A Management Strategy Evaluation for Transboundary Sablefish in the Northeast Pacific Ocean

Maia Sosa Kapur

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2023

Reading Committee:

André E. Punt, Chair

Melissa A. Haltuch

Timothy E. Essington

Program Authorized to Offer Degree:

School of Aquatic and Fishery Sciences

University of Washington

Abstract

A Management Strategy Evaluation for Transboundary Sablefish in the Northeast Pacific Ocean

Maia Sosa Kapur

Chair of the Supervisory Committee:
Professor André E. Punt
School of Aquatic and Fishery Sciences

Fisheries assessment science has long concerned itself about the incorporation of spatial stock structure into the development, testing and deployment of management strategies. The population dynamics of a stock may not match static political boundaries due to migration, ontogenetic changes in demography, and spatially varying exploitation. Correctly specifying spatial structure can reduce bias in management quantities to a greater degree than correcting other parameters or population processes. While the value of developing spatially-structured assessment models is well recognized, there are many barriers to implementation, including the cost of collecting and evaluating datasets at the appropriate spatial resolution to determine how such assessments should be structured. The management strategy evaluation (MSE) approach enables scientists to test how various spatial assessment configurations might impact management performance, before moving to tactical implementation. Sablefish (Anoplopoma fimbria) of the northeast Pacific support a highly mobile, valuable fishery resource worth over $112 million USD. At the federal level, sablefish in this region are currently managed as three separate populations (Alaska, British Columbia and the US West Coast) by the National Oceanic and Atmospheric Administration (NOAA-US) and the Department of Fisheries and Oceans (DFO – Canada). Recent work has shown sablefish to be genetically mixed across the range, tagging studies confirm high movement rates and there is range wide synchrony in biomass

trends, including declines during the last decade. These observations have led scientists and managers to consider whether evaluating the stock as a spatially structured population might reveal transboundary dynamics, and/or preclude poor management outcomes at the regional scale. An investigation of growth rates throughout the northeast Pacific revealed roughly four regions of distinct sablefish growth. These regions correspond with large oceanographic features that characterize large marine ecosystems (LME). A spatially structured statistical catch-at-age model, conditioned to historical catch, discard, survey and age-composition data from the three management regions suggested that the data can be plausibly fit as a spatially structured population, and corroborated other investigations indicating that LMEs delineate sablefish spawning populations within each management area. An MSE including estimation methods across a gradient of spatial complexity revealed that spatial models of intermediate structural complexity, including those that match the current paradigm in terms of having three modeled areas, can satisfy stakeholder objectives and avoid negative outcomes for the sablefish fishery. The MSE tool built for this dissertation can be used to evaluate management performance in the context of climate-induced changes to reproduction, alternative structural hypotheses, or changes to fleet dynamics.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Andre Punt is the unique sort of mentor who lets the student decide what success looks like, and applies super-human effort to ensure that the student reaches their stated target. When I first sat down with Andre to discuss my goals for my graduate education, he encouraged me to participate in a mix of applied and investigative projects to ensure that both my skillset and network grew during my time at SAFS. Thanks to him, I am proud of the research portfolio I produced over the last several years and hope to give back to the global assessment community in his example of rigor and service.

Melissa Haltuch provided the logistical, intellectual and financial support for this degree to happen. She is a masterful communicator and took care of the difficult parts of organizing dozens of international collaborators so that I could focus on the technical challenge of conducting a transboundary MSE. She has the unique ability to design a project that is at the right scale, at the right time, including the right people and asking a question that pertinent to a specific case (sablefish in this part of the world) and beyond (most exploited marine resources under climate change!). I thank her for her example of balanced scientific achievement, strong personal integrity and not losing sight of what's important.

My committee (Tim Essington and Trevor Branch) were the perfect "third parties" to help me think outside my assessment box. Tim allowed me to TA his ecological modeling course, which helped me revisit key math-stat concepts that I'd learned informally, and be inspired by his commitment to student learning. I thank them for their consistent attention to my degree and support of my career direction.

The Punt Lab, particularly Kristin Privitera-Johnson, Grant Adams, Lee Cronin-Fine, John Best and Madi Heller Shipley: these are the types of colleagues one is lucky to have, because they

also make great friends. We stuck together during the pandemic, and made the unspoken decision to be sincerely and completely one another's supporter, never competitor. I thank them for their judgment-free advice and their humor.

Nearly countless "freebie" mentors and collaborators I have been blessed to meet over the last five years, including Rick Methot, Jim Thorson, Cole Monnohan, Aaron Berger and the entire PSTAT, have enriched my knowledge and helped me feel at home in the stock assessment community. I admire these scientists both for their intellectual achievement and their easygoing kindness. I am honored to be their colleague.

Finally, the dedicated employees of Tacoma Metropolitan Parks who maintain the running trails at Point Defiance, and the tireless volunteer community at the Tacoma City Marathon Association enabled me to maintain my health and sanity while completing this degree. I had some of my best research insights on the running trails maintained by and at races staffed by these individuals. Completing a dissertation is an ultramarathon: a challenge, a privilege, and a joy.

DEDICATION

For Alma

INTRODUCTION

Marine resources have sustained coastal populations since time immemorial. The fish of the world's oceans provide sustenance for millions of individuals and produce enormous economic benefits for both industrial and artisan economies. Many rich nations have developed sophisticated programs of scientific marine resource management, particularly for those fisheries exploited at a large scale for domestic and international production. These management enterprises concern themselves with the monitoring and quantitative assessment of fisheries stocks, producing scientific evaluations of stock status that inform limits placed on future harvesting.

