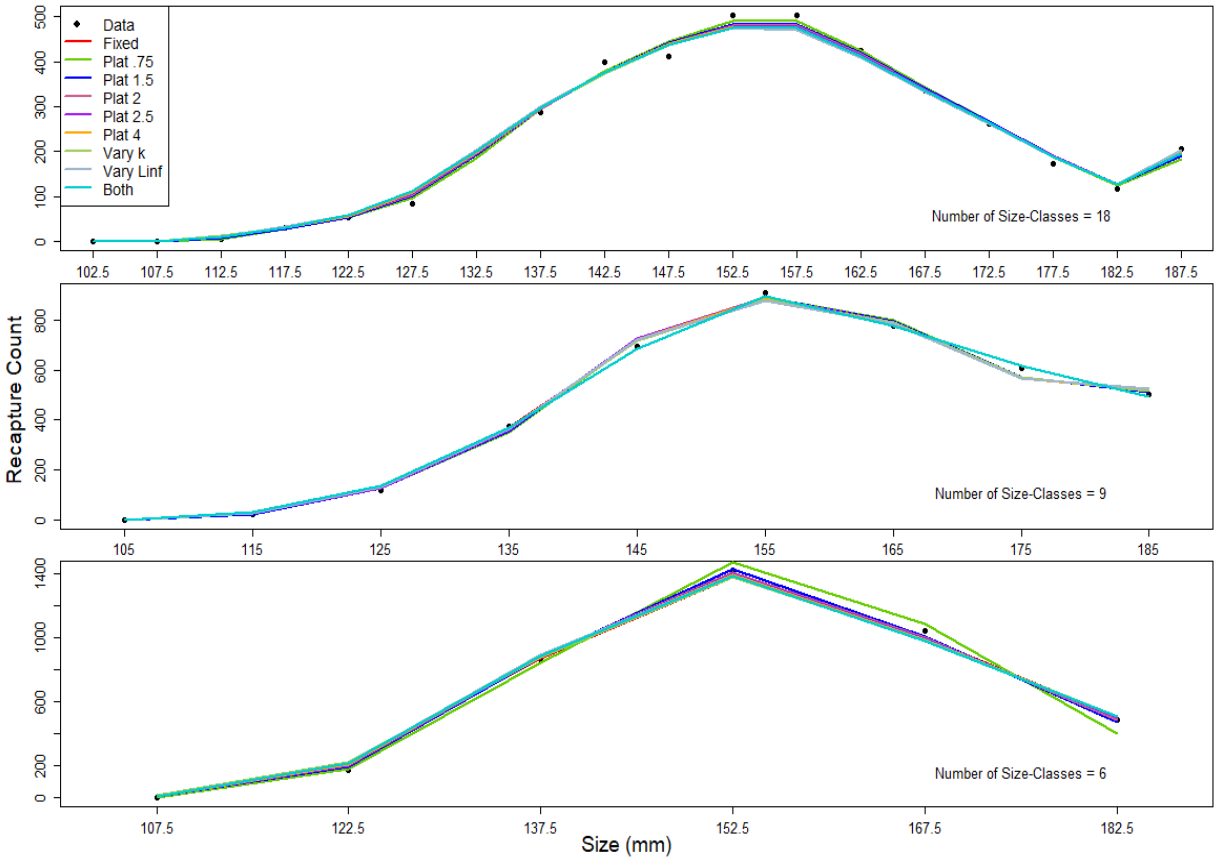

Figure A.5. As for Figure A.4, except that all data sets have 4,000 data points.

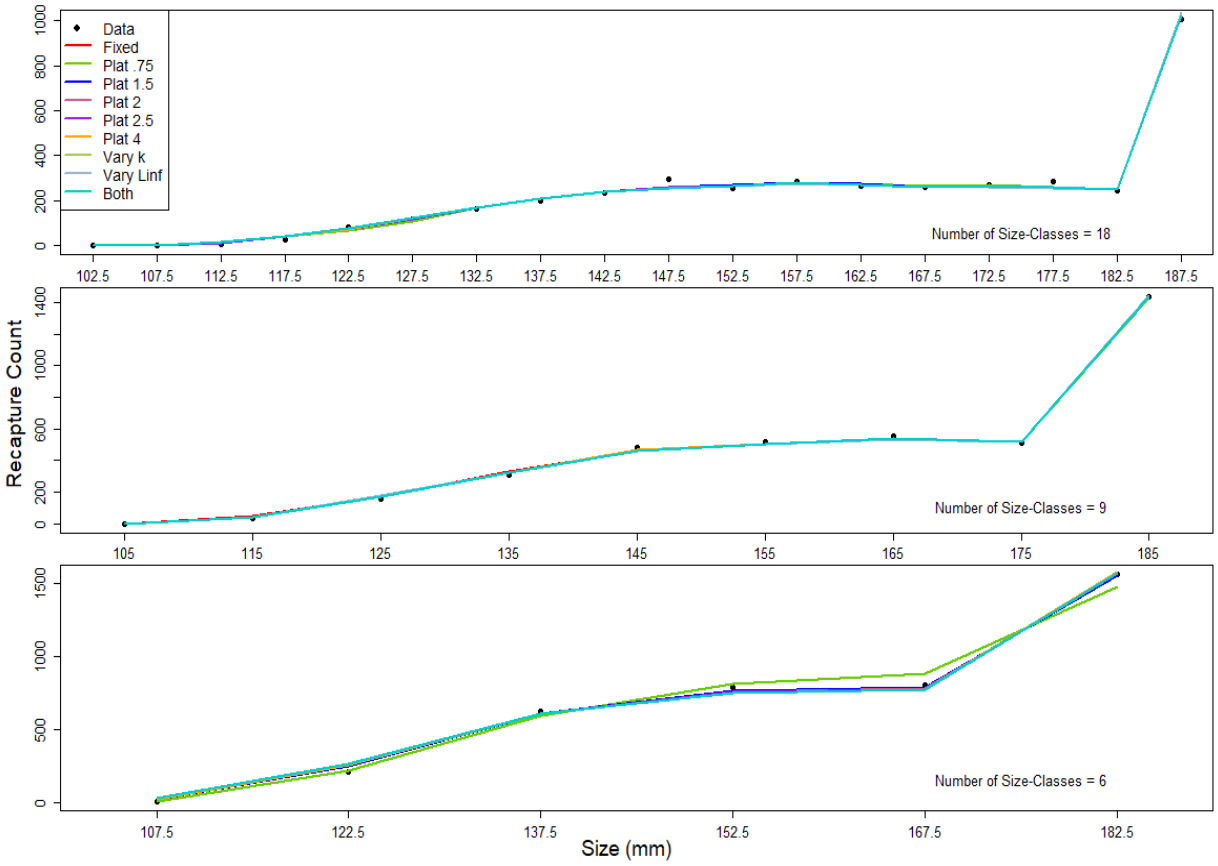

Figure A.6. As for Figure A.5, except that the results pertain to operating model Uniform-7.

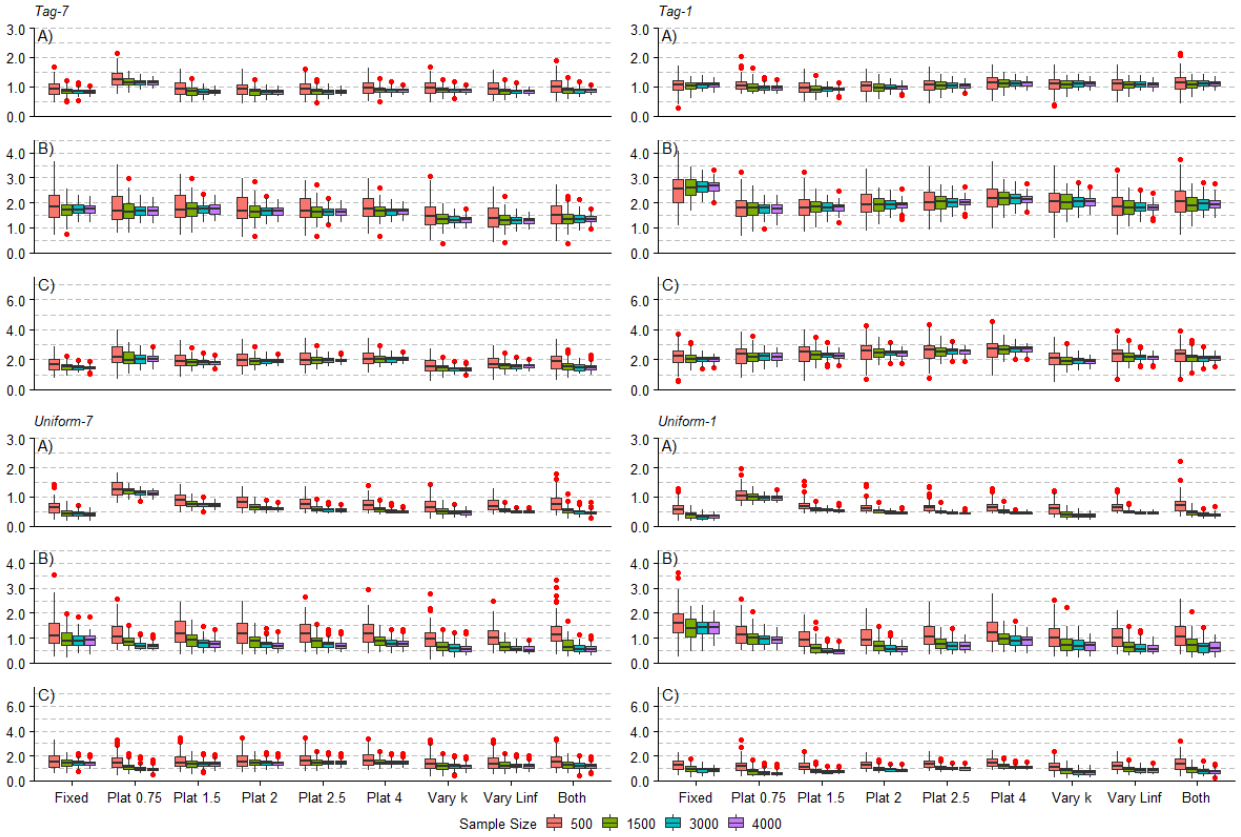

Figure A.7. Boxplots of the matrix diagnostic for OP1. The four blocks of panels depend on initial size distribution (*Tag* for the actual tagging data for golden king crab and *Uniform* for a uniform distribution) and maximum time-at-liberty (*7* for seven years and *1* for one year). Within each block there are three boxplots whose size-transition matrices differ in terms of number of size-classes (A: 18, B: 9; C: 6).

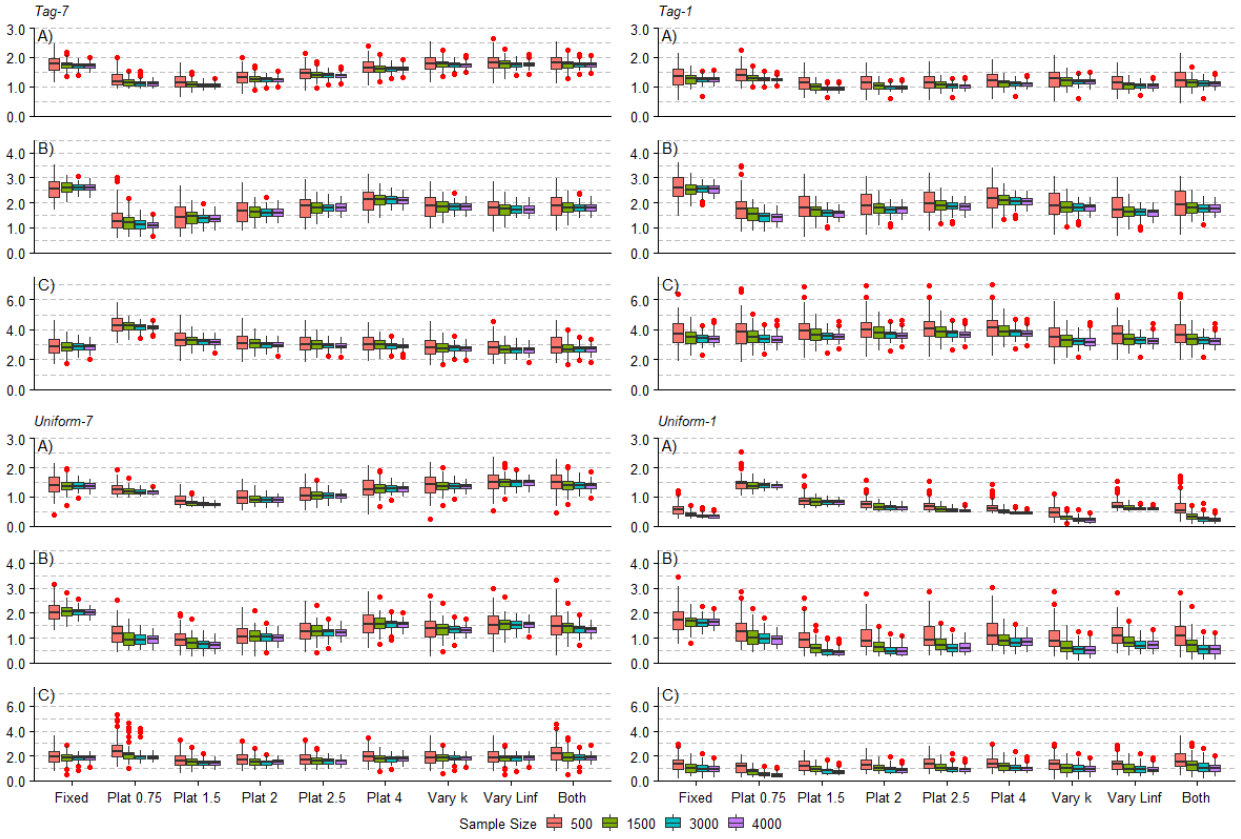

Figure A.8. As for Figure A.7, except that the operating model is OP2.

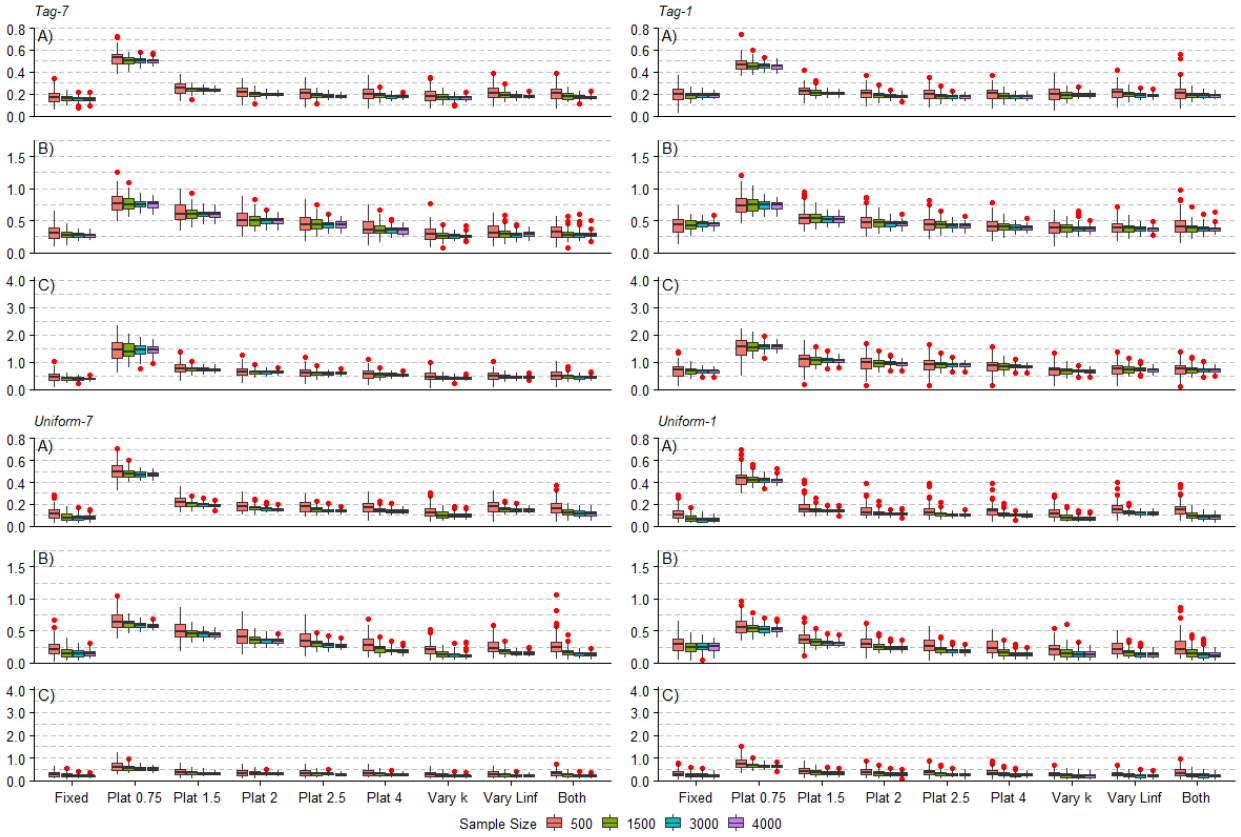

Figure A.9. As for Figure A.7, except that the metric is the equilibrium diagnostic.

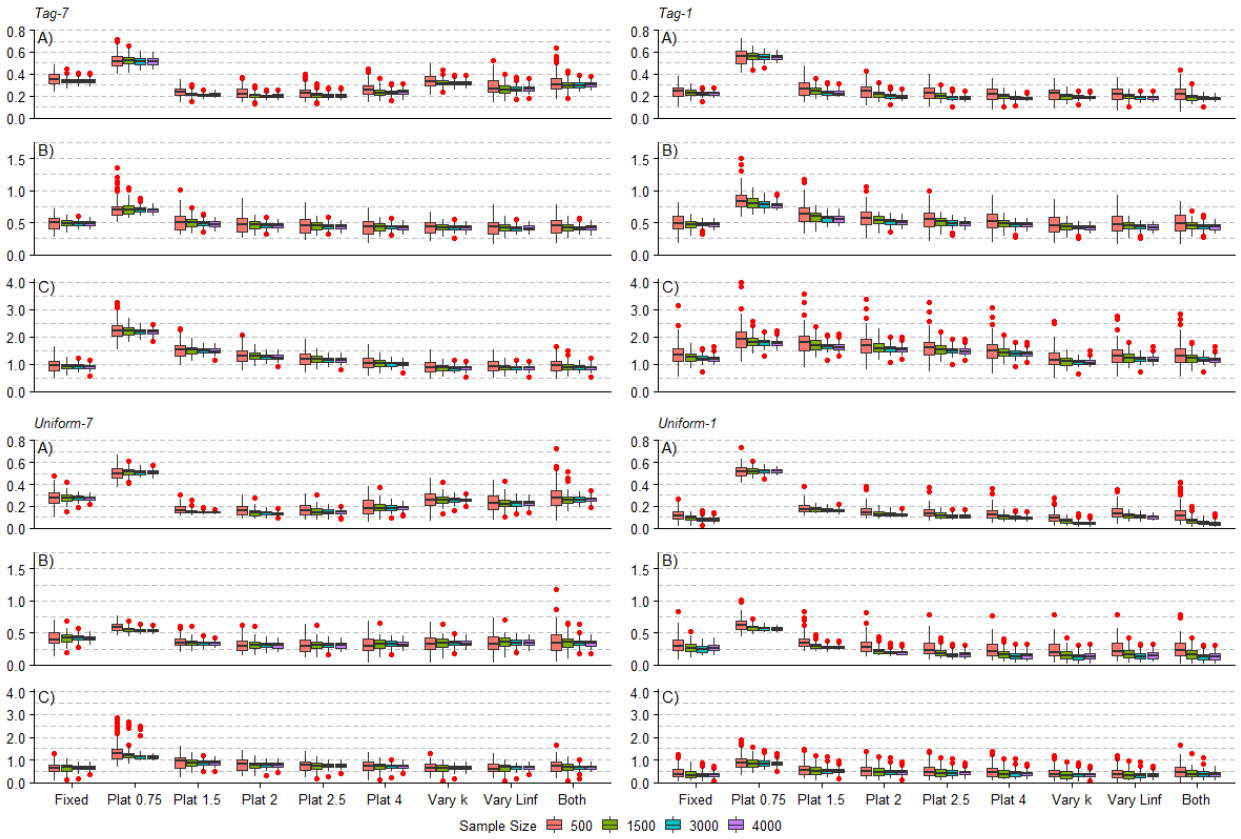

Figure A.10. As for Figure A.9, except that the operating model is OP2.