The definition of a fisheries "stock" varies across nations and agencies, but loosely refers to a population of fish from which catches are taken by a fishery. This population might inhabit a broad geographic range and exhibit variation in demographic traits (e.g., growth, maturity rates); typically, the coarsest perception of a stock corresponds to the scale of management. In the mid 20th century, scientists recognized that these political delineations of stock boundaries were likely violated by ontogenetic changes in fish demography, movement, selective harvesting, and climate effects, among others. It was suspected at the time that inconsistencies between political and population boundaries would induce bias in the statistical evaluation of the stock size and subsequent management quantities. Improvements in computational power allowed researchers in the early 21st century to investigate the impacts of these spatial incongruencies using simulation modelling. The work from this period illustrated several consequences of ignoring or mis-specifying the underlying spatial structure in the assessment process, reviewed thoroughly by Punt (2019a). Of chief concern is the risk that ignoring spatial structure in the assessment process can mask localized depletion, where an individual area is exploited to a greater degree than would be ideal (or allowed) on a system-wide basis. The risk of localized depletion is higher when areas are isolated via unidirectional movement (Ying et al., 2011), are of much smaller scale than other areas (Goethel et al., 2011), and/or are less productive (Denson et al., 2016; Punt, 2019a), whether due to inherent differences in population life histories, or synergies among ontogenetic life history dynamics (Punt et al., 2015), exploitation (Langseth and Schueller, 2016), and ecosystem effects (Berger et al., 2012). Conversely, ignoring spatial stock structure may risk foregone yield, where the perception of the stock is in fact higher when spatial structure is accounted for. The

accumulation of evidence has led scientists and managers to develop spatial stock assessment models for their valuable fisheries, either as a research exercise or for tactical use (Punt et al., 2016b; Shackell et al., 2016; Szuwalski and Punt, 2015, among many others).

The first chapter of this dissertation presents a simulation study to evaluate improvement in assessment model performance with sequential elimination of a series of known mis-specifications of observation and systems model processes. The study aims to quantify the relative importance of correctly specifying spatial structure versus other population processes or parameters in an integrated stock assessment. This contributes to the set of literature that indicates how ignoring spatial structure can lead to undesirable bias in estimated values, and quantifies the value of getting spatial structure "right", in terms of reduced estimation error.

The risks mentioned above are accentuated when biological stock boundaries traverse national borders; as transboundary stocks are assumed to be declining more rapidly than those exploited by a single political entity (Palacios-Abrantes et al., 2020). These concerns have encouraged scientists to evaluate the tractability and tradeoffs in implementing spatial management strategies for exploited fishery populations. Occasionally, these efforts include signing treaties to co-manage transboundary stocks that cross international or other political borders, such as hake (*Merluccius productus*) or halibut (*Hippoglossus stenolepis*) in the Pacific. The challenges in implementing such spatial models should not be understated. Briefly, a spatial assessment model requires the definition of at least two distinct spatial areas, which may or may not be connected by movement, may or may not have distinct demographic regimes (such as growth rates) and/or recruitment paradigms, and require data to inform the removals from each area. A recent "best practice" guide for the whole of fisheries stock assessment indicates that the very first decision point in building models used for management is the specification of how the population is structured through space, if at all (Punt, 2023).

Spatially structured models are necessary, challenging to construct, and inherently less certain than their aggregated counterparts. It is therefore vital that researchers have a means to evaluate the performances of various assessment strategies, potentially including multiple hypotheses about the spatial structure of the stock, before implementing a given assessment model for tactical management. Management Strategy Evaluation (MSE) is a closed-loop simulation framework that enables researchers to do just that. A mathematical representation of the hypothesized "true"

population, called an operating model, is used to project the population forward and generate observations of that population. These projected data are subject to process and observation error, enabling researchers to capture key uncertainties in simulated datasets. These datasets are then fit using one or more estimation methods (EM), and the outputs of the EMs can be used to inform the application of a harvest control rule (HCR) to determine the catch level (either in biomass, or as an exploitation rate) for the subsequent year. A unique OM-EM-HCR combination can be considered a "management strategy": a hypothesis about the reality of the population and fishery given uncertainty (OM); a means of estimating "observed" data simulated from that OM (EM); and a means of translating the EM's perception of that stock into catch limits or harvest rates for future years (the HCR[s]). A completed MSE will contain the time-series of population information and simulated data from the operating model and the quantities estimated by the estimation method(s) for any number of projection years. Researchers can then score the relative performances of each management strategy using pre-defined criteria.

The specification of each component within the MSE can aid scientists in investigating the consequences of given management decisions. For example, researchers have used this technique to show that mis-specifying natural mortality and stock-recruitment steepness can induce variation in management strategy performance (Fay et al., 2011). Many prominent global fisheries have turned to MSE to evaluate potential responses to climate change (A'Mar et al., 2009; Jacobsen et al., 2022), inclusion of reproductive buffers in harvest control rules (Heller-Shipley et al., 2021), and the impact of ignoring spatial structure (Punt, 2019b, and Cadrin, 2020, and references therein).

In the western United States, sablefish (*Anoplopoma fimbria*) are a high-value groundfish species with a similar profile to other stocks that have undergone MSE work over the last 10 years. The US sablefish fishery is worth over $112 million USD and the industry employs 8% of the harvesting workforce in Alaska alone. At the federal level, sablefish in this region are currently managed as separate, closed populations in three separate management areas (Alaska, British Columbia and the US West Coast) by the National Oceanic and Atmospheric Administration (NOAA-US) and the Department of Fisheries and Oceans (DFO – Canada). Recent work has shown sablefish to be genetically mixed across the range (Jasonowicz et al., 2016), tagging studies have confirmed high movement rates (Hanselman et al., 2014), and there is evidence of range wide

synchrony in biomass trends, including declines during the last decade (Fenske et al., 2019). These observations have led scientists and managers to consider whether evaluating the stock as a spatially structured population might reveal transboundary dynamics, and/or preclude poor management outcomes at the regional scale.