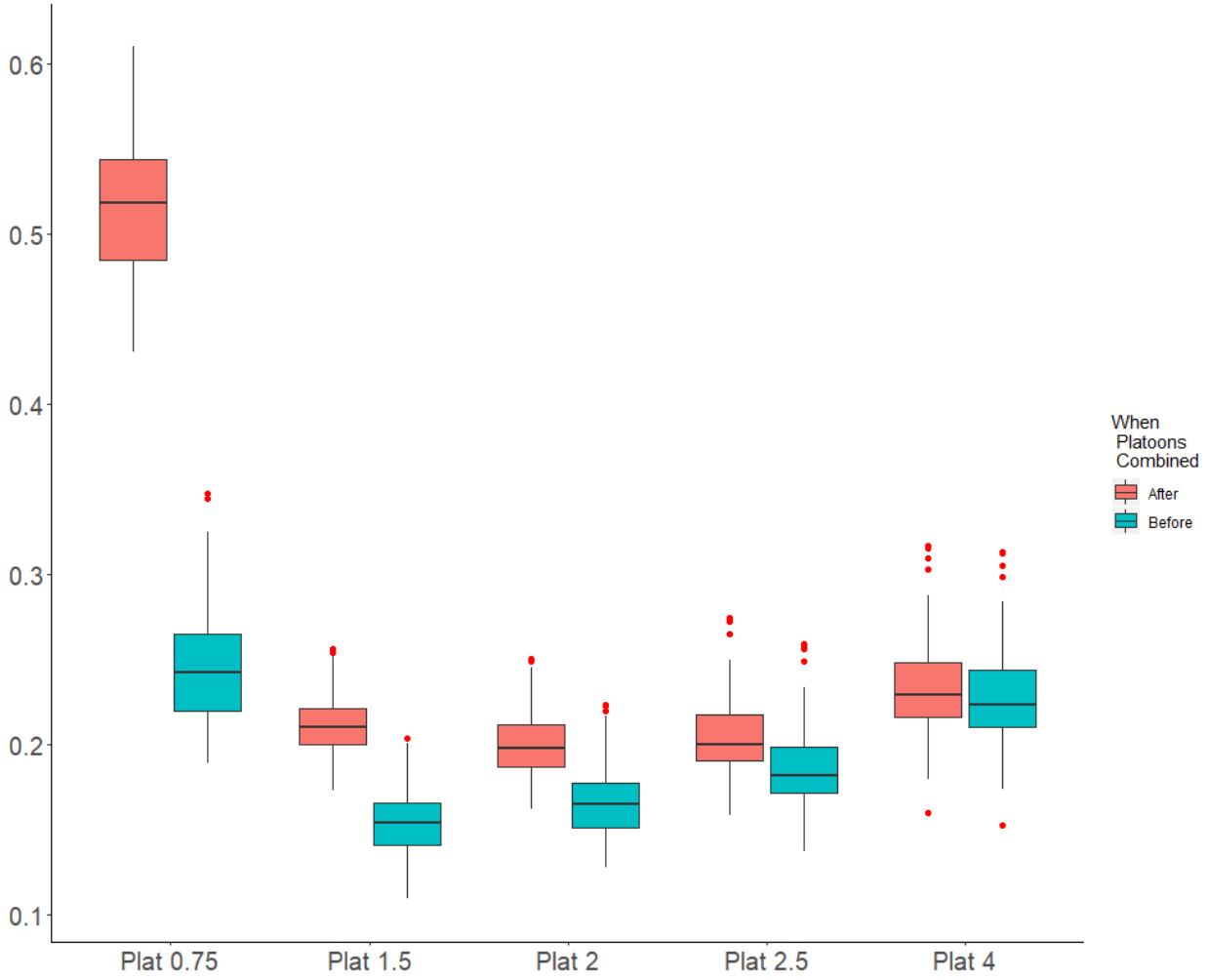

Figure A.11. Boxplots of the equilibrium diagnostic for the platoon methods. The results in this figure are for the scenario in which the initial sizes are selected from the actual tag-recapture data, the maximum time-at-liberty is 7 years, the sample size is 3,000, the number of size-classes is 18 and individuals follow their own growth curve in the operating model (OP2). The red boxes are for when each platoon's equilibrium size-structure is calculated then combined. The blue boxes are for when each platoon's size-transition matrix is combined, and the equilibrium size-structure is calculated.

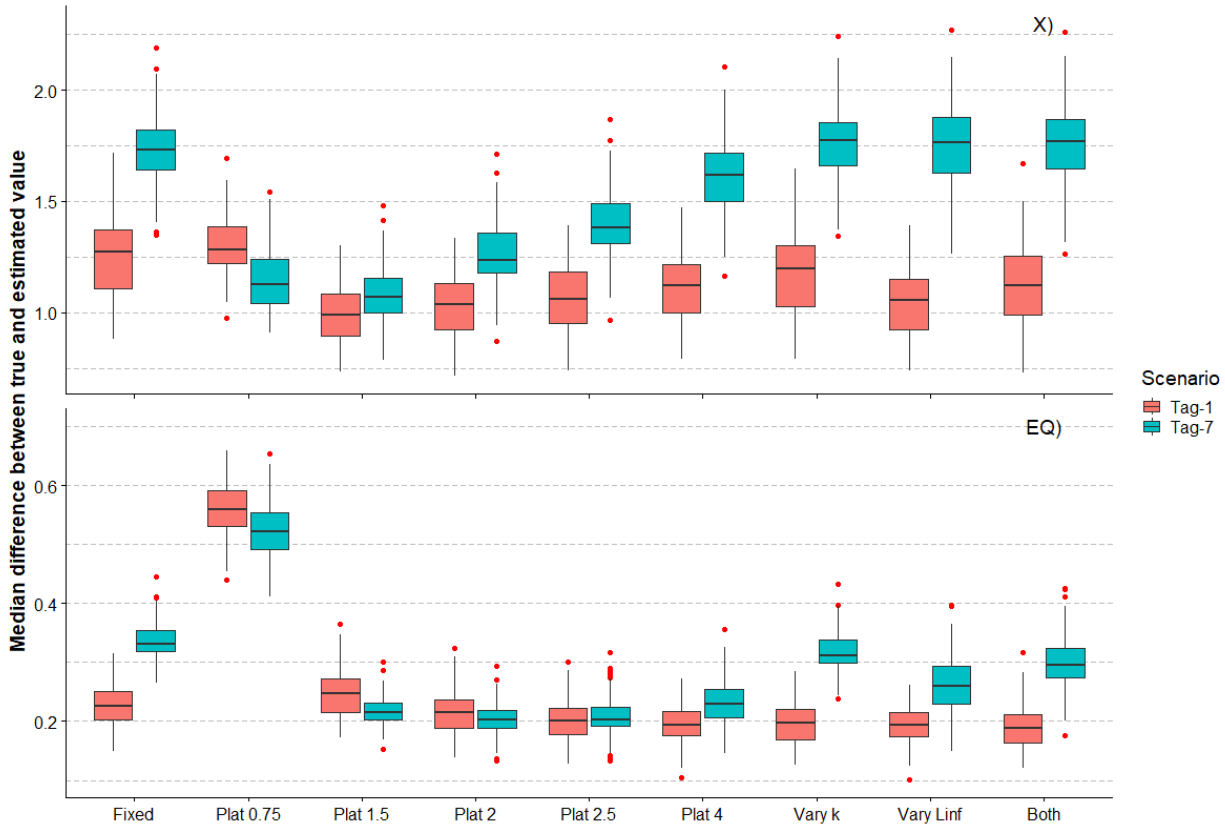

Figure A.12. Boxplots of the matrix (X) and equilibrium (EQ) diagnostics for Tag-7 and Tag-1 when the operating model has individual variation in growth (OP2). The colors indicate how the initial size was generated (Tag – actual tag-recapture data) and the maximum time-at-liberty in years (1 or 7). All estimation methods have 18 size-classes.

Appendix B

Appendix B1: *Likelihood function options for the estimation methods*

The log likelihoods for the catch, discard and index data all follow lognormal distributions. The log-likelihood (ignoring constants independent of the model parameters) for the catch data is:

$$\sum_y \sum_f \left(\frac{log\left(\tilde{C}_{y,f}^{R,obs}\right) - log\left(\tilde{C}_{y,f}^{R,pred}\right)}{\sigma_C} \right)^2 \quad \text{(B1.1)}$$

where $\tilde{C}_{y,f}^{D,obs}$ is the observed retained catch for fleet *f* during year *y*, $\tilde{C}_{y,f}^{R,pred}$ is the predicted retained catch for year *y* and fleet *f*, and σ_C is the standard deviation for the predicted catch. The log-likelihood function for discard data is:

$$\sum_y \sum_f log(\sigma_D) + \frac{\left(log\left(\tilde{C}_{y,f}^{D,obs}\right) - log\left(\tilde{C}_{y,f}^{D,pred}\right)\right)^2}{\sigma_D^2} \quad \text{(B1.2)}$$

where $\tilde{C}_{y,f}^{D,pred}$ is the predicted discarded catch for year *y* and fleet *f*, and σ_D is the standard deviation for the predicted discard. The log-likelihood function for index data is:

$$\sum_y \sum_s log(\sigma_I) + \frac{\left(log\left(N_{y,s}^{obs}\right) - log\left(N_{y,s}^{pred}\right)\right)^2}{\sigma_I^2} \quad \text{(B1.3)}$$

where $N_{y,s}^{pred}$ is the predicted survey index for year *y* and sex *s*, and σ_I is the standard deviation for the predicted survey index.

The loglikelihood for the tagging data is based on the size-transition matrix:

$$\sum_{ntag} \ln\left(S_{f,y,i(rec)} [\boldsymbol{X}^{t_{tag}}]_{s,j(rel),i(rec)}\right) \quad \text{(B1.4)}$$

where *ntag* is the number of tagged individuals, $S_{f,y,i(rec)}$ is the selectivity for a tagged individual recaptured in size-class *i* during year *y* by fleet *f* (only tagged individuals captured by the directed fleet were included) and $[\boldsymbol{X}^{t_{tag}}]_{s,j(rel),i(rec)}$ is the size-transition matrix value for a tagged individual of sex *s*, first tagged and released t_{tag} years ago in size-class *j* and recaptured in size-class *i*.

There are three likelihood function options for the size-composition data. The first option is the multinomial distribution:

$$\sum_y N_y \sum_s \sum_i P_{y,s,i} \ln\left(\frac{\hat{P}_{y,s,i}}{P_{y,s,i}}\right) \quad \text{(B1.5)}$$

where N_y is the effective sample size for year *y* (which is fixed to the actual sample size), $P_{y,s,i}$ is the proportion of the catch that is of sex *s* and size-class *i* during year *y* and $\hat{P}_{y,s,i}$ is the estimated proportion of the catch that is of sex *s* and size class *i* during year *y*:

$$\hat{P}_{y,s,i} = \frac{C_{y,s,i}}{\sum_s \sum_l C_{y,s,i}} \tag{B1.6}$$

where $C_{y,s,i}$ is the model estimate of the catch of animals of sex *s* in size-class *i* during year *y*.

The second option is the Dirichlet-multinomial distribution:

$$\sum_y \left[\begin{array}{c} ln\Gamma(\sigma_1^2) - \ ln\Gamma(\sigma_1^2 + N) - ln\Gamma(N) + \ ln\Gamma(N+N) + \\ \sum_s \sum_i \left\{ ln\Gamma\left(\sigma_1^2 \hat{P}_{y,s,i} + N P_{y,s,i}\right) - \ ln\Gamma\left(\sigma_1^2 \hat{P}_{y,s,i}\right) - ln\Gamma\left(N\left(\hat{P}_{y,s,i} + P_{y,s,i}\right)\right) + ln\Gamma\left(N\hat{P}_{y,s,i}\right) \right\} \end{array} \right] \tag{A.7}$$

where σ_1^2 is the effective sample size and *N* is the sample size. Care needs to be taken when there are zeros in the data. To address this problem, we replaced zeros with a very small number (1.0 x e^{-10}). The third is the multivariate normal distribution:

$$-\frac{1}{2}\sum_y \sum_s ln\left[\left|\boldsymbol{\Sigma}_{y,s}\right|\right] - \frac{1}{2}\left(N(\boldsymbol{P}_{y,s} - \ \widehat{\boldsymbol{P}}_{y,s})\right)^T \boldsymbol{\Sigma}_{y,s}^{-1}\left(N(\boldsymbol{P}_{y,s} - \ \widehat{\boldsymbol{P}}_{y,s})\right) \tag{B1.8}$$

where $\boldsymbol{P}_{y,s}$ is a vector of the proportions of catch by size-class for year *y* and sex *s*, $\widehat{\boldsymbol{P}}_{y,s}$ is a vector of the estimated proportions of catch by size-class for year *y* and sex *s* and $\boldsymbol{\Sigma}_{y,s}$ is the variance covariance matrix for year *y* and sex *s*. The variance-covariance matrix is given by:

$$\boldsymbol{\Sigma}_{y,s,i,j} = \begin{cases} \sigma_2^2 \hat{P}_{y,s,i}\left(1 - \ \hat{P}_{y,s,i}\right) & if\ i = j \\ -\sigma_2^2 \hat{P}_{y,s,i} \hat{P}_{y,s,j} & if\ i \neq j \end{cases} \tag{B1.9}$$

where σ_2^2 is the effective sample size. Equation B1.9 allows the multivariate normal likelihood to approximate a multinomial distribution. The variance-covariance matrix for the multivariate normal is degenerate. This means that one of the proportions of catches by size-class depends on all the others. Therefore, the number of rows and columns in the variance covariance matrix needs to be reduced by one. In addition, the multivariate normal is sensitive to zeros since the determinant of the variance covariance matrix needs to be calculated when determining the likelihood. To solve this problem, remove zeros in the data when calculating the likelihood.

Appendix B2: *Bias in the estimation methods*

Figures such as 3.4 and 3.7 show that the estimates of spawning stock biomass are positively biased at the end of the time series regardless of how selectivity is modeled, and the likelihood function used for the size-composition data. This positive bias also causes the average MRE values for SSB_{2016}/SSB_{1968} to always be positive. The bias is due to a lack of information on the cohorts in the population at the end of the time-series. The size of recruitment is first detected in the size-composition data from the survey fleet, which captures smaller individuals since its selectivity curve is shifted to the left of the other fleets (Figure 3.1a). Increasing the effective sample size for the survey fleet reduces and eventually removes the positive bias (Figure B.3). In contrast, increasing the effective sample size for the other fleets does not eliminate the positive bias. A larger effective sample size for the survey improves the accuracy and precision of the survey fleet's selectivity curve (Figure B.4). This improves estimation of recruitment and hence spawning stock biomass. Other fleets detect recruitment several years after the survey since it takes time for individuals to grow large enough to be caught. An imprecise approximation of the survey selectivity reduced the accuracy of the estimated recruitment for the years towards the end of the time series since no additional information is provided by the other fleets, and hence the bias at the end of the time series. Improved estimates of survey selectivity benefit estimates of spawning stock biomass for all years because this improves the accuracy of the distribution of recruits added to the population over the entire timeline.

Table B.1. The male size-transition matrix used in all operating models. The rows are the initial size, the columns are the ending size.

0.14	0.14	0.18	0.19	0.16	0.11	0.06	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.17	0.13	0.17	0.17	0.15	0.10	0.06	0.03	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.20	0.13	0.15	0.16	0.13	0.10	0.06	0.03	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.23	0.12	0.14	0.14	0.12	0.10	0.07	0.04	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.26	0.12	0.13	0.13	0.12	0.09	0.07	0.04	0.02	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.28	0.11	0.12	0.12	0.11	0.09	0.07	0.04	0.03	0.02	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.30	0.11	0.11	0.11	0.10	0.09	0.07	0.05	0.03	0.02	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.32	0.10	0.11	0.11	0.10	0.08	0.06	0.05	0.03	0.02	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.33	0.10	0.10	0.10	0.09	0.08	0.06	0.05	0.03	0.02	0.01	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.35	0.09	0.10	0.09	0.09	0.08	0.06	0.05	0.04	0.02	0.02	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.36	0.09	0.09	0.09	0.08	0.07	0.06	0.05	0.04	0.03	0.02	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.37	0.08	0.09	0.08	0.08	0.07	0.06	0.05	0.04	0.03	0.02	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.39	0.08	0.08	0.08	0.07	0.07	0.06	0.05	0.04	0.03	0.02	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.40	0.08	0.08	0.08	0.07	0.06	0.06	0.05	0.04	0.03	0.02	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.41	0.07	0.07	0.07	0.07	0.06	0.05	0.05	0.04	0.03	0.02	0.05	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.41	0.07	0.07	0.07	0.07	0.06	0.05	0.05	0.04	0.03	0.02	0.06	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.42	0.07	0.07	0.07	0.06	0.06	0.05	0.04	0.04	0.03	0.02	0.06	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.43	0.07	0.07	0.06	0.06	0.06	0.05	0.04	0.04	0.03	0.02	0.07	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.44	0.06	0.06	0.06	0.06	0.05	0.05	0.04	0.04	0.03	0.03	0.08	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.44	0.06	0.06	0.06	0.06	0.05	0.05	0.04	0.04	0.03	0.03	0.08	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.45	0.06	0.06	0.06	0.05	0.05	0.05	0.04	0.04	0.03	0.03	0.09
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.45	0.06	0.06	0.06	0.05	0.05	0.05	0.04	0.04	0.03	0.12
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.46	0.06	0.06	0.05	0.05	0.05	0.04	0.04	0.04	0.16
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.46	0.05	0.05	0.05	0.05	0.05	0.04	0.04	0.20
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.47	0.05	0.05	0.05	0.05	0.05	0.04	0.24
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.47	0.05	0.05	0.05	0.05	0.04	0.29
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.48	0.05	0.05	0.05	0.05	0.33
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.48	0.05	0.05	0.05	0.38
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.48	0.05	0.05	0.42
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.49	0.05	0.47
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.49	0.51
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00

Table B.2. The female size-transition matrix used in all operating models. The rows are the initial size, the columns are the ending size.