The three main chapters (2-4) of this dissertation are concerned with developing and applying a management strategy evaluation for sablefish that 1) represents a hypothesis of the actual transboundary structure of the sablefish population, and 2) tests estimation methods encompassing a range of spatial complexity. This work involved the collaboration of an international, interdisciplinary team of scientists and the direct contribution of stakeholder (fishery) participants through multi-day workshops (Kapur et al., 2021a). Chapter 2 develops, tests, and applies a data-driven breakpoint detection method to identify key geographic areas between which sablefish growth is the most distinct, to form the basis of the putative transboundary stock structure used in the OM. Chapter 3 synthesizes movement rates from tag-recapture analyses and demographic information into a spatially structured statistical catch-at-age model, conditioned upon catch, discard, survey and age composition data from the three management areas. This model is used to project simulated populations into the future and sample survey observations from those populations, subject to fishing. The OM is the first transboundary representation of sablefish in the northeast Pacific. Chapter 4 develops a MSE for the transboundary sablefish stock, with seven management strategies explored. Five management strategies are based on spatially structured delay-difference models, which are designed to be computationally efficient while fitting simulated survey data from the OM. These EMs encompass a gradient of spatial complexity, ranging from a granular, multi-area model mimicking the OM, to a fully panmictic model in which the entire northeast Pacific is treated as a single stock. The final two management strategies do not consider an estimation method at all, and simply set annual catch limits based upon either changes in projected survey observations, or to the mean catches observed in the recent historical time series. All management strategies are evaluated using criteria developed by fishery stakeholders during a 2021 workshop, concerning the viability, sustainability, and economic value of the sablefish population through the projection period (2020-2040).

Chapter 1. GET THE SPATIAL STRUCTURE CORRECT, FIRST

1.1 Abstract

Sustainable exploitation of renewable natural resources requires quantitative scientific guidance, which has led to the expansion of the use of Integrated Population Models (IPMs). IPMs use mathematical relationships to specify changes in population abundance and to link model predictions to data. However, for many managed species, there are limited empirical estimates of key population processes or parameters, which must be specified or estimated within an IPM, or assessment, framework. A "mis-specification" is a mismatch of a certain model process or parameter between the model on which the assessment is based and reality. The consequences of mis-specification that concerns analysts are bias – where the estimation method produces quantities of management interest that differ systematically from reality – and precision, where the variances of the model outputs are too large for the estimates to be useful. Simulation is used to evaluate improvement in assessment model performance with sequential elimination of a series of known mis-specifications of observation and systems model processes. Correcting spatial structure leads to large improvements in estimation accuracy, particularly for population reproductive output. In addition, synergies between population productivity, natural mortality and size-specific mortality (due to fishing) preclude straightforward gains in estimation accuracy when any of these three processes remain uncorrected. This work encourages use of IPMs and emphasizes the importance of correctly specifying spatial structure for modeled populations, and how to prioritize the correct estimation of various model parameters depending on the estimated quantity of interest.

1.2 Introduction

Sustainable exploitation or the management of natural resources requires sound scientific guidance. To this end, integrated population dynamics models (e.g., Fournier and Archibald (1982)) have flourished as the tool of choice to evaluate the status and possible future outcomes for exploited, threatened or managed populations (Maunder and Punt, 2014; Tempel et al., 2014). Integrated population models use mathematical relationships (processes) to specify how changes in population abundance occur as well as to link model predictions to data (observations). The processes are themselves governed by parameters that can either be estimated during the model fitting process or pre-specified based on independent studies. In the United States and many other

developed nations, integrated population dynamics modelling has enjoyed a long history in fisheries management (typically referred to as "stock assessments"), whereas this practice has only more recently spread to terrestrial and marine wildlife applications, there known as Integrated Population Modeling (IPM) (Schaub et al., 2007). The difference in funding and research years dedicated to these fields has led to several discrepancies between fisheries assessment and other IPM studies in the construction and evaluation of models. For well-funded management agencies, fisheries stock assessment is typically performed using standardized, peer-reviewed software (e.g., Stock Synthesis, Methot and Wetzel, 2013), maintained by a team of paid scientists. These software packages often have built-in model diagnostics and require many model specifications. Finally, the scientific management advice for commercial fisheries relies on "reference points", which are calculated quantities mandated by law, and used to indicate the status of a commercially exploited stock in a given year (Gabriel and Mace, 1999).

The state of IPMs is wholly different: custom models are generally built as one-off endeavors for a specific study (which could involve one species at a given site), often characterized by sparse data or little life history information. There is a general shortage of general tools for model validation or testing beyond parameter sensitivities performed for the study at hand because few IPMs get re-used for subsequent analyses. For example, Eacker et al. (2017) conducted a Bayesian life-stage simulation analysis for elk (*Cervus canadensis*) in Montana and completed a very detailed sensitivity analysis of population status to demographic parameters. That study concluded that because parameter estimates from their IPM were similar to vital-rate-only models, the population parameters (vital rates) must have been most influential to their study – an interesting finding, yet challenging to verify without simulation testing the original model or applying it to other species. For example, it could have been that their priors, which were based on pre-existing vital rate studies, were simply more informative in the Bayesian model than the monitoring data. Analyses for similar species (woodland caribou, Johnson et al. (2019), and boreal caribou, Rudolph et al. (2017)) involved essentially building new IPMs and precluded what could have become a validation of that finding, even though Eacker et al. (2017) have published their IPM as a web-based tool. Finally, because IPMs are deployed to answer specific research questions, there is not a formalized "reference point" system to compare outcomes among various models or to support management actions.