0.12	0.12	0.17	0.19	0.17	0.12	0.07	0.03	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.16	0.13	0.16	0.17	0.15	0.11	0.07	0.03	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.20	0.13	0.15	0.15	0.14	0.10	0.07	0.04	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.23	0.12	0.14	0.14	0.13	0.10	0.07	0.04	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.26	0.12	0.13	0.13	0.12	0.09	0.07	0.04	0.02	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.28	0.11	0.12	0.12	0.11	0.09	0.06	0.04	0.03	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.31	0.11	0.12	0.11	0.10	0.08	0.06	0.04	0.03	0.02	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.33	0.10	0.11	0.11	0.10	0.08	0.06	0.04	0.03	0.02	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.35	0.10	0.10	0.10	0.09	0.08	0.06	0.04	0.03	0.02	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.37	0.10	0.10	0.09	0.08	0.07	0.06	0.04	0.03	0.02	0.01	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.39	0.09	0.09	0.09	0.08	0.07	0.06	0.04	0.03	0.02	0.01	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.40	0.09	0.09	0.08	0.08	0.07	0.06	0.04	0.03	0.02	0.02	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.42	0.08	0.08	0.08	0.07	0.06	0.05	0.04	0.03	0.02	0.02	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.43	0.08	0.08	0.08	0.07	0.06	0.05	0.04	0.03	0.02	0.02	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.44	0.08	0.08	0.07	0.07	0.06	0.05	0.04	0.03	0.03	0.02	0.04	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.45	0.07	0.07	0.07	0.06	0.06	0.05	0.04	0.03	0.03	0.02	0.04	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.46	0.07	0.07	0.07	0.06	0.06	0.05	0.04	0.03	0.03	0.02	0.05	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.47	0.07	0.07	0.06	0.06	0.05	0.05	0.04	0.03	0.03	0.02	0.05	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.48	0.07	0.06	0.06	0.06	0.05	0.05	0.04	0.03	0.03	0.02	0.05	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.49	0.06	0.06	0.06	0.06	0.05	0.04	0.04	0.03	0.03	0.02	0.06	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.50	0.06	0.06	0.06	0.05	0.05	0.04	0.04	0.03	0.03	0.02	0.06
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.50	0.06	0.06	0.06	0.05	0.05	0.04	0.04	0.03	0.03	0.09
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.51	0.06	0.06	0.05	0.05	0.05	0.04	0.04	0.03	0.12
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.52	0.06	0.05	0.05	0.05	0.04	0.04	0.04	0.15
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.52	0.05	0.05	0.05	0.05	0.04	0.04	0.19
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.53	0.05	0.05	0.05	0.05	0.04	0.23
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.53	0.05	0.05	0.05	0.04	0.28
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.54	0.05	0.05	0.05	0.32
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.54	0.05	0.05	0.36
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.55	0.05	0.41
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.55	0.45
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00

Table B.3. The percent difference from the lowest median absolute relative error (MARE) for each management quantity for the simulation study exploring time-varying selectivity. A zero means the associated estimation method had the lowest value for that particular metric for the time-invariant operating model.

Run		MARE		
Operating Model	**Estimation Method**	SSB_0	SSB_{2016}/SSB_0	SSB_{2016}/SSB_{1968}
Time-Invariant	EM_{invar}	0.00	2.18	0.00
	EM_{block}	0.72	0.00	1.73
	EM_{RP}	9.26	1.54	9.84
Low Time-Varying	EM_{invar}	27.05	23.92	-12.73
	EM_{block}	26.33	28.25	-16.79
	EM_{RP}	30.01	24.17	-10.01
Medium Time-Varying	EM_{invar}	19.30	4.03	28.38
	EM_{block}	18.48	3.03	11.16
	EM_{RP}	13.79	4.99	15.95
High Time-Varying	EM_{invar}	30.89	44.16	38.35
	EM_{block}	26.26	24.67	-13.83
	EM_{RP}	30.53	26.77	-1.48
Very High Time-Varying	EM_{invar}	28.17	42.35	68.40
	EM_{block}	33.91	33.09	-13.46
	EM_{RP}	38.58	32.90	-4.00

Table B.4. The mean, median and standard deviation (SD), expressed as percentages, of Mohn's ρ for spawning stock biomass (SSB) and recruitment for the simulation study exploring how best to model time-varying selectivity.

Run		Mohn's ρ value					
Operating Model	Estimation Method	SSB			Recruitment		
		Mean	Median	SD	Mean	Median	SD
Time-Invariant	EM_{invar}	2.43	2.48	5.24	8.33	7.72	17.92
	EM_{block}	2.32	2.51	5.11	8.51	7.83	18.10
	EM_{RP}	2.16	3.00	5.21	8.95	7.64	18.10
Low Time-Varying	EM_{invar}	3.61	3.71	5.99	10.87	9.33	17.97
	EM_{block}	3.39	3.19	5.89	10.99	8.93	17.93
	EM_{RP}	3.15	3.31	5.84	11.44	9.97	18.09
Medium Time-Varying	EM_{invar}	2.82	2.86	6.28	10.87	9.75	17.73
	EM_{block}	2.34	2.04	5.88	10.90	9.41	17.62
	EM_{RP}	2.15	1.77	5.76	11.41	9.83	17.91
High Time-Varying	EM_{invar}	3.37	3.11	7.35	10.29	8.69	19.17
	EM_{block}	2.59	2.23	6.78	9.95	7.20	19.27
	EM_{RP}	2.37	2.32	6.52	10.43	7.81	19.38
Very High Time-Varying	EM_{invar}	4.44	3.55	8.04	10.46	7.64	18.56
	EM_{block}	3.23	2.74	6.80	10.11	7.81	18.82
	EM_{RP}	2.90	1.93	6.31	10.55	7.92	19.65

Table B.5. The percent difference times 100 from the lowest median absolute relative error (MARE) for a particular operating model and management quantity for the simulation study exploring likelihood functions. A zero means the associated estimation method had the lowest value for that particular quantity and operating model.

Run		MARE		
Operating Model	**Estimation Method**	SSB_0	SSB_{2016}/SSB_0	SSB_{2016}/SSB_{1968}
I-M	I-M	0	0.62	2.57
	I-D	3.78	0.96	0
	I-MN	7.09	23.21	7.48
	V-M	9.26	0	12.66
	V-D	10.87	0.37	16.79
	V-MN	4.14	17.63	8.30
I-D	I-M	0	18.01	8.35
	I-D	14.98	2.21	0
	I-MN	11.98	24.23	18.52
	V-M	6.28	10.86	16.35
	V-D	12.59	0	13.49
	V-MN	12.10	19.02	17.46
I-MN	I-M	2.37	0	0
	I-D	8.96	2.54	7.74
	I-MN	0.03	44.19	2.60
	V-M	0.78	5.30	11.17
	V-D	9.59	4.12	17.28
	V-MN	0	44.69	33.80
V-M	I-M	4.85	2.17	10.72
	I-D	3.98	0.09	7.16
	I-MN	1.27	22.87	13.55
	V-M	0	3.11	0
	V-D	2.51	0	3.93
	V-MN	3.32	31.82	4.71
V-D	I-M	0.33	18.84	16.32
	I-D	9.59	1.46	0
	I-MN	0	9.91	16.69
	V-M	12.26	15.60	14.10
	V-D	10.80	8.24	11.19
	V-MN	13.32	0	17.74
V-MN	I-M	24.11	19.66	10.90
	I-D	37.05	13.13	6.87
	I-MN	0	40.86	24.32
	V-M	36.03	0	0
	V-D	38.19	7.25	1.99
	V-MN	4.39	25.32	9.31

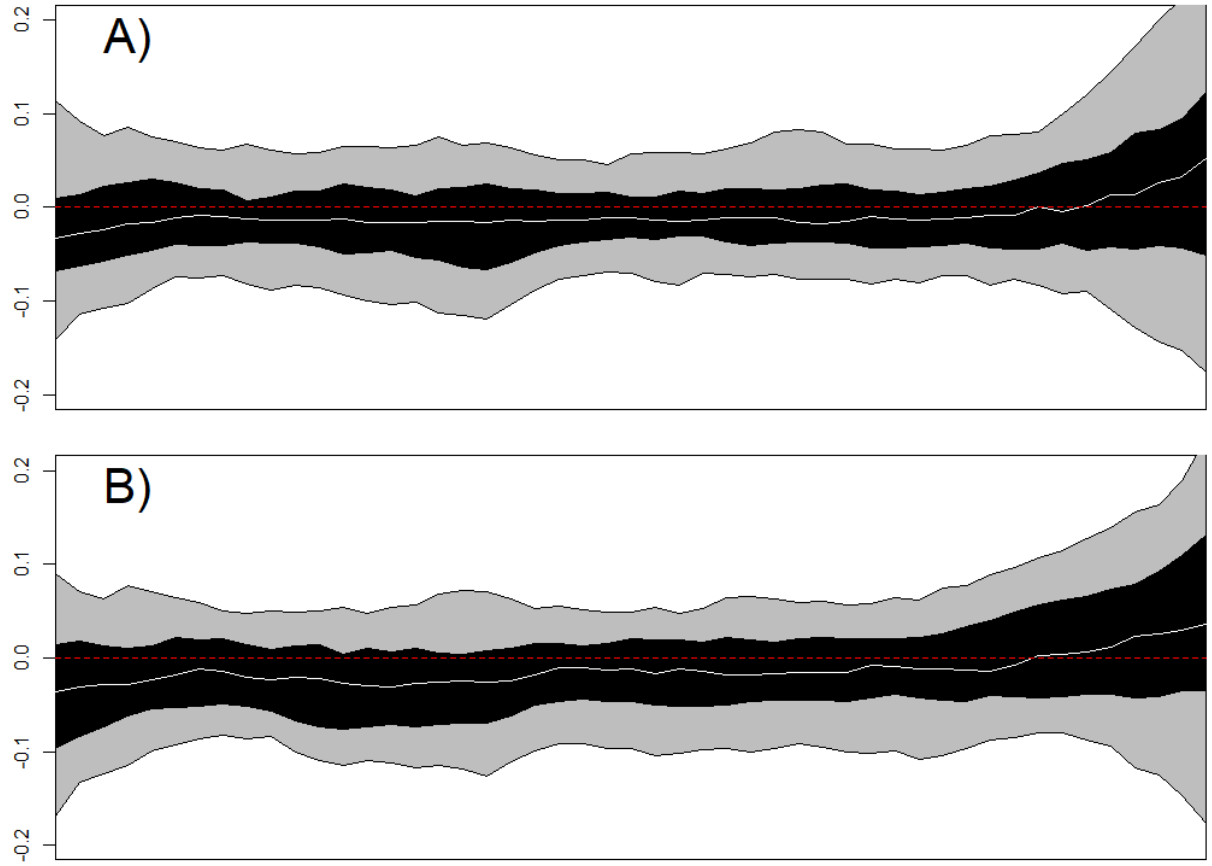

Figure B.1. The relative error in spawning stock biomass (SSB) from 100 simulations in which the estimation method uses the random parameters (EM_{RP}) approach. The two graphs differ by operating model. Figure A) uses a random walk while B) a random parameter process to generate the data. Both operating models have very high time-varying selectivity. The white line in panels A and B is the median relative error, the red dash line is the zero line, the black shaded region is the 50% quantile and the grey shaded region is the 90% quantile.

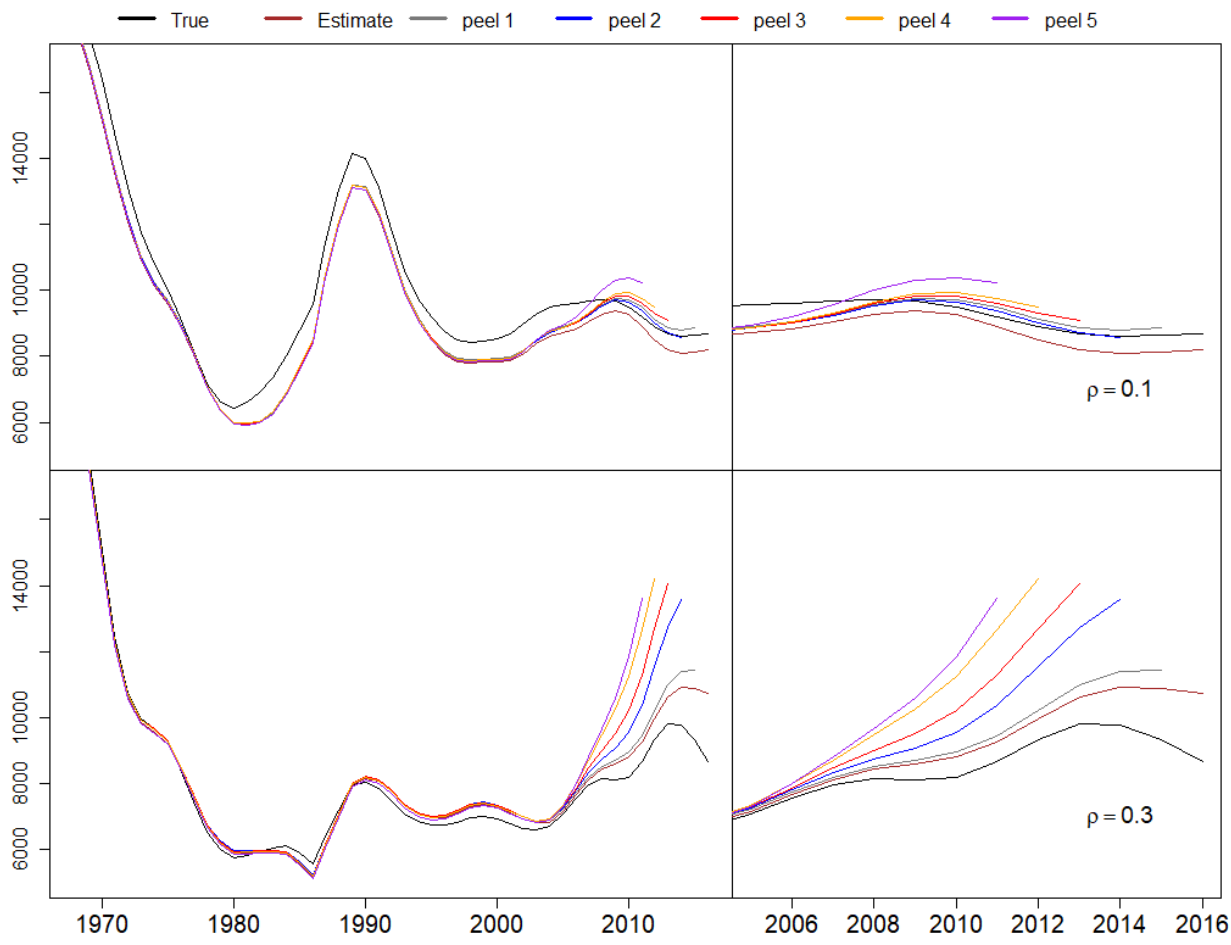

Figure B.2. An example of the retrospective patterns for spawning stock biomass from two simulations with the same operating model (time-invariant selectivity and multinomial generating function) and estimation method (time-invariant selectivity and multivariate normal likelihood). The first column shows the full timeline. The second column shows the last 10 years. The ρ's are the Mohn's ρ value for that simulation.

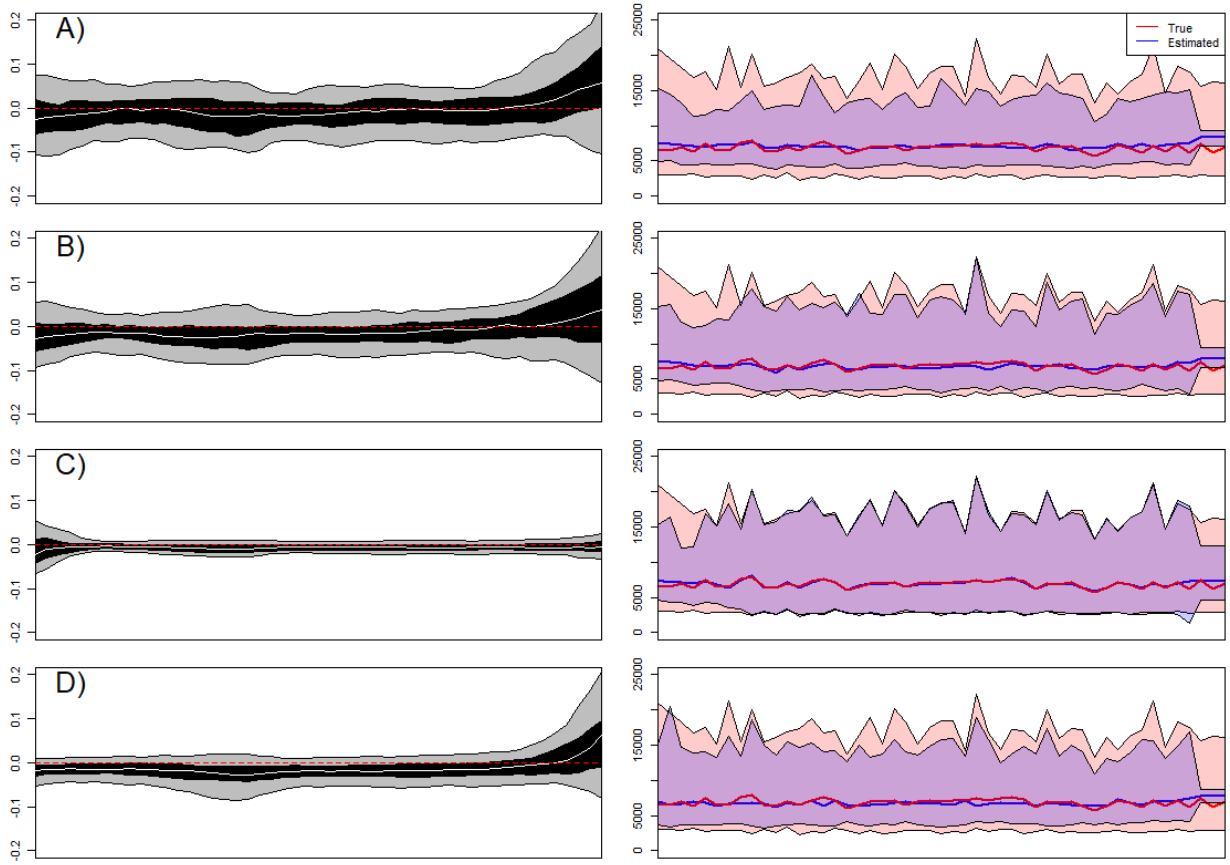

Figure B.3. The relative error in spawning stock biomass (SSB) and recruitment from 100 simulations in which the operating model and estimation method assume time-invariant selectivity. Each row represents a different scenario in which the effective sample size for the size-composition data for either the directed fleet or survey is altered. Row A is the same as the OP I and EM_{invar} scenario in Figure 3.4. The effective sample size for the survey in rows B and C is increased to 1,000 and 100,000 respectively. The effective sample size for the directed fleet in Row D is increased to 100,000. The white line in the left panels is the median relative error, the red dash line is the zero line, the black shaded region is the 50% quantile and the grey shaded region is the 90% quantile. The light red shaded region in the right panels is the range of true recruitments from all simulations with the dark red line representing true median recruitment. The blue shaded region is the range of estimated recruitments from all simulations with the blue line representing estimated median recruitment.