Despite these differences, scientists conducting fisheries stock assessment and IPMs are unified in their goal to estimate quantities relevant to the dynamics of biomass (or individuals) in the population through time. This is accomplished by either specifying or estimating the parameters governing those processes that link data to the population dynamics (Gauthier et al., 2007; Maunder and Piner, 2015). In fisheries, a model provides the basis to reasonably predict the effects of future fishing when the production function is accurately represented (Conn et al., 2010). Accurate estimation of the population dynamic processes only occurs when the model is correctly specified, which can be challenging if the underlying population dynamics have, for example, spatial processes not accounted for in the model – an observation made many years ago by fisheries scientists (Schaefer, 1968) and continually revisited (Booth, 2000; Fay et al., 2011; Punt, 2019b). Indeed, IPMs have more rapidly become spatially explicit to account for, for example, transboundary movement in bears (Chandler and Clark, 2014), koalas (Rhodes et al., 2011) and birds (Pacifici et al., 2017), although funding remains an ongoing and limiting factor in data collection for managed or at-risk species. For the biological and observation processes within spatial units, the decision to pre-specify or estimate a parameter often depends on which biological processes have been well studied or which parameters are thought to be estimable within an integrated model (Lee et al., 2012, 2011). Simulation work in fisheries has suggested that spatial approaches, when applied appropriately, will reduce bias and increase precision of estimated quantities (Goethel and Berger, 2016). It is useful to evaluate the relative importance of correctly specifying population processes in the context of spatial variability to population estimation considering these advances in both fields, and the paucity of information to inform other model processes for most IPMs.

A well-designed approach to model specification is needed to ensure that the assessment outcomes accurately represent the system. Model specification decisions are related to the functional form of the process, the variables they depend on (e.g., age or length, Lee et al. (2019)) and potential temporal variability in those processes. Inappropriate specification of a population dynamics model can occur in the observation (data) and/or population processes (Maunder and Piner, 2017). In contrast to population processes, the parameters of the observation model are nearly always estimated because they address sampling uncertainties that are largely unknowable without an estimate of the population dynamics. Mis-specification occurs when a process is assumed to be governed by the wrong functional form, a parameter is set to the wrong value, or a process is

modelled such that temporal variability is not correctly accounted for (or even ignored). Mis-specification of the population processes can lead to biased estimates of the parameters and hence quantities of management interest while mis-specification of observation processes can lead to the data not providing the correct information about the estimated parameters (e.g., Langseth et al., 2016; Maunder et al., 2023). Moreover, mis-specification in one process can lead to poor fits to data directly linked to that process and to data indirectly linked via the population dynamics because all data and model processes are linked through the population dynamics equations (Lee et al., 2019; Taylor et al., 2013). The linkage of all data via the population dynamics equations is the strength of integrated modelling, but this strength also makes locating mis-specifications challenging.

Diagnostics (e.g., Carvalho et al., 2021) are used to detect model mis-specification. Work in fisheries has shown that candidate diagnostic tests are not equally reliable at detecting model mis-specification (Carvalho et al., 2017). Some diagnostics are good at detecting mis-specification of observation components while far fewer appear to be good at identifying the more severe problem of mis-specification of system processes. It is not clear if any diagnostics are good at isolating the process that is mis-specified or the exact nature of that mis-specification for complicated models, such as those with many fisheries and types of data. Much of the work that has tested model diagnostics has evaluated only a single source of mis-specification in relatively simple models (e.g., Carvalho et al. (2017); Piner et al. (2011)). Almost certainly, diagnosing and correcting model mis-specification becomes more difficult when multiple processes are simultaneously mis-specified.

Researchers constructing IPMs hope that each successive version of a model reduces the number and severity of mis-specifications, and gradually improves model performance, typically by identifying and implementing more appropriate model components. This ‘sequential elimination’ approach is a hypothesis about model development, and it is unclear which components should be addressed first. Ideally, the most critical mis-specifications are corrected in the earliest versions, while later versions gradually ‘sneak up’ on a well-specified model.

The present study implements the sequential elimination of mis-specified processes leading to a gradual improvement in model performance in models subject to multiple sources of mis-specification, specifically in the context of spatial variation. Model performance is measured by

quantities that describe the accuracy of the estimates of the population biomass and reproductive output; to emphasize the ability to translate the results to species other than fishes, comparison amongst management reference quantities specific to fisheries is avoided. The most critical processes to focus on are determined, and whether correcting such processes in the face of incorrect spatial structure is detrimental to model accuracy. This study is based on a common fishery stock assessment modeling package because the model's structure is well-documented, it runs quickly, and is comprised of population and observation components that encompass those used in most recent IPMs (Abadi et al., 2010). The results of this work will inform priorities for how model specification and evaluation can be addressed for integrated population dynamics models that use similar components. One hypothesis is that gradual improvement of a model with incorrect spatial structure assumptions will reach a point of diminishing returns in relative error, until the spatial structure of the model is correct. The order of corrections that yields the best outcome emphasizes the value of getting certain aspects of the model right, which should be of great interest to reviewers, review panels and management agencies with limited funding.

1.3 Methods

1.3.1 Overview

Simulation is used to evaluate the improvement in the performance of an integrated population model resulting from the sequential correction of a series of model mis-specifications. The sequences, here called "strings", fall into one of two "Experiments", which differ in whether the spatial structure is corrected first ("Experiment 1: Space First") or last ("Experiment 2: Space Last"). The simulation procedure first involves the specification of a model of true population dynamics, the operating model (OM). The OM is used to generate typical data for a fish population (time-series of catches in weight from a fishery fleet, an index of abundance from a survey for each of two areas, and the proportions-at-length for both the fishery and surveys, also known as the composition data). All three types of data are generated from the OM for "1980-2016", with a period of early recruitment deviations extending for 10 years prior to "1980" (Figure A.1). These generated data are used in a set of estimation methods (EMs). The EMs fit to the data and estimate the quantities of management interest. These estimates are then compared to the true values from the OM.

The first EM for both Experiments contains five mis-specifications, including spatial structure. In Experiment 1: Space First, the second EM corrects the spatial structure, whereas in Experiment 2: Space Last the second EM randomly eliminates one of the other mis-specifications. The simulation procedure continues by sequentially removing each of the remaining mis-specifications until the last EM is well-specified ("mis-specification free"). In Experiment 2, the final mis-specification to be corrected is spatial structure; for Experiment 1 the last mis-specification to be corrected varies among the other four mis-specifications. A total of 24 unique combinations, each involving six models (five corrected in a unique order, one correct) are used to describe the effects of the gradual reduction in mis-specification and impact of eliminating different types of mis-specifications. The number of combinations (24) is based on the number of unique ways to order four mis-specifications, with space either first or last. Detailed descriptions of the OM, EMs, types of mis-specifications, and the performance metrics are given below.