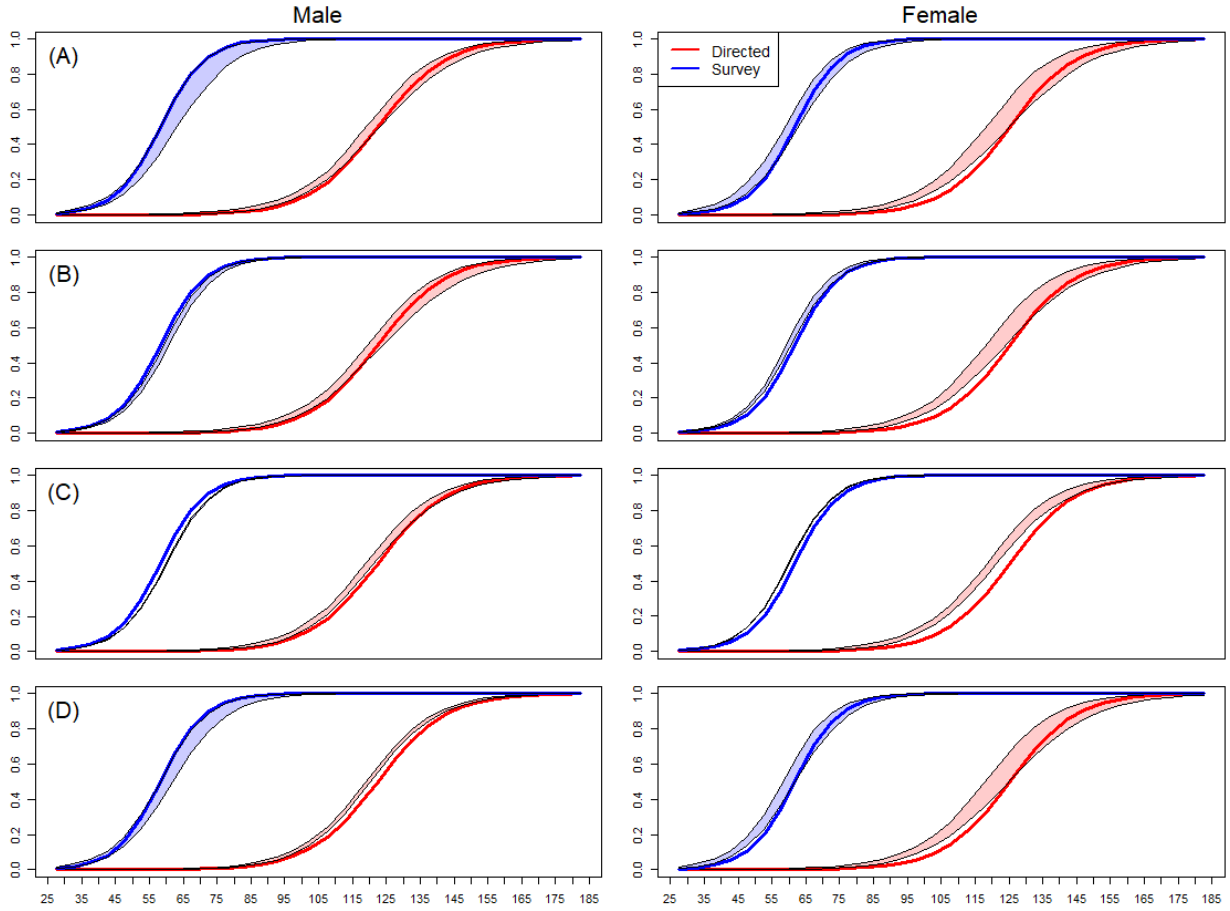

Figure B.4. The selectivity curves for the directed fishery and the survey from 100 simulations in which the operating model and estimation method both assume time-invariant selectivity. Each row represents a different scenario, with the order mirroring the order in Figure B.3. The light red shaded region is the 95% quantile of the estimated selectivity curves for the directed fleet for all simulations with the dark red line representing true selectivity curve for the directed fleet. The light blue shaded region is the 95% quantile of the estimated selectivity curves for the survey for all simulations with the dark blue line representing true selectivity curve for the survey fleet. Male and female selectivity are shown in the left and right panels, respectively.

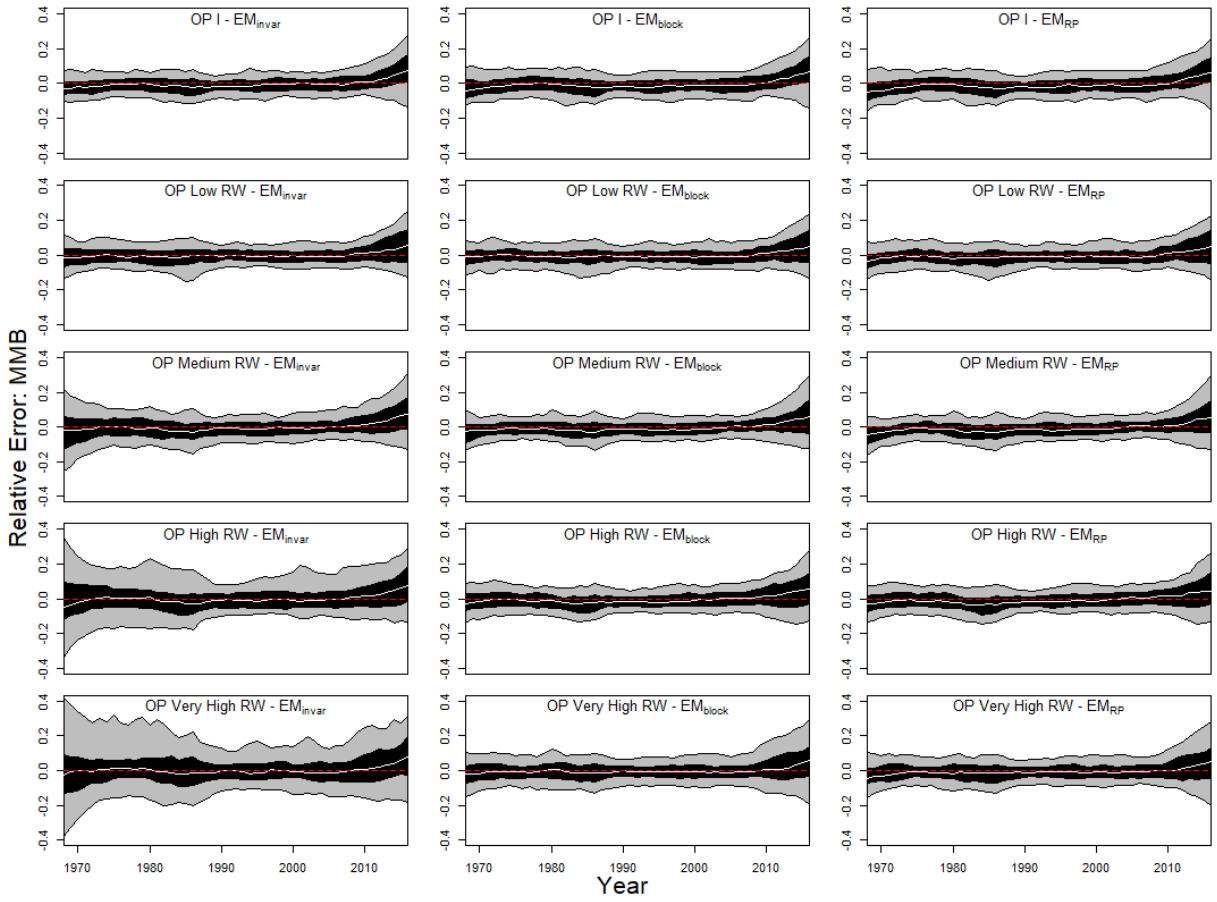

Figure B.5. Relative error for mature male biomass (MMB) for the simulation study exploring how best to model time-varying selectivity. The five operating models (OP) are: time-invariant (I), low random walk (Low RW), medium random walk (Medium RW), high random walk (High RW) and very high random walk (Very High RW). The three estimation methods are: time-invariant (EM_{invar}), blocking (EM_{block}) and random parameters (EM_{RP}). The white line is the median relative error, the red dash line is the zero line, the black shaded region is the 50% quantile and the grey shaded region is the 90% quantile.

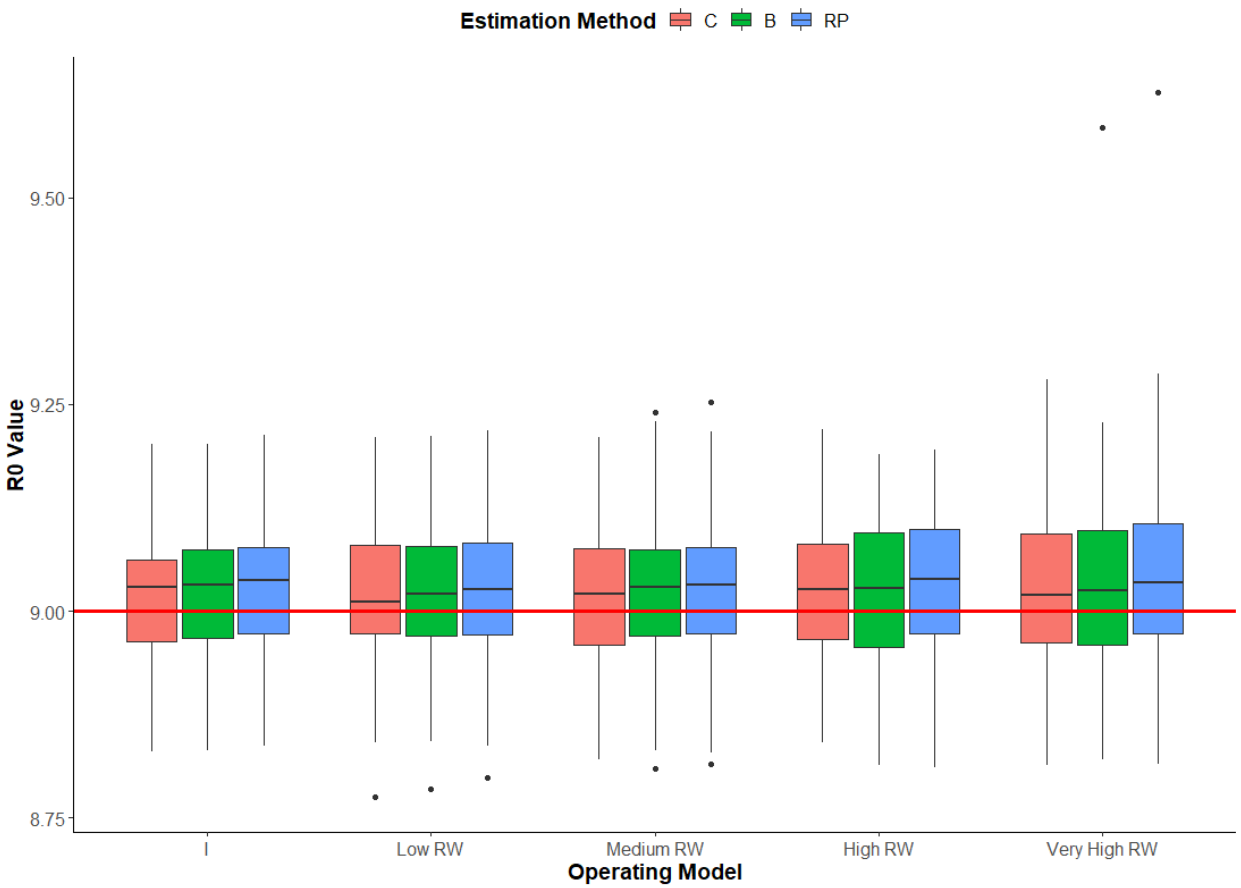

Figure B.6. Boxplots of the estimates of R_0 from the study exploring how best to model time-varying selectivity. The colors represent the estimation methods (EM_{invar} – time-invariant, EM_{block} – blocking, EM_{RP} – random parameters). The x-axis indicates operating model (I – time-invariant, Low RW – low random walk, Medium RW – medium random walk, High RW – high random walk, Very High RW – very high random walk). The red solid line is the true R_0 value.

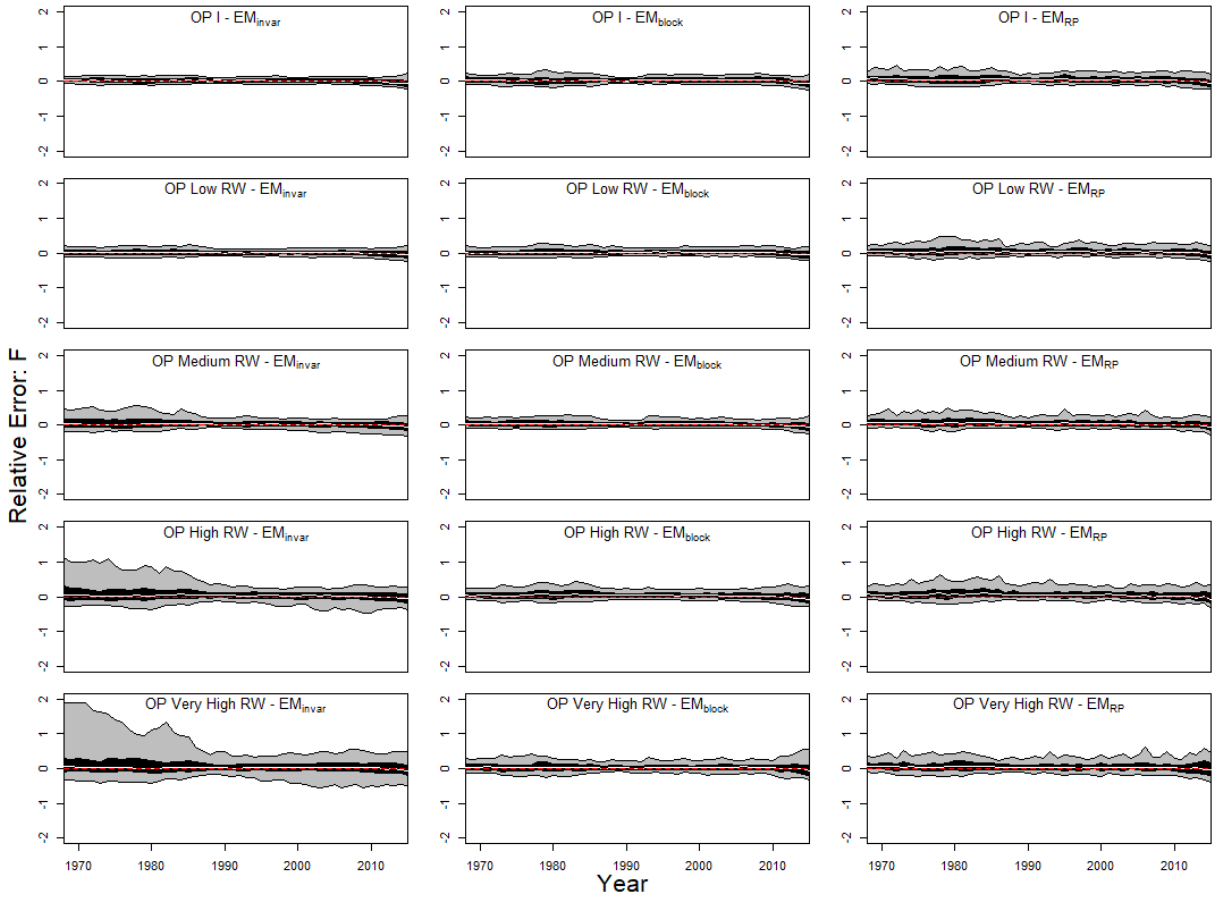

Figure B.7. Relative error for fully selected fishing mortality (F) for the simulation study exploring how best to model time-varying selectivity. The five operating models (OP) are: time-invariant (I), low random walk (Low RW), medium random walk (Medium RW), high random walk (High RW) and very high random walk (Very High RW). The three estimation methods are: time-invariant (EM_{invar}), blocking (EM_{block}) and random parameters (EM_{RP}). The white line is the median relative error, the red dash line is the zero line, the black shaded region is the 50% quantile and the grey shaded region is the 90% quantile.

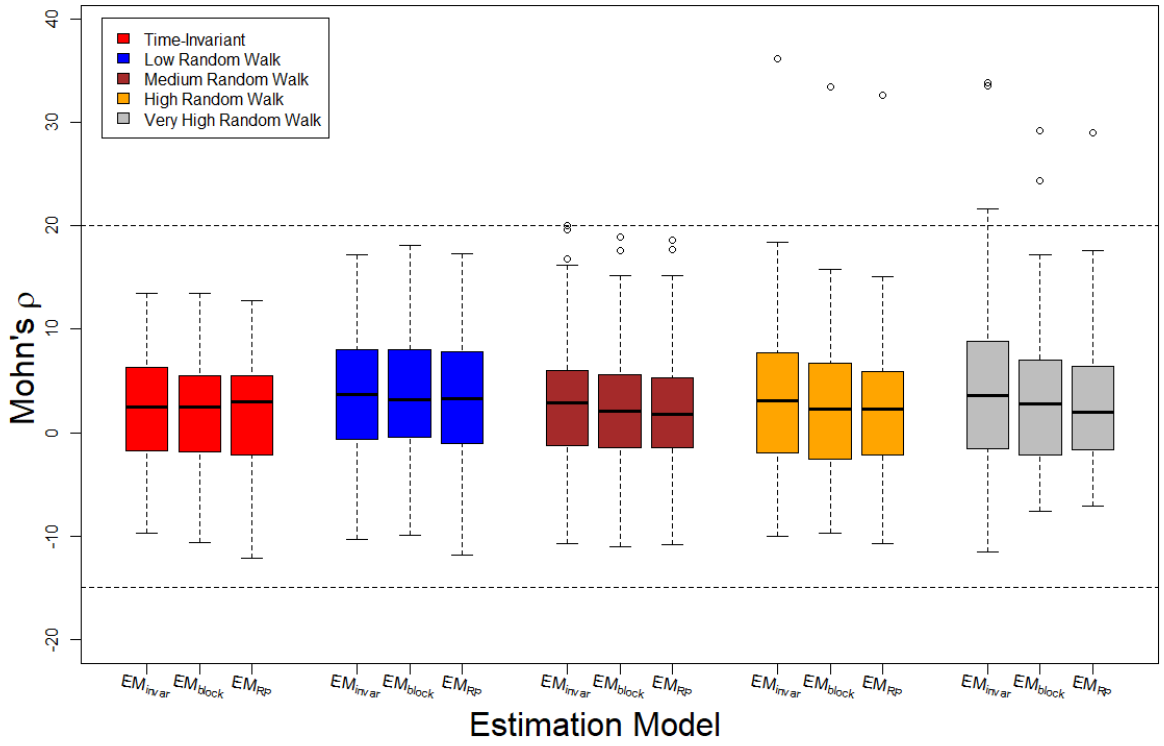

Figure B.8. Boxplots of Mohn's ρ for spawning stock biomass times 100 from the study exploring how best to model time-varying selectivity. The colors represent the operating model scenario. The x-axis indicates the estimation method (EM_{invar} – time-invariant, EM_{block} – blocking, EM_{RP} – random parameters). The dashed lines are the bounds for Mohn's ρ that do not suggest retrospective patterns.