1.3.2 Operating Model

The OM is an age-structured population dynamics model implemented in the Stock Synthesis software (SS version 3.24z, Methot and Wetzel (2013)). Key systems and observation processes and their parameter values are listed in Table 1.1, and some biological assumptions (e.g., stock-recruitment steepness, natural mortality, and growth) were originally estimated for Pacific Hake (*Merluccius productus*, Taylor et al. (2015)). A general description of the OM is as follows: it is a two-area model with time-varying uni-directional age-based movement (Figure 1.1d) that leads to an area of primarily juveniles (area 1) and an area of primarily adults (area 2). The population is single-sex, with time-invariant length-weight and maturity-at-age relationships, natural mortality (M), and length-at-age. Recruitment is assumed to follow a time-invariant Beverton and Holt (1957) relationship with steepness (expected recruitment at 20% of the expected pre-fishery biomass, h) fixed at 0.86 and randomly-generated stock-recruitment deviations with a standard deviation (σ_R) of 1.4 in log space. The observation process involves a single fishing fleet operating in area 2 (adult area) and two survey fleets, one in each area. The relative probability of capture by length (selectivity) for the fishery fleet and surveys is time-invariant. All ages are available to the survey and fishery fleets. Annual fishing mortality rates for all OMs were generated for each simulated population using the functional form described in Appendix A of Carruthers et al. (2012), which allows fishing mortality to vary randomly about an increasing trend between 1980

and 2000, and randomly thereafter given an overall mean fishing mortality rate equal to the rate of natural mortality (Figure A.2). The initial conditions were specified so that there was no impact of fishing prior to the first year. Process error in each OM replicate arises from sampling $\tilde{R}'_y \sim N(0, \sigma_R^2)$, where $\tilde{R}'_y$ is the recruitment deviate, and σ_R is the variance of annual recruitment deviates (Figure A.1).

Data used in the EMs are the time-series of catches in weight from the fishing fleet, a time series of relative abundance from the survey, and length-composition data that provide a measure of the size structure of the index in each area and fishery (Figure A.3; see below for a description of how spatial data were aggregated for use in single-area EMs). The catch observations were assumed to be known with minimal error, as is reasonable for well-monitored fisheries. Each abundance observation was assumed to be proportional to the available absolute abundance, called "catchability" in fisheries applications, and was generated using a log-normal distribution with a coefficient of variation of 0.2 (Figure A.4). Each length-composition observation was generated from a multinomial distribution with variability described by an effective sample size of 50. The values for the parameters for data generation follow the operating model of Lee et al. (2019).

The model components that were manipulated in the simulation experiments and how the mis-specifications were implemented are described below.

1.3.3 Mis-specified Processes

1.3.3.1 Spatial Structure

The OM is a two-area model that is fished by a fishery in the adult area and surveyed by a survey in each area. Each fishery and survey produce a length-composition time series, and each survey produces an index of relative abundance (Table 1.1). Movement increases a logistic function of age (Figure 1.1d). Estimation methods with spatial structure mis-specified require the data (catches, survey abundance indices, and compositions for each area) to be "collapsed" into those for a single area, leading to a single-area model with a single survey index and a length-composition time series for the survey and fishery. There are many reasons why spatial structure may be mis-specified; an incorrect configuration such as that represented in this study could be a misunderstanding of the underlying demographic variation, a political requirement to manage a harvested population as a single stock, or simply not thinking about spatial structure, among others.

Converting survey index information from each area and year $I_{y,A}$ into a single index is achieved using a summation across areas among years:

$$I_y = q \sum_{A=1}^{2} \frac{I_{y,A}}{q_A} \qquad (1.1)$$

where q_A is the catchability coefficient for a survey in the area A, and q is the overall catchability coefficient for each survey (fixed to 1.0).

Length compositions are annual vectors of the proportion of fish observed in each length class, of which there are 23 (2-cm increments from 8cm to 52cm). For the spatially structured OMs, there are proportions for all ages through the plus group for both surveys (one in each area). To combine these for the single-area (mis-specified) EMs, the area-specific proportions $P_{l,y,A}$ are weighted by the total model-estimated number in that area in that year $N_{y,A}$, which is calculated within the OM (see the Stock Synthesis Manual for technical details, Methot and Wetzel 2013). These weighted values are then summed across areas for each year and divided by the total number observed in both areas, returning the proportions-at-age for a single survey.

$$P_{l,y} = \frac{\sum_{A=1}^{2} P_{l,y,A} N_{y,A}}{\sum_{A=1}^{2} N_{y,A}} \qquad (1.2)$$

EMs with spatial structure mis-specified are denoted by the letter S.

1.3.3.2 Growth

The growth curve in the OM is modeled using the von Bertalanffy (1957) growth function, a common relationship used in fisheries assessment to model the length (cm) of an average fish with respect to its age (years). The model is parameterized using asymptotic length (L_∞, the inferred length at infinite age) and the growth rate K (the rate at which the average fish reaches asymptotic length). As in many IPMs, including those for mammals (Regehr et al., 2018) and reptiles (Eaton and Link, 2011), assumptions about the growth rate of individuals can be influential in determining the population growth rate, as many animals must meet a minimum size before reaching reproductive maturity. Researchers may obtain inaccurate input values of this parameter via unrepresentative or imprecise sampling, which fails to capture or correctly measure individuals at large lengths and/or older ages (Shelton and Mangel, 2012). In fisheries, this can occur because

older, larger fish reside at depths outside the range of the gear used to conduct surveys (Taylor and Methot, 2013). Furthermore, the relationship between fish length and weight is typically governed by an allometric relationship, such that incorrect estimates of average fish length would lead to incorrect weight estimates resulting in a bias in the estimated total biomass. The mis-specification examined assuming L_∞ values different from those in the OM; growth parameters ($L_{age=1}$, L_∞, K, and CVs of length at ages) are estimated when this process is not mis-specified (Figure 1.1a). The values of the other growth parameters (($L_{age=1}$, K, and CVs of length at ages) are estimated when L_∞ is mis-specified. Estimation methods with L_∞ mis-specified are denoted by the letter L.