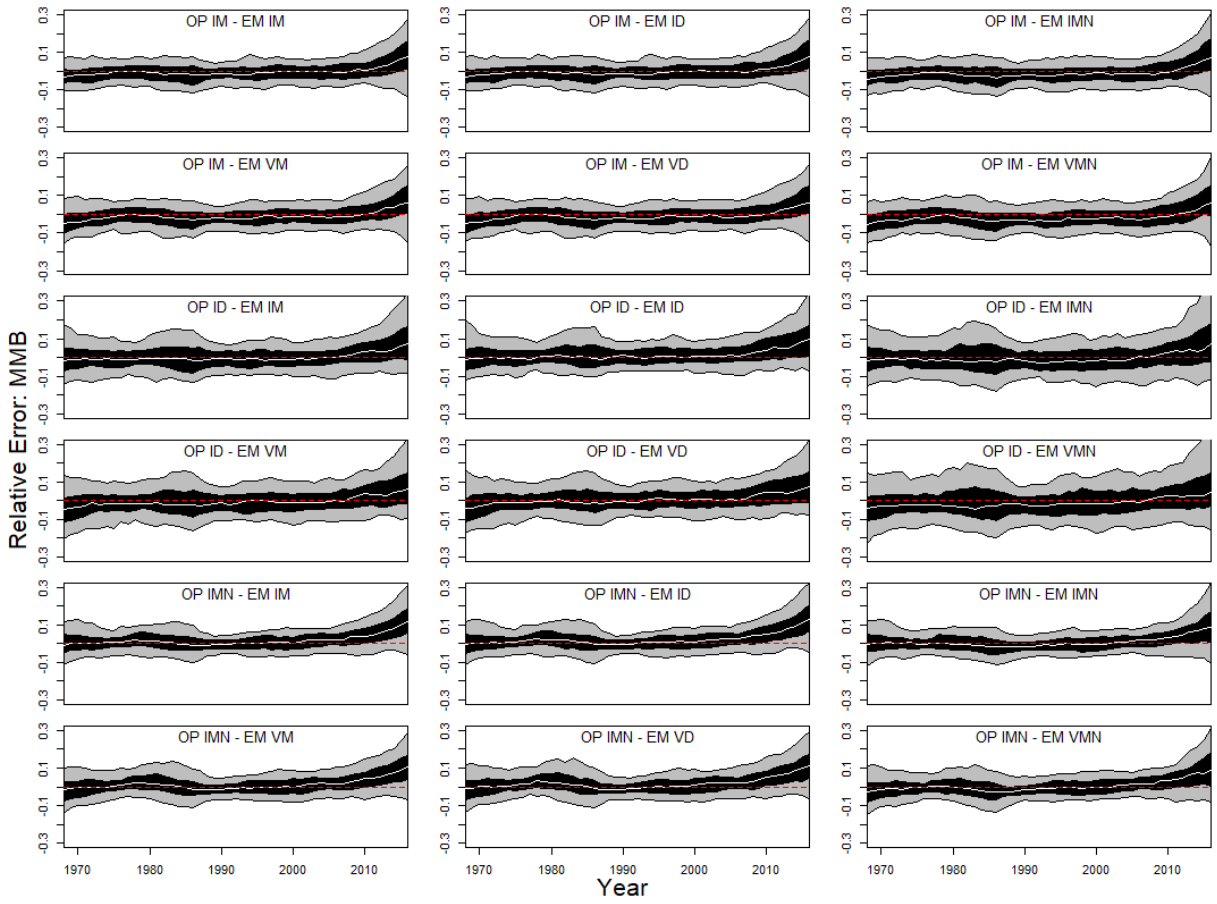

Figure B.9. Relative errors for mature male biomass (MMB) for the simulation study examining likelihood functions in relation to time-varying selectivity when the operating model (OP) has time-invariant (I) selectivity. The generating functions are either multinomial (M), Dirichlet-multinomial (D), or multivariate-normal (MN). The estimation methods (EM) assume either time-invariant (I) or time-varying (V) selectivity. The likelihood functions for the estimation methods are either multinomial (M), Dirichlet-multinomial (D), or multivariate-normal (MN). The white line is the median relative error, the red dash line is the zero line, the black shaded region is the 50% quantile and the grey shaded region is the 90% quantile.

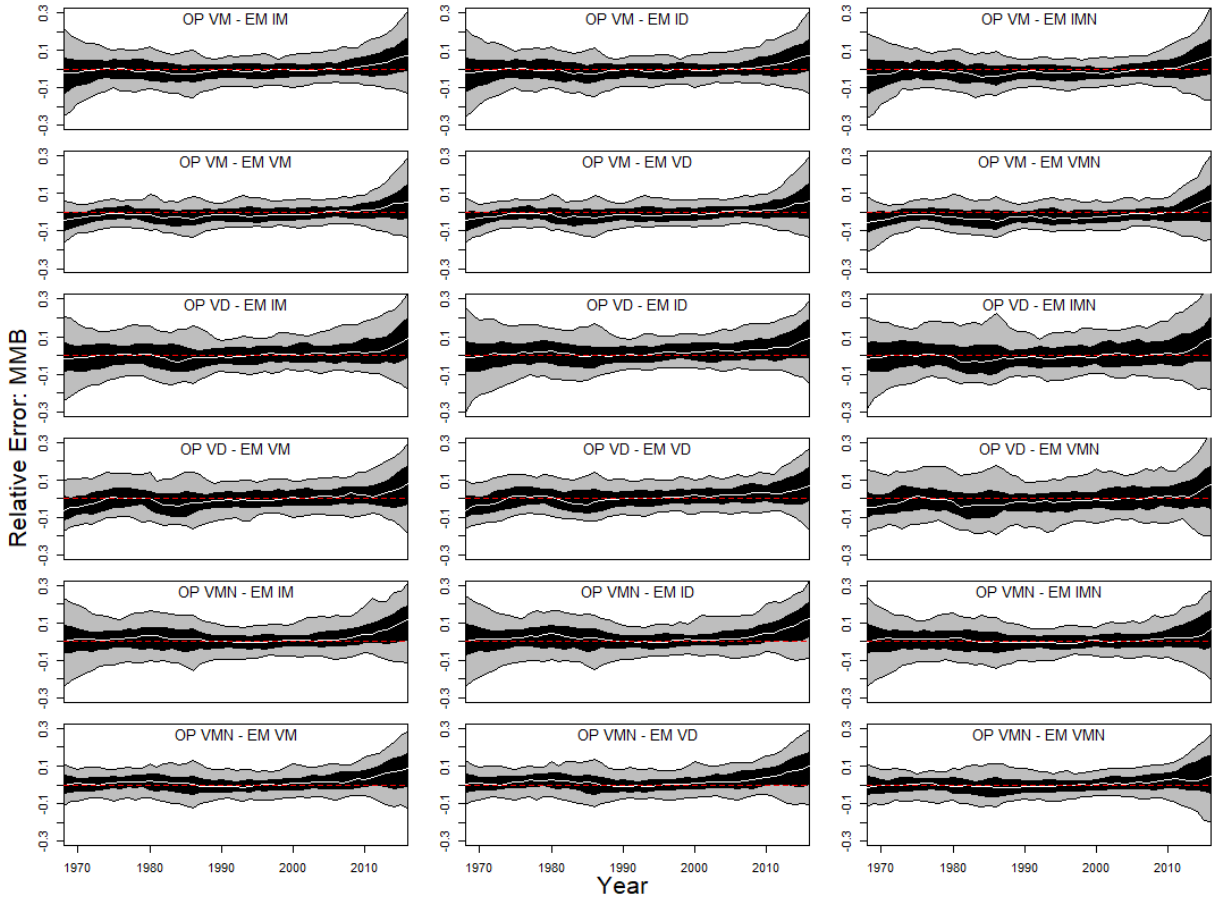

Figure B.10. Relative errors for mature male biomass (MMB) for the simulation study examining likelihood functions in relation to time-varying selectivity when the operating model (OP) has time-varying (V) selectivity. The generating functions are either multinomial (M), Dirichlet-multinomial (D), or multivariate-normal (MN). The estimation methods (EM) assume either time-invariant (I) or time-varying (V) selectivity. The likelihood functions for the estimation methods are either multinomial (M), Dirichlet-multinomial (D), or multivariate-normal (MN). The white line is the median relative error, the red dash line is the zero line, the black shaded region is the 50% quantile and the grey shaded region is the 90% quantile.

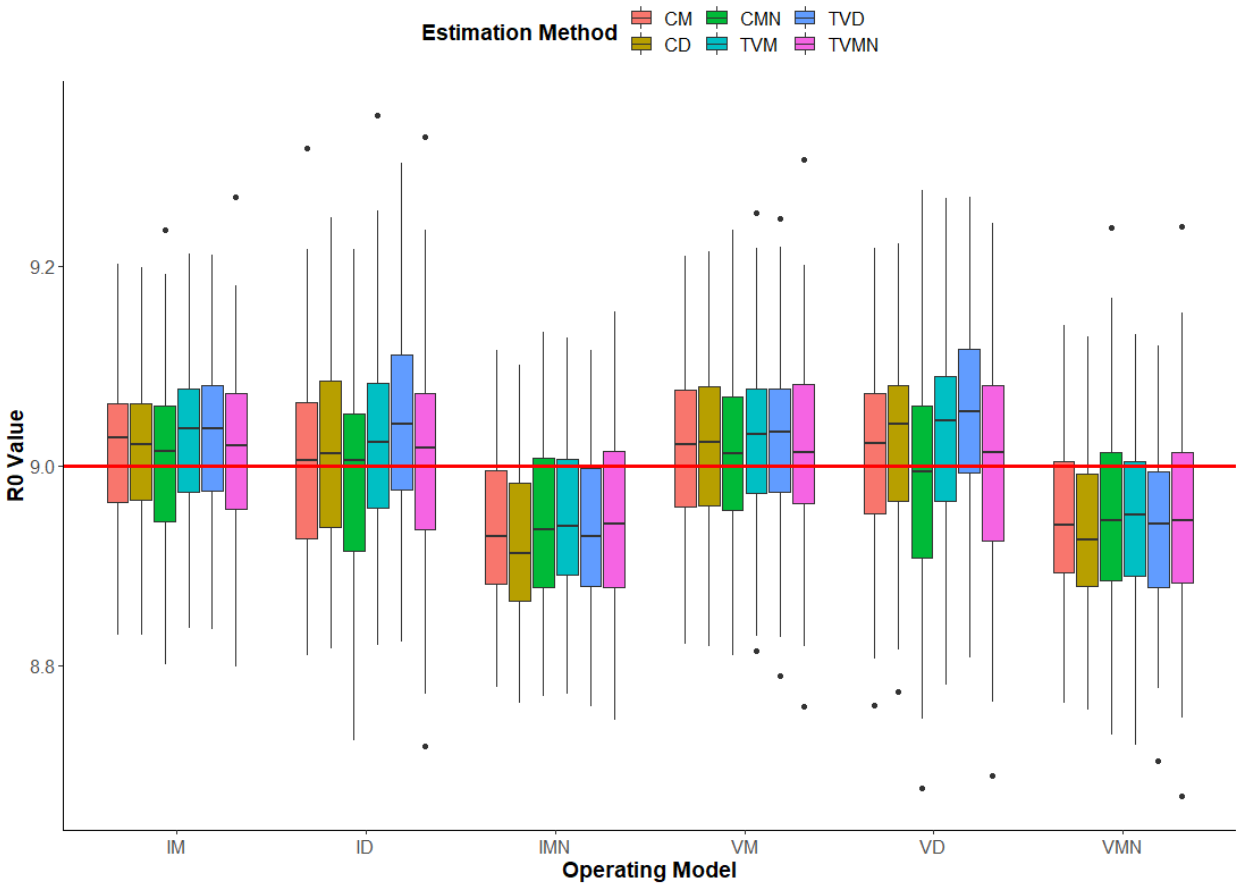

Figure B.11. Boxplots of the estimates of R_0 from the study examining likelihood functions in relation to time-varying selectivity. The colors indicate the estimation method. The x-axis specifies the operating model. For both the colors and x-axis, the first letter indicates the selectivity assumption (I – time-invariant, V – time-varying). The rest specifies the generating function/ likelihood (M – multinomial, D – Dirichlet-multinomial, MN – multivariate-normal). The red solid line is the true R_0 value.

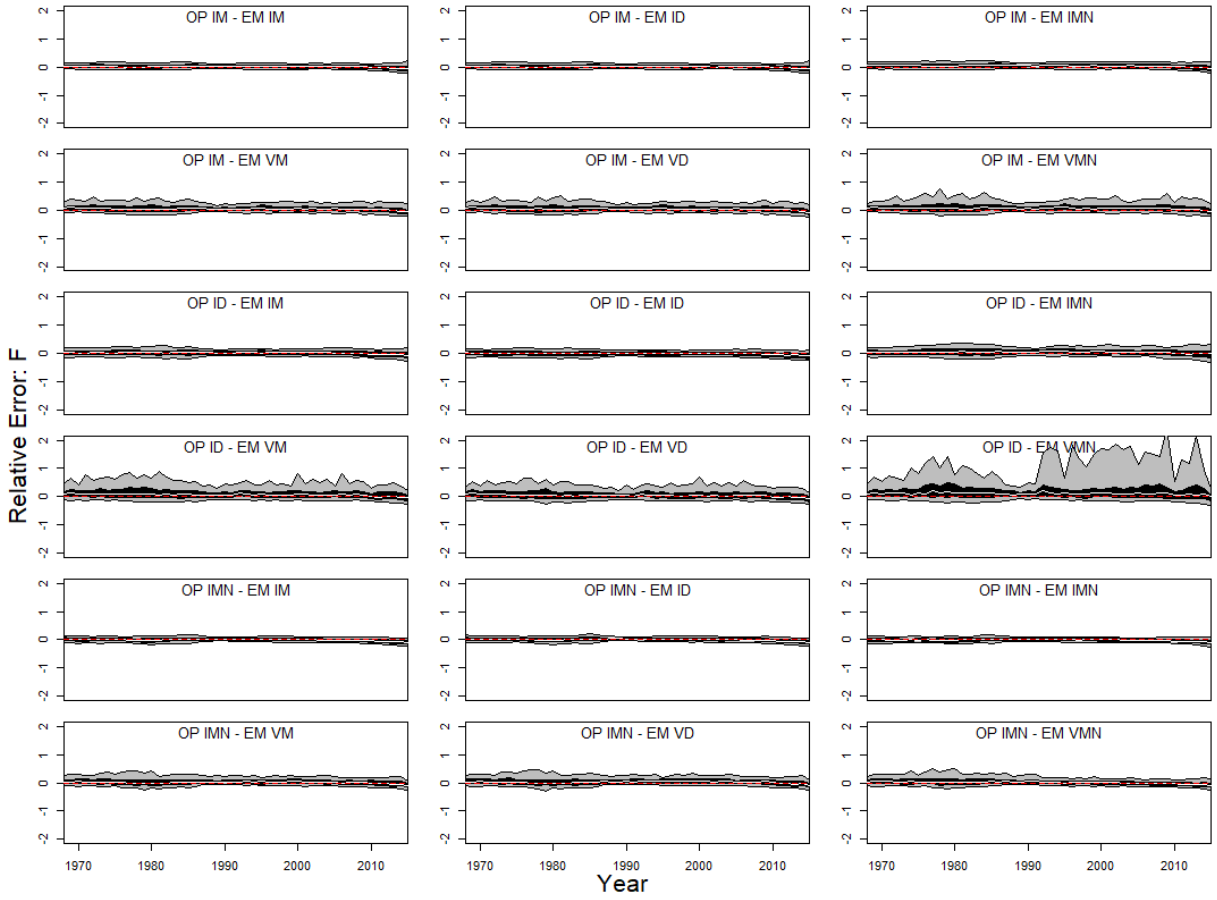

Figure B.12. Relative error for fully selected fishing mortality (F) for the simulation study examining likelihood functions in relation to time-varying selectivity when the operating model (OP) has time-invariant (I) selectivity. The generating functions are either multinomial (M), Dirichlet-multinomial (D), or multivariate-normal (MN). The estimation methods (EM) assume either time-invariant (I) or time-varying (V) selectivity. The likelihood functions for the estimation methods are either multinomial (M), Dirichlet-multinomial (D), or multivariate-normal (MN). The white line is the median relative error, the red dash line is the zero line, the black shaded region is the 50% quantile and the grey shaded region is the 90% quantile.

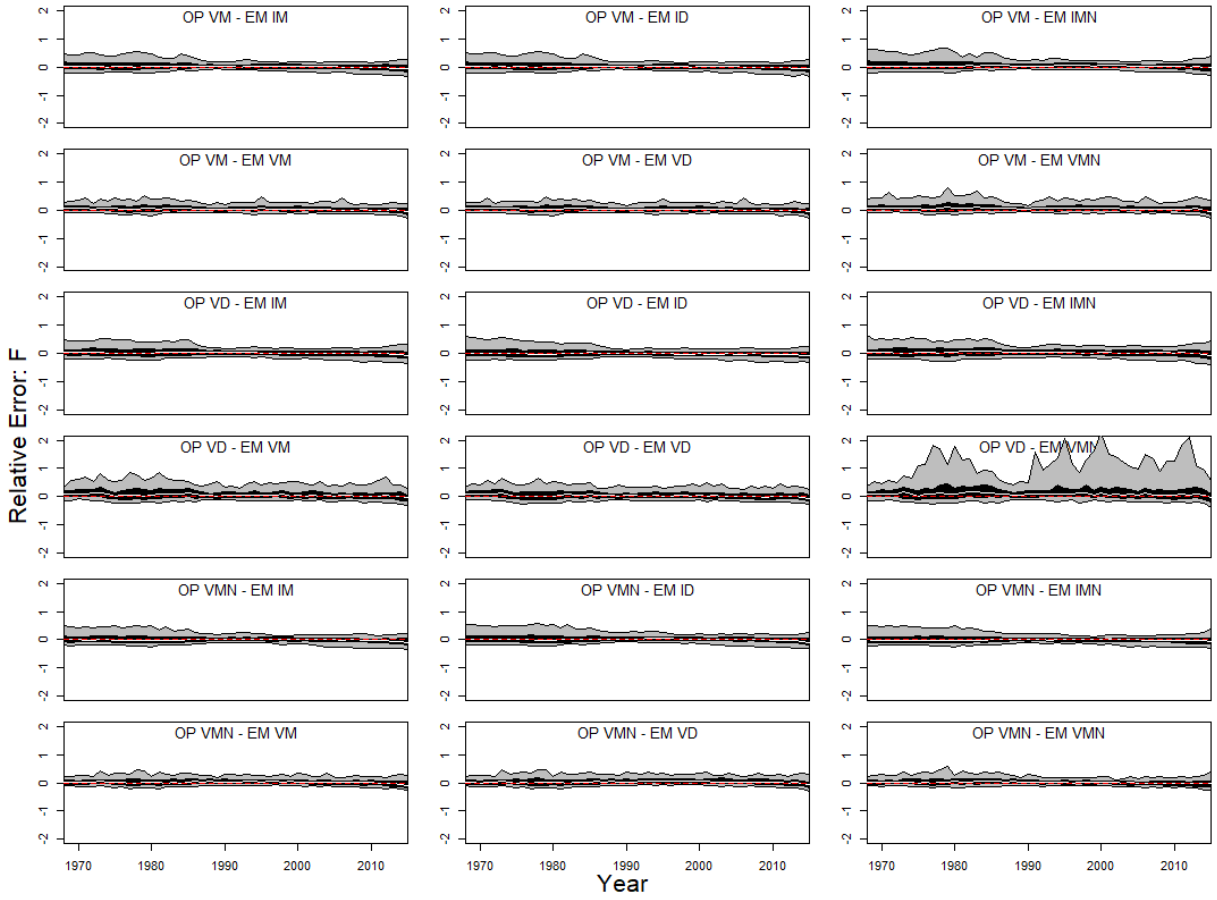

Figure B.13. Relative error for fully selected fishing mortality (F) for the simulation study examining likelihood functions in relation to time-varying selectivity when the operating model (OP) has time-varying (V) selectivity. The generating functions are either multinomial (M), Dirichlet-multinomial (D), or multivariate-normal (MN). The estimation methods (EM) assume either time-invariant (I) or time-varying (V) selectivity. The likelihood functions for the estimation methods are either multinomial (M), Dirichlet-multinomial (D), or multivariate-normal (MN). The white line is the median relative error, the red dash line is the zero line, the black shaded region is the 50% quantile and the grey shaded region is the 90% quantile.