1.3.3.3 Natural Mortality

Natural mortality, the instantaneous loss of population numbers due to causes other than fishing is time- and age-invariant and set at 0.2 yr^{-1} (Table 1.1). In IPMs, the analogous process is often survivorship (which can be stage-specific); there were insufficient empirical data to parameterize age-specific natural mortality for the case study species. Generally, it is difficult to obtain empirical estimates of natural mortality for any fish species (e.g., Hamel, 2014; Punt et al., 2021; Maunder et al., 2023). In fisheries, several methods have been developed that infer this value from the maximum age or length (Then et al., 2015) or via a meta-analysis of similar species within a genus (Thorson et al., 2017). Investigations of generalized IPMs have found heterogeneity in the mortality process (natural or otherwise) to lead to substantial differences in population parameter estimates (e.g., Riecke et al. (2019)), and emphasize that uncertainty around natural mortality is likely a result of biological processes and observation error. Estimation methods with natural mortality mis-specified at incorrect values are denoted by the letter M.

1.3.3.4 Reproduction

The regenerative capacity of a population influences its ability to recover from exploitation. In fisheries, the common measure of stock resilience is the steepness of the stock-recruitment relationship. Annual reproduction R in the OM is calculated based on a Beverton-Holt function (Equation 1.3) of the system-wide reproductive biomass in a given year (SB), expected unfished recruitment R_0 and biomass SB_0 and h, i.e.:

$$R_y = \frac{4hR_0SB_y}{SB_0(1-h) + SB_y(5h-1)} e^{-0.5\sigma_R^2 + \tilde{R}_y} \; ; \tilde{R}_y \;\; N(0, \sigma_R^2) \qquad (1.3)$$

Annual recruitment deviates $\tilde{R}_y$, governed by σ_R (set at 1.4), measure the log-distance from the deterministic curve given by Equation 1.3 and is a source of process error in the OM. Therefore, these recruitment deviates are randomly generated once for each OM replicate to reflect the process variability, but the steepness and R_0 are fixed (Table 1.1). In the estimation methods, the recruitment deviates and R_0 are estimated, with steepness fixed at either the correct or a mis-specified value. Estimation methods with steepness mis-specified are denoted by the letter H (Table 1.1).

1.3.3.5 Fishery selectivity

In fisheries assessment models, "selectivity" refers to the proportional length- or age-based availability of fish to surveys or fisheries. The type of gear used, the timing of the survey, or/and depths fished can all contribute to a preferential sampling of a given age- or size-classes, which is reflected in the compositional data collected during monitoring. In the IPM literature, concern regarding such preferential sampling is typically accounted for in the observation model, such as via stage-specific detection probabilities (Schaub and Abadi, 2011). The fishery, which only operates in the adult area, and the surveys in each area, all have a length-based asymptotic selectivity pattern in the OM, meaning that all individuals above a certain size have a close to equal probability of being captured (Table 1.1; Figure 1.1b,c,e,f). However, it is possible to assume a "dome shaped" selectivity curve, which presumes that the gear and/or survey design selectively targets an intermediate range of fish sizes (or ages), with declining proportions captured both above and below an ideal size. Analysts may mis-specify the form of the selectivity curve because they mistakenly assume certain sizes/ages of animals are less preferentially sampled, particularly those above a given size. This could be caused by a faulty assumption that older/larger individuals are less abundant in the surveyed area, or have learned to avoid or escape the sampling gear more frequently than younger/smaller individuals (Ahrestani et al., 2017; Schaub and Abadi, 2011). When the selectivity is correct, estimation methods estimate the selectivity parameters under the assumption that selectivity is asymptotic function of length for the fishery and survey(s). Estimation methods with selectivity mis-specified are denoted by the letter X, indicating that parameters for a dome-shaped (double normal) selectivity curve for the fishery fleet are estimated while the selectivity curve for the survey(s) is correctly assumed to be asymptotic; in the case that

spatial structure is mis-specified as well, estimation proceeds with the incorrect selectivity form for a single fishery and survey.

1.3.4 Determining mis-specification thresholds

Instead of arbitrarily choosing mis-specified parameter values, the two nearest value(s) that led to a 20% change in the final-year depletion (the ratio between final-year biomass and expected unfished biomass) were solved for. This detection protocol was applied to mis-specifications in parameter values only (steepness, mortality, and L_∞). The threshold detection was performed by fitting a series of estimation methods to the same 2-area dataset generated by the OM across a broad range of fixed values for each parameter in turn: for example, 19 estimation methods with steepness *h* fixed at 0.05, 0.10, 0.15,…,0.95 and other parameter values estimated. The relative error in final-year depletion was calculated between the estimates from each estimation method and the values in the OM and used to find the parameter values nearest to the OM values that corresponded to relative errors of -20% and 20%. This step ensured that the mis-specifications implemented in the experimental design are known to impact estimated outputs to the same extent. This led to two mis-specified parameter values, one above and one below the values used in the OM, for all parameters except for steepness. In the case of steepness, no values above the OM value met the mis-specification threshold criteria; instead, two values below that in the OM corresponding to changes in final depletion of -20% and -10% were selected. Selectivity was mis-specified by enabling the estimation of a descending limb (approaching dome-shaped, instead of asymptotic as in the OM). The spatial structure mis-specification was not treated as a gradient, so threshold levels of aggregation for that category were not determined (models were either spatially correct (2 areas) or mis-specified (one area)).

1.3.5 Estimation methods and experimental design

The EMs were implemented in Stock Synthesis version 3.24 (Methot and Wetzel, 2013). The experimental design followed a systematic procedure (Figures 1.2 and 1.3), which enabled an evaluation of the benefits of correcting mis-specifications besides spatial structure versus correctly specifying all processes and spatial structure, either from the outset (Experiment 1) or as a final correction (Experiment 2). The mis-specifications examined are detailed in the fourth column of

Table 1.1. A simulation protocol was used for both experiments. The experimental workflow was as follows:

1. Generate an operating model "replicate" with process errors (recruitment deviations and fishing mortalities) and observation errors (generation of survey abundance indices and compositional data).

2. Sample a vector of three values for each replicate, each with even probability of being either a 0 or 1. This vector determines how each mis-specification is implemented. For example, the first OM replicate may have the draw [0, 0, 0] in which all three parameters would be specified below the true value for all EMs fit to those OM data. The next OM replicate may have a different vector draw, ensuring that variation caused by differences in process and observation errors are balanced against the directionality of mis-specifications.

3. Fit EMs for each of the 24 "strings" corresponding to the mis-specified categories (Figure 1.2) to each replicate. Each string sequentially eliminates one letter at a time. For the string "LHXMS" in Experiment 2, EM2 will correct the L_∞ (L) with steepness (H), selectivity (X), natural mortality (M), and space (S) mis-specified; EM3 will correct the L_∞ (L) and steepness (H) with X, M, and S mis-specified, and so on. EM5 will have only S mis-specified, and EM6 will have all components correct, using the correctly-stratified data from the corresponding OM replicate (labeled "correct").

4. Repeat step 3 for all possible strings (24) and two Experiments, each fitting to data from the same OM replicate.

5. Repeat steps 1-4 for a total of fifty resampled OM replicates.

This protocol ensures the effect of the mis-specifications was not influenced by the high/low nature of the random vector assigned to each string. In total, the study design fit 14,440 EMs, 7,200 for each Experiment (6 estimation methods for each of 24 strings and each fitted to 50 OM replicates).

Convergence was evaluated for each estimation method by whether the Hessian matrix was invertible, and the final gradient (slope of the objective function at the maximum likelihood estimate). This step was taken to avoid comparing the performances of models that analysts would normally discard from consideration, regardless of their belief in the accuracy of the specification.

Because our analysis relied on the comparison of averages across OM replicates and EM combinations, balance was retained across simulations by identifying and removing OM replicates for which the mean gradient (across all 24 strings fit to that replicate) was above a threshold (0.0025), which resulted in the removal of 13 out of 50 OM replicates. Individual EM runs that returned a non-invertible Hessian matrix were removed after this step. Figures and tables only display results for model runs that passed these filters.

1.3.6 Performance measures

The results were summarized by the deviation between the estimates of the management quantities and the corresponding OM values. In lieu of fisheries-specific management quantities (e.g., the ratio of current biomass to the biomass that corresponds to maximum sustainable yield), values common across many IPMs were examined, namely the time series of reproductive biomass (here, spawning stock biomass, SSB) and reproductive output (here, recruitment). In addition to the general trend in these estimated values, results based on the mean SSB over the last ten years were evaluated. Together, these statistics aim to capture temporal variation in estimation performance as well as model performance during the recent period, which is typically of more interest to managers. The deviations between the EM and OM values by year, replicate, string, and model were summarized using relative or absolute relative errors and then averaged across replicates (Equations (1.4) and (1.5)). These measures indicate the magnitude of difference between estimated quantities and the OM values. Relative error (positive or negative) enables an investigation of whether there are systematic and/or directional biases induced by the various mis-specifications. Using absolute relative error disregards the direction of the difference, and is useful for highlighting the scale of the effects of various mis-specifications. The mean relative (MRE) and mean absolute relative (MARE) errors for SSB are calculated via:

$$MRE_{SSB_y} = \sum_i \frac{\hat{S}SB_y^{EM_{ijk}} - SSB_y^{OM}i}{SSB_y^{OM}i} / 50 \qquad (1.4)$$

$$MARE_{SSB_y} = \sum_i \frac{\hat{S}SB_y^{EM_{ijk}} - SSB_y^{OM}i}{SSB_y^{OM}i} / 50 \qquad (1.5)$$

where i indexes OM replicate ($i = 1,2, \ldots ,50$[1]), j indexes model process ($j = 1,2, \ldots ,6$), k indexes EM string ($k = 1,2, \ldots ,24$), and y indexes year ($y = 1980,1981, \ldots ,2016$). The same formula was used for recruitment. To summarize the results, the statistics for all EMs were calculated that 1) share the same number of mis-specifications, or 2) have the identical correction present first (such as two strings from Experiment 2: Space Last, which both correct natural mortality first), or 3) have the identical corrections present in a given order (such as two EMs from Experiment 2 that corrected selectivity and steepness first).

1.4 Results

1.4.1 Generated data

The mean biomass trend across OM replicates for the simulated species descends from an average high of 3.8 million mt in 1980 to an average low of 670 thousand mt in 2016 (Figure A.4) due to the trend in fishing mortality rate. The operating model introduced sufficient noise into datasets (e.g., Figure A.4) and hence variation in relative error. The filtration step (based on invertibility of the Hessian and mean gradient across OMs) resulted in the discarding of results for 13 of the 50 replicates from both experiments. Most of the EMs that produced non-invertible Hessian matrices were highly mis-specified, with 80% and 63% of failures occurring in fully-mis-specified EMs from Experiments 1 and 2 respectively.