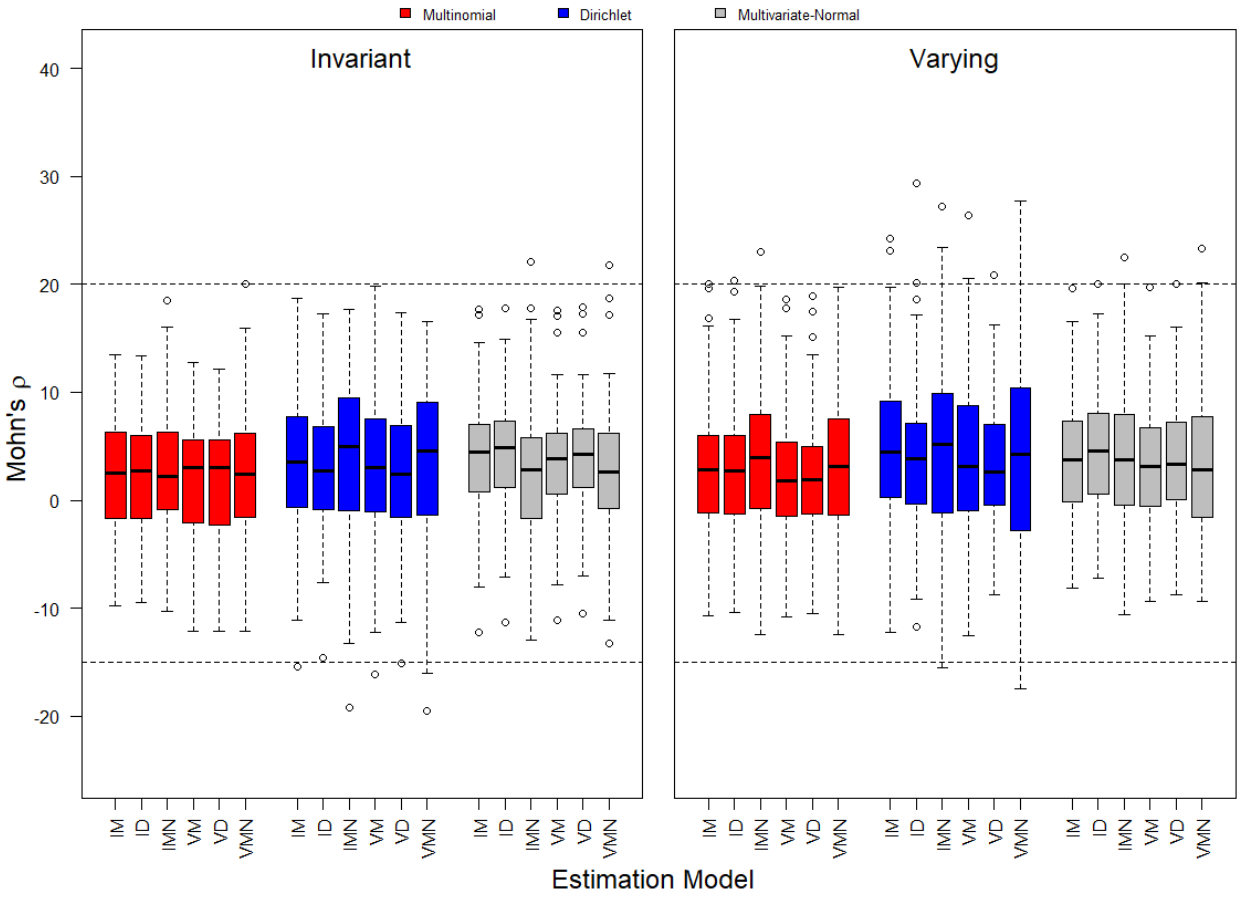

Figure B.14. Boxplots of Mohn's ρ for spawning stock biomass times 100 from the study examining likelihood functions in relation to time-varying selectivity. The left boxplots are for the operating models with time-invariant selectivity. The right boxplots are for the operating models with time-varying selectivity. The colors indicate the generating function. The x-axis specifies the estimation method. The first letter indicates the selectivity assumption (I – time-invariant, V – time-varying). The rest specifies the likelihood (M – multinomial, D – Dirichlet-multinomial, MN – multivariate-normal). The dashed lines mark the limits for Mohn's ρ that do not suggest retrospective patterns.

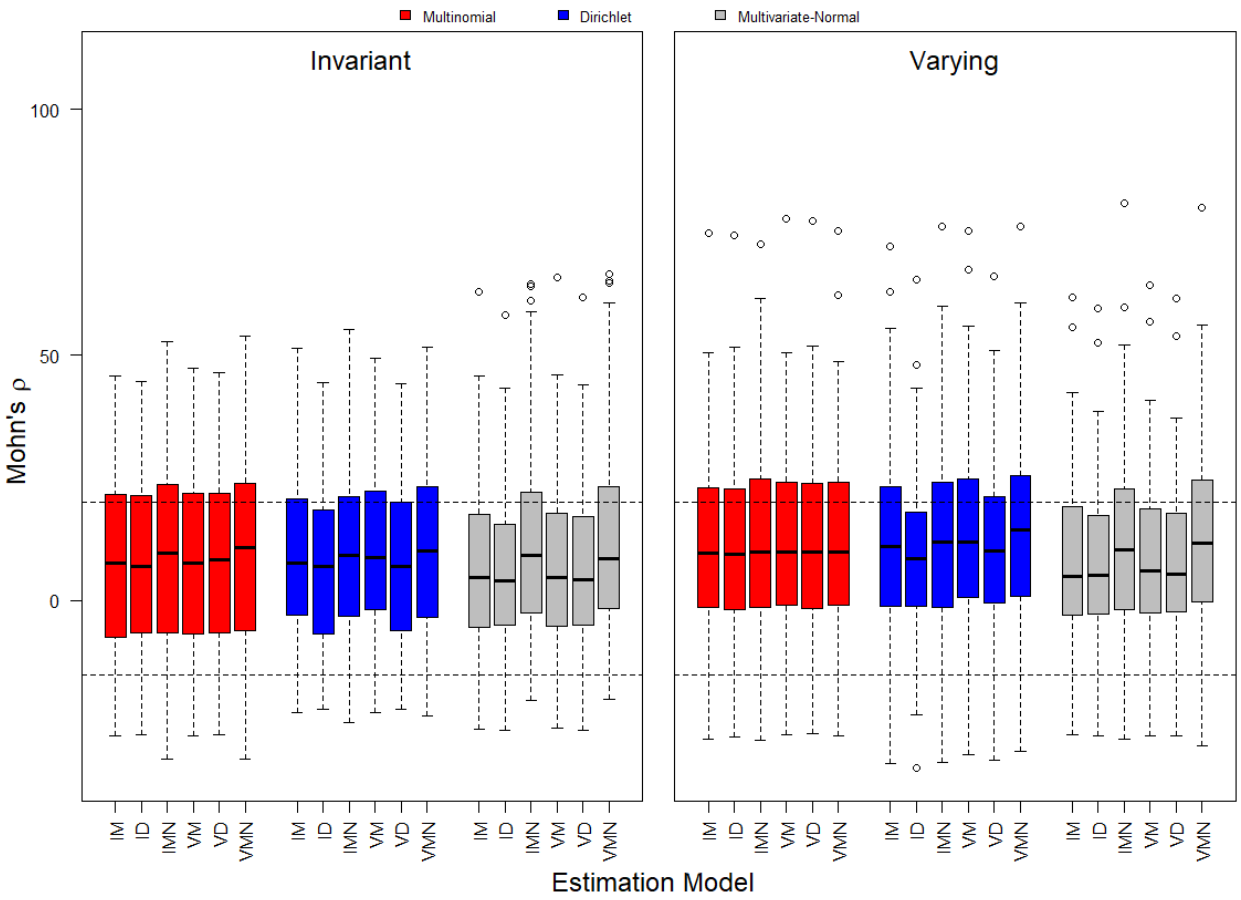

Figure B.15. As for Figure B.14, except the results pertain to the Mohn's ρ values for recruitment, times 100.

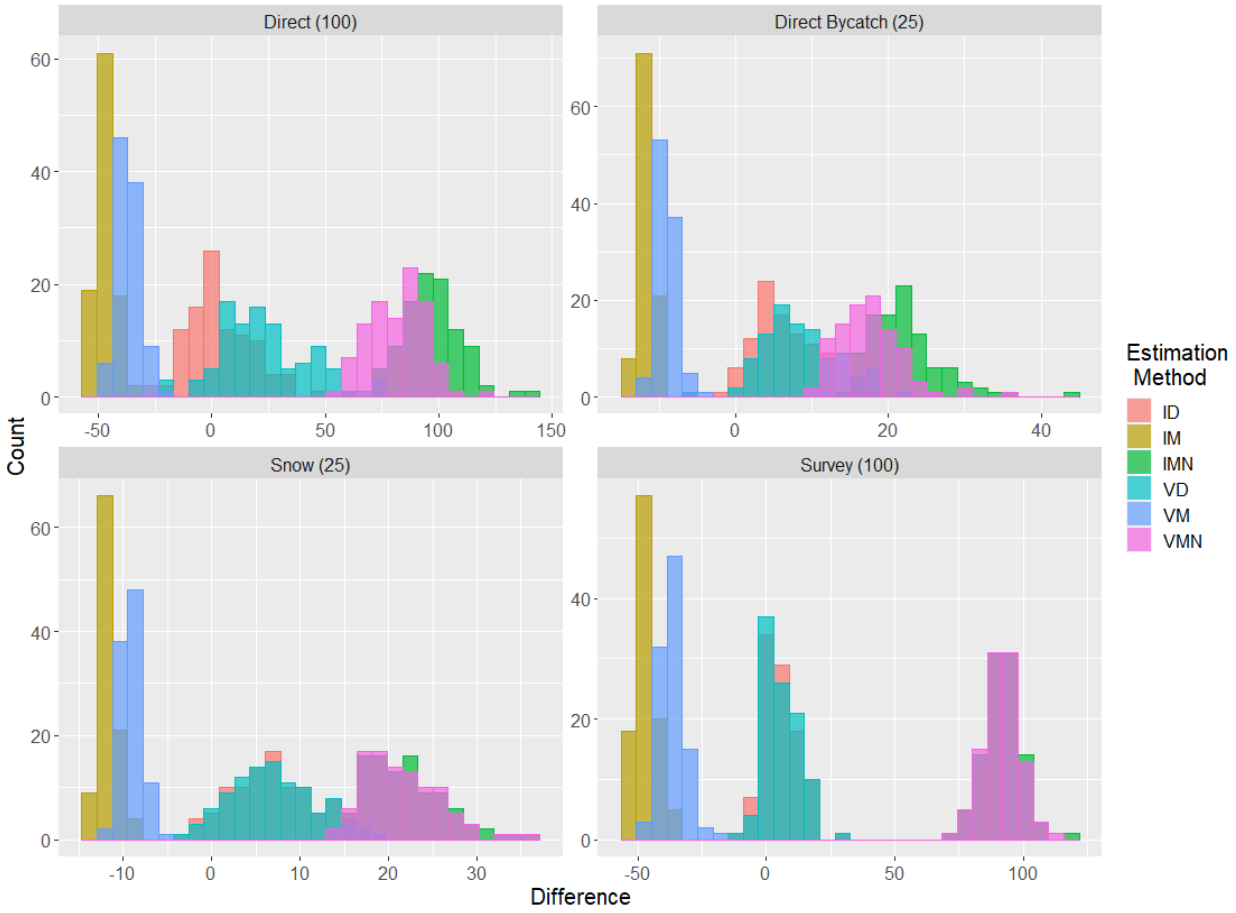

Figure B.16. A histogram of the difference between the estimated and true effective sample size when the operating model has a Dirichlet-multinomial generating function and time-invariant selectivity. Each plot is for a different size-composition data set with the true effective sample size in the header. The colors indicate the estimation method. The first letter in the legend indicates the selectivity assumption (I – time-invariant, V – time-varying) and the rest of the legend specifies the likelihood (M – multinomial, D – Dirichlet-multinomial, MN – multivariate-normal). The closer the values are to zero on the x-axis, the better the performance.

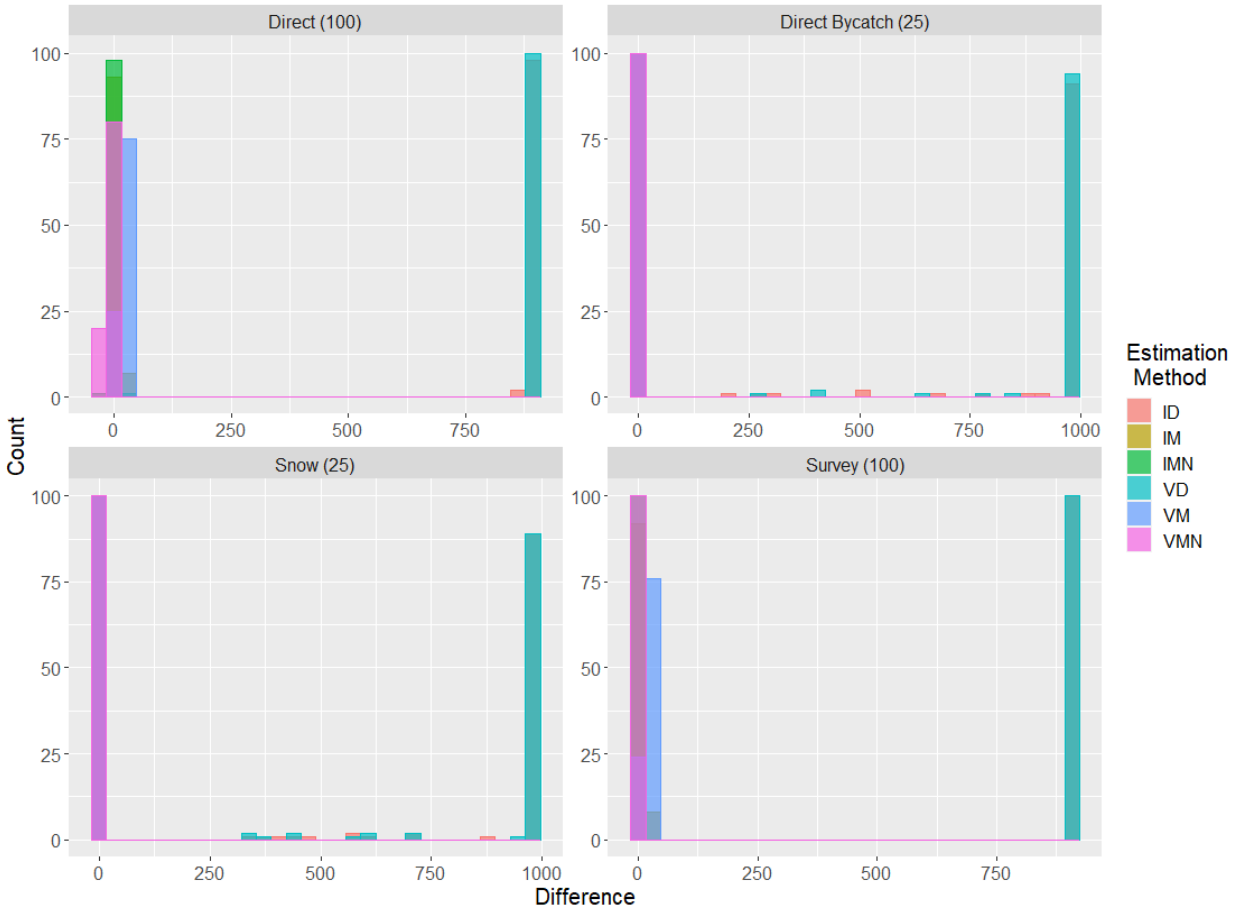

Figure B.17. As for Figure B.16, except the operating model generating function is multinomial.

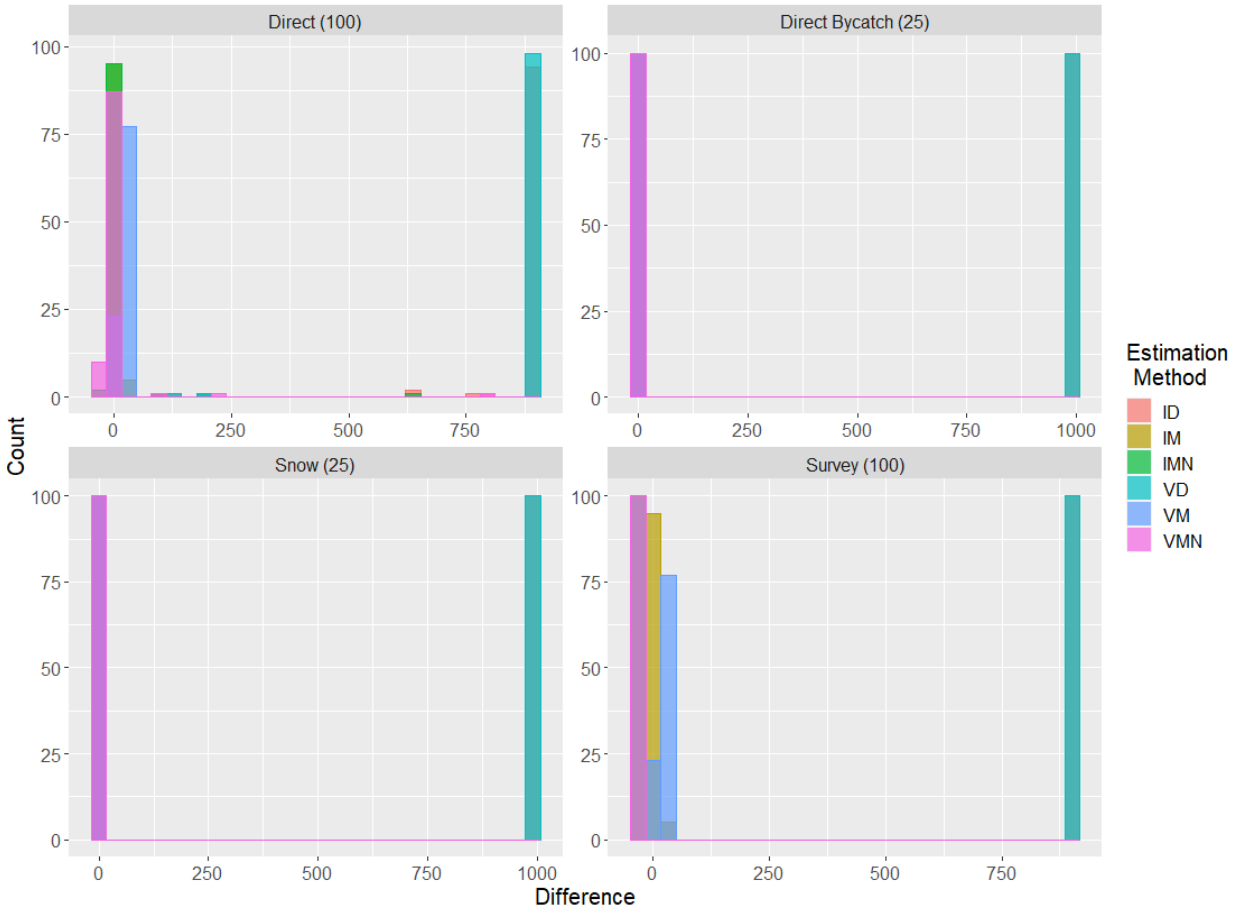

Figure B.18. As for Figure B.16, except the operating model generating function is multivariate normal.

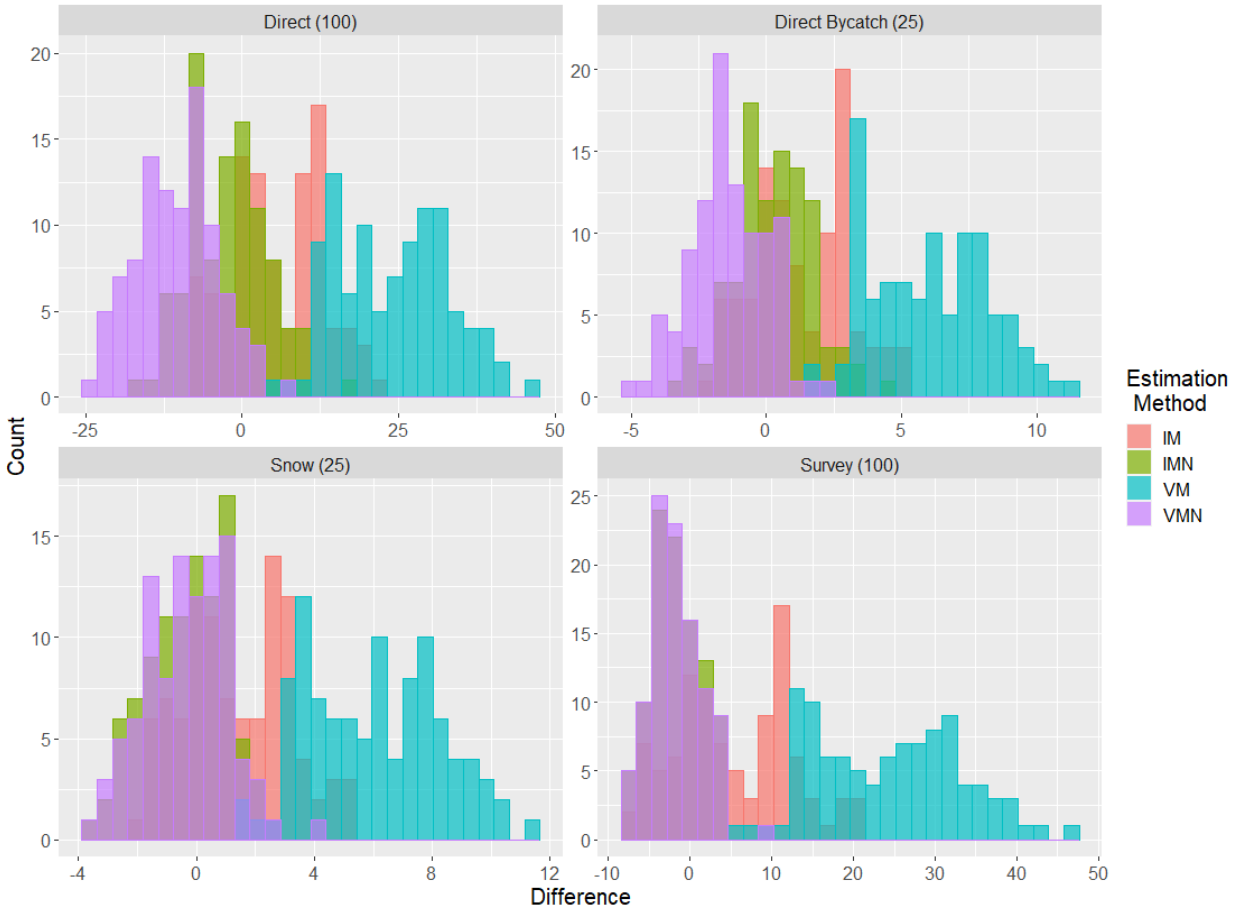

Figure B.19. As for Figure B.17, except estimation methods using the Dirichlet-multinomial likelihood were removed.

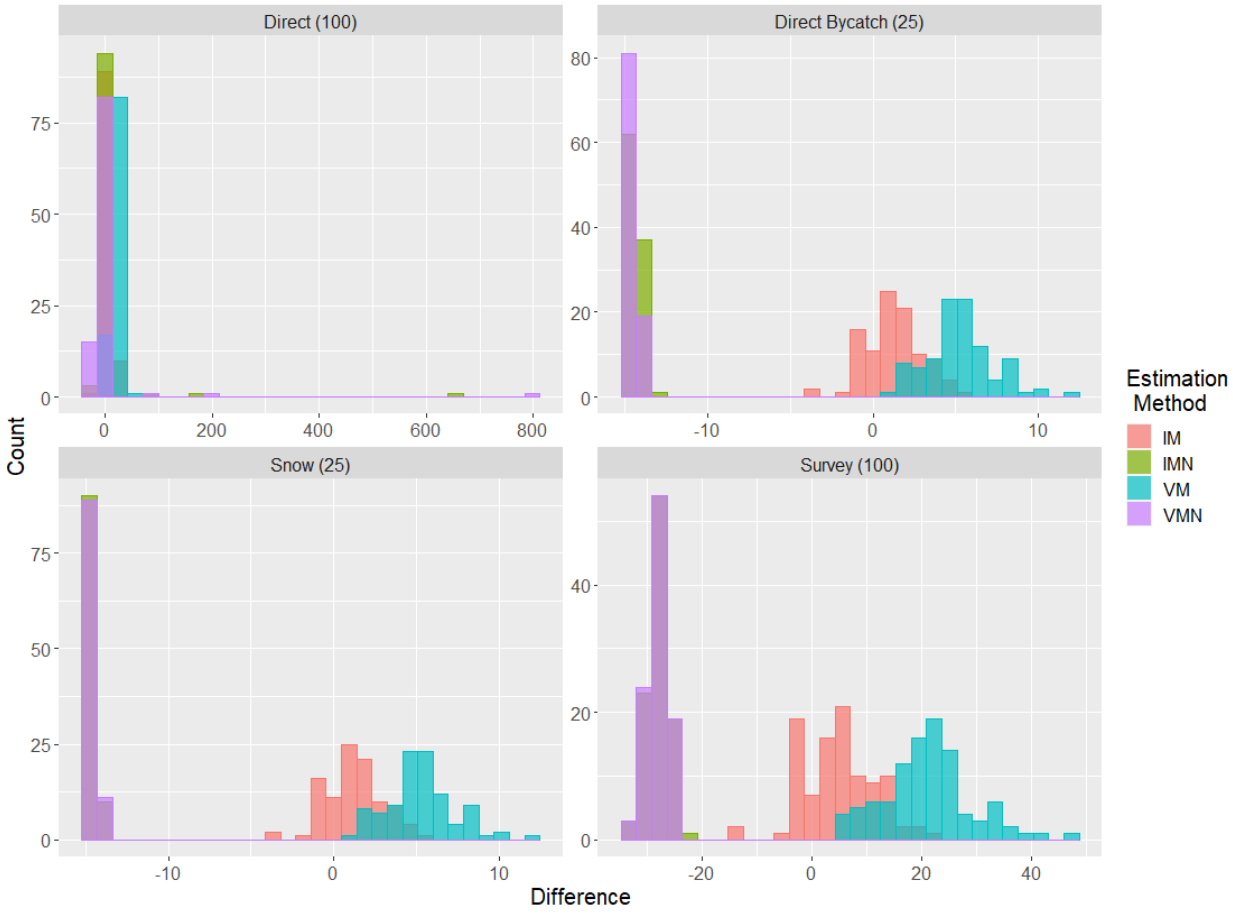

Figure B.20. As for Figure B.18, except estimation methods using the Dirichlet-multinomial likelihood were removed.

Appendix C

Appendix C1: *Maintaining population maturity curve*

In population dynamics models, it is assumed that the population follows a single maturity curve (typically logistic). A size-structure model that tracks the population by maturity stage must have an equation to determine what proportion of immature individuals transitioning from size-class *j* to size-class *i* must mature for the proportion mature in size-class *i* to be unchanged from year to year. After each year, a proportion of individuals that started the year in size-class *i* remain in that size-class. The rest transition into a larger size-class (we assume individuals cannot shrink). We assume that the proportion mature in size-class *i* matches the maturity curve at the beginning of each year. Therefore, the proportion mature of the proportion of individuals remaining in a size-class *i* already mirror the maturity curve. To maintain this proportion, we need to ensure that individuals transitioning from size-class *j* to size-class *i* have the same proportion mature as in size-class *i*. This is done by solving for $Pmat_{j,i}$ in the following equation:

$$Mat_i = \frac{N_j X_{j,i} Mat_j + N_j X_{j,i}(1 - Mat_j) Pmat_{j,i}}{N_j X_{i,j}} \quad \text{(C1.1)}$$

where N_j is the number of individuals in size-class *j*. Equation C1.1 can be simplified to:

$$Pmat_{j,i} = \frac{Mat_{s,i} - Mat_{s,j}}{(1 - Mat_{s,i})} \quad \text{(C1.2)}$$

Equations C1.2 only works when we assume the individuals remaining in the same size-class after one year cannot mature. Also, the proportion mature in size-class *i* must be bigger than the proportion mature in size-class *j* or $Pmat_{j,i}$ will be negative which is biologically impossible. Equation C1.2 does not ensure that the population follows the maturity curve when there is terminal molt. The assumption that the proportion of mature individuals staying in size-class *i* after one year matches the maturity curve is violated since mature individuals do not grow. The equation for $Pmat_{j,i}$ becomes significantly more complicated with a higher chance of producing biological impossibilities.

Table C.1. Median relative errors (MREs) and median absolute relative errors (MAREs) (expressed as percentages) for *M* estimates for the Single operating model for the simulation study exploring model misspecification. Estimation method indicates how *M* is estimated. The numbers are the time-blocks from the blocking design. E-SS estimates three values for *M*: mature males (M), mature females (F) and immature males/female (I). The bold MAREs are exceptions to growth pre-specified leading to smaller MAREs than growth estimated.

Estimation method	MRE				MARE			
	No terminal molt		Terminal molt		No terminal molt		Terminal molt	
	Growth estimated	Growth pre-specified	Growth estimated	Growth pre-specified	Growth estimated	Growth pre-specified	Growth estimated	Growth pre-specified
P-1	0	0	0	0	0	0	0	0
P-0.5	-50	-50	-50	-50	50	50	50	50
P-1.5	50	50	50	50	50	50	50	50
E-M	3.21	-1.16	-0.93	-1.93	4.40	2.19	4.42	3.40
E-SS (M)	-4.49	-5.48	-1.60	-1.45	5.92	5.56	5.77	4.81
E-SS(F)	6.15	2.23	-0.94	-3.94	6.18	2.89	5.72	4.71
E-SS (I)	24.21	11.25	-0.45	-11.79	24.21	11.41	**10.29**	**11.92**
1	-1.03	-3.81	-5.24	-5.68	5.85	5.45	8.68	6.89
2	4.75	1.11	2.30	3.95	12.95	11.21	**11.29**	**11.98**
3	3.75	-0.71	0.86	0.99	11.17	8.84	10.36	9.57
4	3.96	-0.21	-3.05	-3.20	8.17	6.64	8.53	7.91
5	2.33	-2.23	1.45	-1.74	5.59	5.46	7.10	6.93
6	3.40	-0.75	-2.52	-2.64	8.04	6.95	8.37	7.16
7	3.71	0.06	-0.48	-3.4	**7.89**	**7.92**	9.54	8.70
8	2.80	-0.76	-3.23	-4.6	8.57	6.87	10.67	9.78
9	4.76	0.98	0.57	-2.63	10.00	8.50	8.89	7.46
10	8.87	5.14	-3.66	-2.43	**18.71**	**19.15**	15.97	15.20

Table C.2. Median relative errors (MREs) and median absolute relative errors (MAREs) (expressed as percentages) for *M* estimates when the operating model is Sex + Stage for the simulation study exploring model misspecification. The bold MARE values are exceptions to growth pre-specified producing smaller values than growth estimated.

Run			MRE			MARE		
Terminal molt	**Growth estimated?**	**Estimation method**	**Male *M***	**Female *M***	**Immature *M***	**Male *M***	**Female *M***	**Immature *M***
Both	Both	P-1	5.17	-4.69	27.08	5.17	4.69	27.08
		P-0.5	-47.41	-52.34	-36.46	47.41	52.34	36.46
		P-1.5	57.76	42.97	90.62	57.76	42.97	90.62
No	Yes	E-M	5.23	-4.64	27.15	5.30	**4.67**	27.15
		E-SS	-4.14	5.23	25.08	**5.96**	5.83	25.32
	Pre-specified	E-M	-2.34	-11.49	18.01	2.65	**11.49**	18.01
		E-SS	-6.17	1.20	10.17	**6.17**	2.38	11.29
Yes	Yes	E-M	5.89	-4.03	27.96	5.93	6.43	27.96
		E-SS	-1.13	2.68	6.62	5.97	5.70	14.82
	Pre-specified	E-M	5.66	-4.24	27.68	5.68	4.37	27.68
		E-SS	-1.40	-3.40	-8.49	4.68	4.52	9.77

Table C.3. Median relative errors (MREs) and median absolute relative errors (MAREs) (expressed as percentages) between E-B *M* estimates and the average true *M* for each time-block when the operating model is Time-varying for the simulation study exploring model misspecification. The bold MARE values are exceptions to growth pre-specified producing smaller values than growth estimated.

Time block	MRE				MARE			
	No terminal molt		Terminal molt		No terminal molt		Terminal molt	
	Growth estimated	Growth pre-specified	Growth estimated	Growth pre-specified	Growth estimated	Growth pre-specified	Growth estimated	Growth pre-specified
1	14.78	11.38	3.36	4.90	16.22	15.12	**11.20**	**13.05**
2	0.84	-1.65	0.71	1.82	**16.41**	**16.75**	16.93	16.21
3	3.62	-0.43	-3.22	-1.93	**9.56**	**9.62**	**11.03**	**11.07**
4	3.49	0.95	-4.82	-3.44	8.38	7.34	**11.39**	**11.69**
5	4.17	0.81	-2.47	-1.69	6.47	4.64	9.63	7.87
6	3.43	-0.21	-3.55	-3.20	8.14	7.56	8.17	7.44
7	3.66	0.46	-3.31	-3.71	7.40	5.73	9.13	6.90
8	1.75	-2.32	-3.34	-2.21	6.99	6.59	7.61	7.36
9	4.92	3.39	-1.41	-0.15	7.93	7.82	**7.26**	**7.85**
10	3.52	3.58	0.39	-0.94	18.14	18.03	19.72	18.37

Table C.4. Median absolute relative errors (MAREs) for the growth parameter estimates for the Single operating model for the simulation study exploring model misspecification.

Estimation method	Male		Female	
	No terminal molt	**Terminal molt**	**No terminal molt**	**Terminal molt**
		Growth Parameter = L_∞		
P-1	17.21	19.49	6.77	8.88
P-0.5	46.47	25.64	51.64	53.38
P-1.5	291.30	19.33	43.35	10.12
E-M	18.18	19.02	5.52	8.89
E-SS	14.72	19.49	8.60	9.31
E-B	19.82	19.93	6.21	8.50
		Growth Parameter = k		
P-1	22.01	27.7	11.61	12.53
P-0.5	37.14	23.55	36.07	24.98
P-1.5	72.17	34.2	22.42	27.36
E-M	23.72	30.3	10.87	14.78
E-SS	21.77	28.85	11.66	17.52
E-B	23.12	28.87	10.87	14.58
		Growth Parameter = $\tilde{\sigma}$		
P-1	4.68	2.98	3.81	3.15
P-0.5	13.13	14.31	25.07	41.04
P-1.5	5.98	5.35	5.07	12.82
E-M	5.31	3.33	3.74	4.02
E-SS	6.12	3.47	3.88	4.65
E-B	5.18	3.39	3.79	4.22

Table C.5. Median absolute relative errors (MAREs) for the growth parameter estimates for the Sex + Stage operating model for the simulation study exploring model misspecification.

Estimation method	Male		Female	
	No terminal molt	**Terminal molt**	**No terminal molt**	**Terminal molt**
		Growth Parameter = L_∞		
P-1	19.88	15.41	9.30	10.69
P-0.5	36.10	17.74	57.61	59.04
P-1.5	220.16	12.41	43.01	11.46
E-M	17.84	15.35	9.45	12.54
E-SS	14.62	16.28	10.62	9.56
		Growth Parameter = k		
P-1	26.42	22.29	12.81	19.57
P-0.5	28.09	17.86	31.20	30.89
P-1.5	66.85	45.50	29.42	21.30
E-M	24.39	23.21	13.31	16.24
E-SS	25.88	20.98	13.72	16.65
		Growth Parameter = $\widetilde{\sigma}$		
P-1	5.96	2.82	4.72	4.63
P-0.5	8.18	12.38	31.83	48.05
P-1.5	5.06	3.53	4.05	8.77
E-M	6.23	3.59	4.77	4.82
E-SS	4.63	3.13	4.21	4.26

Table C.6. Median absolute relative errors (MAREs) for the growth parameter estimates for the Time-varying operating model for the simulation study exploring model misspecification.

Estimation method	Male		Female	
	No terminal molt	**Terminal molt**	**No terminal molt**	**Terminal molt**
		Growth Parameter = L_∞		
P-1	13.65	19.84	7.84	10.21
P-0.5	69.27	59.07	63.74	64.36
P-1.5	291.3	13.86	63.73	6.98
E-M	21.23	20.42	6.99	10.9
E-B	18.4	20.03	7.43	10.86
		Growth Parameter = k		
P-1	18.37	23.2	12.98	21.18
P-0.5	143.7	30.83	93.92	33.82
P-1.5	72.56	39.25	33.02	38.36
E-M	21.48	22.26	11.64	18.57
E-B	20.48	22.63	12.41	18
		Growth Parameter = $\widetilde{\sigma}$		
P-1	4.24	2.75	4.72	2.93
P-0.5	30.73	24.44	40.33	37.52
P-1.5	4.66	4.49	5.78	12.13
E-M	4.92	2.95	4.57	3.44
E-B	4.81	3.18	4.67	3.7

Table C.7. Median relative errors (MREs) (expressed as percentages) for the management quantities from the Single operating model for the simulation study exploring model misspecification.

Terminal molt	Growth estimated?	Estimation method	SSB_0	$SSB_{16/0}$	$SSB_{16/68}$	SSB_{16}	SSB_{avg}
No	Yes	P-1	1.92	-3.16	0.47	0.05	-1.69
		P-0.5	49.71	-32.51	41.01	4.35	-9.68
		P-1.5	-8.85	4.74	-22.56	-3.10	3.45
		E-M	1.41	-1.72	-1.43	0.03	-1.21
		E-SS	8.59	-8.38	2.59	-0.37	-1.97
		E-B	8.61	-10.88	-4.76	-3.94	-1.57
	Pre-specified	P-1	2.36	-3.80	-2.03	-0.54	-0.66
		P-0.5	64.07	-24.05	86.97	26.09	-15.46
		P-1.5	-6.40	1.19	-30.3	-4.05	12.84
		E-M	2.69	-3.73	-0.79	-0.31	-0.93
		E-SS	6.66	-7.38	1.80	-1.29	-2.36
		E-B	8.94	-13.01	-4.33	-3.94	-1.18
Yes	Yes	P-1	2.42	-0.64	-0.95	1.36	-0.79
		P-0.5	46.92	-24.97	14.72	10.72	0.04
		P-1.5	-12.21	12.63	-6.50	-1.42	-0.79
		E-M	3.27	-1.15	-1.37	1.69	-0.77
		E-SS	2.44	-0.99	-0.83	0.99	-0.37
		E-B	7.12	-12.04	-2.57	-1.02	-0.03
	Pre-specified	P-1	2.59	-0.06	-1.73	1.79	1.37
		P-0.5	52.37	-16.00	20.1	32.54	8.15
		P-1.5	-11.88	13.49	-8.23	-1.70	1.77
		E-M	4.72	-0.54	-1.43	2.65	1.46
		E-SS	0.31	1.81	-0.80	1.03	-0.39
		E-B	7.88	-10.02	-2.59	0.24	2.19

Table C.8. The percent difference from the lowest median absolute relative error (MARE) by terminal molt scenario for each management quantity from the Single operating model for the simulation study exploring model misspecification.

Terminal molt	Growth estimated?	Estimation method	SSB_0	$SSB_{16/0}$	$SSB_{16/68}$	SSB_{16}	SSB_{avg}
No	Yes	P-1	13.81	57.81	4.09	0	43.16
		P-0.5	964.71	556.93	706.52	61.78	545.47
		P-1.5	95.28	31.84	343.63	27.76	154.65
		E-M	4.56	51.41	13.86	4.94	28.38
		E-SS	108.20	116.00	41.36	4.12	55.36
		E-B	142.90	294.16	135.54	173.64	48.08
	Pre-specified	P-1	0	57.25	7.22	8.32	0
		P-0.5	1272.11	418.18	1610.41	475.04	931.08
		P-1.5	48.86	0	495.83	10.87	755.86
		E-M	15.97	62.87	0	13.76	28.13
		E-SS	73.47	89.54	18.21	21.83	81.46
		E-B	155.22	293.34	134.81	183.36	37.81
Yes	Yes	P-1	0	25.48	6.35	8.73	3.82
		P-0.5	759.74	228.3	120.02	161.14	19.87
		P-1.5	123.70	66.07	26.32	15.28	2.61
		E-M	7.75	16.64	8.44	16.81	0
		E-SS	33.92	0	3.82	21.62	1.18
		E-B	105.64	147.85	55.21	136.46	0.69
	Pre-specified	P-1	8.99	8.67	0	0	17.28
		P-0.5	859.65	123.75	172.76	638.65	326.49
		P-1.5	117.71	77.39	30.22	19.43	18.66
		E-M	12.25	9.50	5.02	4.59	15.12
		E-SS	9.34	6.43	1.48	1.53	10.83
		E-B	94.36	148.74	40.96	106.99	37.89

Table C.9. As for Table C.7 but for the Sex + Stage operating model.

Terminal molt	Growth estimated?	Estimation method	SSB_0	$SSB_{16/0}$	$SSB_{16/68}$	SSB_{16}	SSB_{avg}
No	Yes	P-1	8.45	-5.29	-0.94	1.46	0.16
		P-0.5	58.28	-34.57	36.36	3.77	-8.64
		P-1.5	-5.05	4.24	-22.43	-0.02	6.4
		Est-M	8.67	-5.43	-0.81	1.57	0.17
		E-SS	8.72	-8.64	4.32	0.00	-1.98
	Pre-specified	P-1	15.49	-7.28	-5.62	6.15	9.81
		P-0.5	73.52	-29.17	63.94	24.73	-9.14
		P-1.5	7.06	4.94	-28.17	13.51	29.77
		Est-M	18.46	-8.66	0.40	7.54	6.96
		E-SS	8.63	-6.67	5.05	-0.38	-2.34
Yes	Yes	P-1	8.23	-2.93	-0.67	3.71	1.97
		P-0.5	56.27	-30.65	10.56	9.74	1.42
		P-1.5	-8.22	15.07	-5.39	3.60	2.94
		Est-M	7.39	-2.96	-0.96	3.84	1.99
		E-SS	4.45	-2.70	1.85	2.49	0.35
	Pre-specified	P-1	12.03	-0.42	-0.36	12.12	11.81
		P-0.5	64.49	-19.85	18.1	35.55	14.30
		P-1.5	-2.43	16.86	-5.79	14.61	15.34
		Est-M	12.29	0.03	0.37	12.17	11.29
		E-SS	0.93	1.26	1.25	2.63	0.80

Table C.10. As for Table C.8 but for the Sex + Stage operating model.

Terminal molt	Growth estimated?	Estimation method	SSB_0	$SSB_{16/0}$	$SSB_{16/68}$	SSB_{16}	SSB_{avg}
No	Yes	P-1	32.33	14.57	7.38	24.10	0
		P-0.5	794.78	390.22	411.76	65.15	441.53
		P-1.5	0	1.33	215.74	0	301.16
		E-M	36.69	15.13	0	24.93	0.09
		E-SS	64.03	51.56	15.81	11.45	39.21
	Pre-specified	P-1	137.83	24.75	2.43	47.25	514.96
		P-0.5	1028.75	313.63	800.03	398.78	473.39
		P-1.5	12.94	0	296.57	172.54	1766.65
		E-M	183.42	44.65	2.92	57.39	336.25
		E-SS	49.18	47.25	22.32	16.70	54.73
Yes	Yes	P-1	45.94	38.33	20.59	29.37	1.00
		P-0.5	852.71	419.91	78.10	92.41	9.71
		P-1.5	39.16	155.61	19.45	24.56	37.5
		E-M	40.41	30.68	21.96	27.34	1.28
		E-SS	47.40	78.94	21.69	8.86	4.86
	Pre-specified	P-1	103.72	9.52	8.27	130.13	425.02
		P-0.5	991.99	236.70	151.52	575.19	535.58
		P-1.5	0	185.97	0	177.47	581.94
		E-M	108.01	0	4.96	131.14	401.88
		E-SS	17.38	46.81	4.51	0	0

Table C.11. As for Table C.7 but for the Time-varying operating model.

Terminal molt	Growth estimated?	Estimation method	SSB_0	$SSB_{16/0}$	$SSB_{16/68}$	SSB_{16}	SSB_{avg}
No	Yes	P-1	9.92	23.55	-0.31	35.96	1.35
		P-0.5	26.28	7.48	33.73	41.10	-5.38
		P-1.5	1.72	23.97	-22.82	26.22	6.15
		E-M	9.16	21.89	-1.91	34.79	1.70
		E-B	39.76	-30.68	-10.05	-2.27	-1.27
	Pre-specified	P-1	9.96	22.37	0.20	34.96	1.85
		P-0.5	59.92	49.51	239.89	134.46	-13.08
		P-1.5	1.96	17.33	-41.79	18.11	15.27
		E-M	9.78	22.22	-2.80	34.30	2.09
		E-B	40.55	-31.30	-8.36	-2.79	-1.41
Yes	Yes	P-1	5.61	17.79	10.71	25.28	-0.28
		P-0.5	37.00	-0.19	23.74	37.67	-0.20
		P-1.5	-4.03	27.01	8.08	22.76	-0.72
		E-M	5.47	16.28	10.92	25.35	-0.52
		E-B	45.19	-30.97	-12.88	-1.50	0.89
	Pre-specified	P-1	5.83	17.71	10.07	25.45	0.17
		P-0.5	43.37	15.42	40.68	70.67	6.49
		P-1.5	-3.08	25.86	4.74	21.40	1.05
		E-M	5.80	16.91	10.81	25.45	0.32
		E-B	43.77	-30.33	-14.68	-1.15	1.38

Table C.12. As for Table C.8 but for the Time-varying operating model.

Terminal molt	Growth estimated?	Estimation method	SSB_0	$SSB_{16/0}$	$SSB_{16/68}$	SSB_{16}	SSB_{avg}
No	Yes	P-1	77.38	38.50	0	204.01	19.71
		P-0.5	369.90	0	435.57	247.48	137.14
		P-1.5	0	40.94	262.34	121.69	182.64
		E-M	63.85	28.75	11.49	194.10	24.53
		E-B	611.09	89.53	158.25	0	5.96
	Pre-specified	P-1	78.08	31.55	10.98	195.57	7.05
		P-0.5	971.57	191.14	3709.51	1036.82	476.35
		P-1.5	11.89	1.94	563.61	53.12	572.91
		E-M	74.94	30.65	5.38	190.01	25.05
		E-B	625.19	91.22	141.96	3.75	0
Yes	Yes	P-1	35.92	9.10	34.40	119.08	4.16
		P-0.5	587.67	10.12	151.04	226.42	12.73
		P-1.5	0	62.71	8.60	97.19	4.82
		E-M	31.72	0	38.43	119.64	0
		E-B	739.86	89.19	91.65	0	36.87
	Pre-specified	P-1	29.04	11.99	37.08	120.50	9.34
		P-0.5	706.17	17.19	330.15	512.33	203.72
		P-1.5	0.89	55.81	0	85.42	16.80
		E-M	31.10	2.15	37.79	120.50	5.41
		E-B	713.49	84.00	103.16	2.25	47.19

Table C.13. Median absolute relative errors (MAREs) for the management quantities for the Single operating model for the simulation study exploring data quality.

Terminal molt	Growth estimated?	Estimation method	Data	SSB_0	$SSB_{16/0}$	$SSB_{16/68}$	SSB_{16}	SSB_{avg}
No	Yes	P-1	Base	5.31	7.81	5.29	4.54	2.15
			H-SS	5.58	6.54	5.64	3.97	2.31
			S-CV	6.27	7.05	5.39	2.76	1.30
		E-M	Base	4.88	7.49	5.79	4.76	1.93
			H-SS	5.00	6.84	6.78	4.01	1.86
			S-CV	5.79	6.83	5.56	2.90	1.33
		E-SS	Base	9.72	10.69	7.19	4.72	2.33
			H-SS	8.43	10.05	7.52	4.48	2.71
			S-CV	8.99	10.09	5.61	3.06	1.27
	Pre-specified	P-1	Base	4.67	7.78	5.45	4.91	1.50
			H-SS	5.84	6.26	6.28	4.45	1.41
			S-CV	5.41	6.64	4.29	2.83	1.32
		E-M	Base	5.41	8.06	5.08	5.16	1.92
			H-SS	5.85	6.53	6.24	4.66	1.68
			S-CV	5.60	6.37	4.84	2.91	1.38
		E-SS	Base	8.10	9.38	6.01	5.53	2.72
			H-SS	7.05	8.15	6.83	4.80	3.13
			S-CV	6.65	7.49	5.22	3.18	1.32
Yes	Yes	P-1	Base	5.46	9.55	7.84	4.79	1.98
			H-SS	6.13	8.76	7.44	4.72	1.94
			S-CV	6.23	7.68	6.51	3.69	1.30
		E-M	Base	5.88	8.87	7.99	5.15	1.91
			H-SS	5.39	8.47	7.83	4.61	1.93
			S-CV	5.93	6.96	6.83	3.65	1.36
		E-SS	Base	7.31	7.61	7.65	5.36	1.93
			H-SS	6.76	9.49	7.57	4.48	1.92
			S-CV	7.44	6.80	7.68	3.60	1.61
	Pre-specified	P-1	Base	5.95	8.27	7.37	4.40	2.24
			H-SS	6.47	7.93	7.66	4.71	2.02
			S-CV	5.90	7.09	6.96	2.96	1.80
		E-M	Base	6.13	8.33	7.74	4.61	2.20
			H-SS	5.61	8.40	6.93	4.93	1.80
			S-CV	5.93	6.79	7.47	2.74	1.95
		E-SS	Base	5.97	8.10	7.48	4.47	2.12
			H-SS	5.99	9.72	7.07	4.47	2.06
			S-CV	5.76	7.78	6.55	3.30	1.62

Table C.14. Median absolute relative errors (MAREs) for the management quantities for the Sex + Stage operating model for the simulation study exploring data quality.

Terminal molt	Growth estimated?	*M* method	Data	SSB_0	$SSB_{16/0}$	$SSB_{16/68}$	SSB_{16}	SSB_{avg}
No	Yes	P-1	Base	8.62	8.08	7.63	6.15	1.59
			H-SS	8.42	8.22	7.09	6.21	1.53
			S-CV	8.36	7.48	5.28	3.35	1.46
		E-M	Base	8.9	8.12	7.1	6.19	1.6
			H-SS	8.13	7.95	7.63	6.43	1.73
			S-CV	8.21	7.95	5.9	3.51	1.47
		E-SS	Base	10.68	10.69	8.23	5.53	2.22
			H-SS	9.26	9.61	9.15	4.99	2.34
			S-CV	9.47	8.3	5.81	3.22	1.34
	Pre-specified	P-1	Base	15.49	8.8	7.28	7.3	9.81
			H-SS	13.58	8.6	7.54	6.76	8.98
			S-CV	13.4	9.52	7.77	3.69	7.21
		E-M	Base	18.46	10.2	7.31	7.8	6.96
			H-SS	16.68	9.93	5.99	7.77	5.8
			S-CV	17.83	10.84	4.9	5.23	5.62
		E-SS	Base	9.72	10.38	8.69	5.79	2.47
			H-SS	7.67	8.59	8.79	5.05	2.76
			S-CV	8.96	8.13	6.04	3.27	1.29
Yes	Yes	P-1	Base	8.62	8.16	8.68	6.81	2.27
			H-SS	6.95	6.73	7.19	6.2	2.33
			S-CV	6.94	7.08	6.85	4.49	2.88
		E-M	Base	8.29	7.7	8.78	6.7	2.28
			H-SS	6.39	6.98	7.6	6.01	2.34
			S-CV	6.42	6.58	7.07	4.2	3.09
		E-SS	Base	8.71	10.55	8.76	5.73	2.36
			H-SS	7.09	8.83	7.21	4.88	2
			S-CV	7.83	8.34	7.16	4.03	1.8
	Pre-specified	P-1	Base	12.03	6.46	7.79	12.12	11.81
			H-SS	8.98	8.55	6.57	11.85	10.98
			S-CV	8.95	6.33	6.25	8.89	9.13
		E-M	Base	12.29	5.9	7.55	12.17	11.29
			H-SS	8.38	7.68	6.61	11.21	11.06
			S-CV	8.71	5.28	6.43	9.12	9.73
		E-SS	Base	6.93	8.65	7.52	5.27	2.25
			H-SS	6.34	8.8	6.46	4.72	2.07
			S-CV	7.38	8.01	5.97	3.55	1.7

Table C.15. Difference between estimated and pre-specified growth median absolute relative errors (MAREs) (expressed as percentages) for the management quantities for the simulation study exploring model misspecification. Positive number means estimated growth has a bigger MARE than pre-specified growth.

Terminal Molt	Estimation method	SSB_0	$SSB_{16/0}$	$SSB_{16/68}$	SSB_{16}	SSB_{avg}
		Operating Model = Single				
No	P-1	0.64	0.03	-0.16	-0.37	0.65
	P-0.5	-14.36	6.87	-45.96	-18.75	-5.78
	P-1.5	2.17	1.57	-7.74	0.77	-9.02
	E-M	-0.53	-0.57	0.71	-0.4	0.01
	E-SS	1.62	1.31	1.18	-0.81	-0.39
	E-B	-0.58	0.04	0.04	-0.44	0.15
Yes	P-1	-0.49	1.28	0.47	0.39	-0.26
	P-0.5	-5.45	7.95	-3.89	-21.04	-5.86
	P-1.5	0.33	-0.86	-0.29	-0.18	-0.31
	E-M	-0.25	0.54	0.25	0.54	-0.29
	E-SS	1.34	-0.49	0.17	0.89	-0.19
	E-B	0.61	-0.07	1.05	1.3	-0.72
		Operating Model = Sex + Stage				
No	P-1	-6.87	-0.72	0.35	-1.15	-8.22
	P-0.5	-15.24	5.4	-27.58	-16.54	-0.5
	P-1.5	-0.85	0.1	-5.74	-8.55	-23.37
	E-M	-9.56	-2.08	-0.21	-1.61	-5.36
	E-SS	0.96	0.31	-0.46	-0.26	-0.25
Yes	P-1	-3.41	1.7	0.89	-5.31	-9.54
	P-0.5	-8.22	10.8	-5.29	-25.42	-11.83
	P-1.5	2.31	-1.79	1.39	-8.05	-12.25
	E-M	-4	1.8	1.23	-5.47	-9.01
	E-SS	1.78	1.9	1.24	0.46	0.11
		Operating Model = Time-Varying				
No	P-1	-0.04	1.18	-0.69	1	0.29
	P-0.5	-33.64	-32.51	-206.16	-93.36	-7.7
	P-1.5	-0.67	6.64	-18.97	8.11	-8.85
	E-M	-0.62	-0.33	0.38	0.49	-0.01
	E-B	-0.79	-0.29	1.02	-0.44	0.13
Yes	P-1	0.37	-0.48	-0.25	-0.17	-0.11
	P-0.5	-6.37	-1.17	-16.94	-33	-4.08
	P-1.5	-0.05	1.15	0.81	1.36	-0.26
	E-M	0.04	-0.36	0.06	-0.1	-0.11
	E-B	1.42	0.86	-1.08	-0.26	-0.22

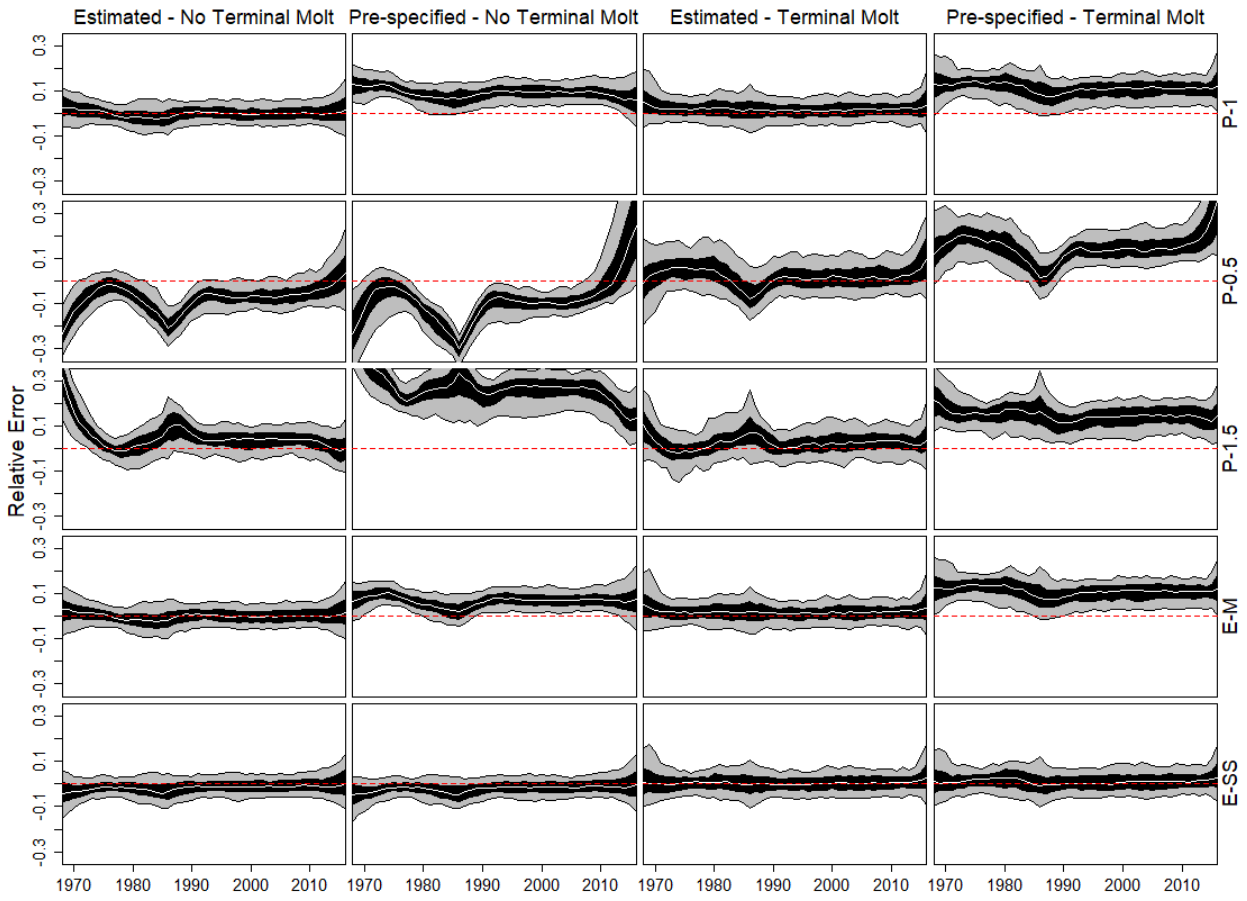

Figure C.1. Relative errors for spawning stock biomass (SSB) for the Sex + Stage operating model for the simulation study exploring model misspecification. The column titles indicate whether growth is estimated or pre-specified and if there is terminal molt. The row titles are the estimation method for *M*. The white line is the median relative error, the red dash line is the zero line, the black shaded region is the 50% quantile, and the grey shaded region is the 90% quantile.