1.4.2 Error across mis-specified models

Spatial structure was corrected first in Experiment 1. A substantial improvement (reduction) in relative errors in SSB and recruitment occurred when spatial structure was corrected first (Figure 1.4; Figure A.6). The relative errors continued to decline with decreasing numbers of mis-specifications. Relative errors in model estimates of SSB were the smallest and least variable through time for correctly specified EMs (Figure 1.5). Most EMs with four or fewer mis-specifications had MREs for SSB at or near zero after 2000, but not for recruitment. The average absolute relative errors for the last 10 years of SSB and recruitments were drastically reduced with spatial structure corrected first (Figure 1.4).

[1] While 50 OM replicates were generated for this study, the final number used to generate results was 33 after excluding replicates that did not achieve an invertible hessian matrix.

Spatial structure was corrected last in in Experiment 2. The relative errors in SSB and recruitment are reduced by half when the spatial structure was corrected (Figures 1.5 and A.7). Overall, relative errors in model estimates of SSB were the smallest and least variable through time for correct EMs. Errors in estimating recruitment were quite extensive in Experiment 2 until spatial structure was corrected, producing many "spikes" during the time series because of the true value being very small (Figure A.9). Most mis-specified EMs in Experiment 2 had low or near-zero MRE for SSB after "2000", although models with spatial structure mis-specified wrongly estimated an increase followed by a sharp decline in SSB during the onset of a period with low catches during the late "1990"s. It is possible that, in the absence of spatial structure, the EM is unable to capture the age-based dynamics (movement) and instead relies on the length-based observations (via selectivity), which is mis-specified and/or poorly informed due to lack of data. The correction of the spatial structure led to a marked improvement (reduction) in the MARE for the last 10 years of SSB and recruitments (Figures 1.6 and A.9) and eliminated this phenomenon.

Estimation performance was considered to "improve" when the MARE decreased from one group of identically structured EMs to the next. The most dramatic improvement in estimation performance for SSB and recruitment occurred in both experiments when spatial structure was corrected (Figures 1.4, 1.6, 1.7, A.6, and A.7). On average, correcting spatial structure in Experiment 1 reduced MARE during the last 10 years of SSB by 82%, and reduced MARE during the last 10 years of recruitment by 65%. Correctly specifying spatial structure reduced error for every replicate. For comparison, the next-most influential parameters in Experiment 1 were steepness (14% improvement in SSB, on average) and natural mortality (8% improvement in the MARE for SSB, on average). Correcting natural mortality alone had a similar average improvement of 9% in MARE SSB for EMs in Experiment 2.

1.4.3 Error across corrected features

There were several emergent patterns across both experiments. First, correction of spatial structure consistently led to improvements in model accuracy as measured by the MARE in the final 10 years of SSB and recruitment, and led to the largest reduction in MARE. After spatial structure, natural mortality (*M)* had a systematic and marked effect on the direction and magnitude of error in recruitment estimates, producing estimates that were consistently either too high (if *M* was mis-

specified high) or too low, even if the spatial structure was correct (Figure 1.7). This effect is present to a lesser degree in estimates of SSB (Figure A.10).

There was not a clear, singular driver of the MRE in SSB aside from spatial structure. There appear to be synergistic effects among certain mis-specifications (steepness, selectivity and asymptotic length) across the two experiments. Correcting only one of these three while the other two parameters remain mis-specified was insufficient to systematically improve estimation of the population's reproductivity (recruitment) or spawning biomass. In other words, there was no order to the correction of these components that consistently and systematically improved performance measures across all simulations. Instead, the apparent outliers – the worst performing models, even with only two or three mis-specifications not necessarily including spatial structure – consistently include some combination of these components mis-specified. This is illustrated by the prevalence of the letters H, X and L in the labels of darker (less accurate) estimation methods in Figures 1.4 and 1.6.

1.5 Discussion

1.5.1 The critical nature of spatial structure in IPMs

The rise of IPM use in realms outside of fisheries presents an opportunity to investigate the influence of major population parameters in the context of spatial structure on estimation accuracy. This study was designed to present a minimal set of illustrative simulations to depict whether there are non-linear gains in estimation accuracy from correcting mis-specifications in a given order, and whether correcting spatial structure systematically improves an otherwise well-specified model. There are many ways in which spatial structure could be mis-specified. This study was designed to represent the most simple case, where ontogenetic movement proceeds in a single direction. Analyses of more sophisticated or complex spatial frameworks have found similarly strong biases induced by spatial mis-specification, particularly in the case where a reproductive unit of the population is over-exploited (Okamoto et al., 2018). The mechanism by which a spatially-mis-specified estimation method fails to capture the population dynamics is that the spatial components of the model (namely movement between areas) is age-based. In the spatially mis-specified EM, the only means for the observation model to account for these observed differences is via length-based selectivity; if both movement and selectivity were length-based, it is possible the EM could better handle both gear preference as well as the true underlying

availability (c.f. Taylor and Methot, 2013). As anticipated, correcting spatial structure in estimation methods led to the greatest reduction in error when estimating recruitment and population biomass. However, it was surprising that this positive effect was not diminished even if all other model components were correct, meaning that correcting this aspect of the model should be the high priority.

The generalizability of our findings are reinforced by previous work in both the fisheries and terrestrial IPM literature indicating that spatial structure is highly influential for estimation of population quantities (e.g. Chandler and Clark, 2014; Thorson et al., 2015a); the present study provides a more nuanced understanding of the degree of bias reduction associated with getting spatial structure "right" in addition to, or in the absence of, other key population processes. A secondary mechanism by which spatial structure exerts high influence on model outcomes is the fact that, unlike a simple change in a parameter value, introducing spatial structure modifies data inputs and the dimensions of the likelihood equations. Given this, it is understandable how mis-specifying spatial structure leads to large bias in study systems including and beyond fisheries. An extension of this work could include a meta-analysis evaluating the relative impact of spatial structure across a suite of life histories. For example, spatial structure should be less crucial for highly-mixed populations, or for species whose life history strategies lead to reduced spatial variation in state variables.

1.5.2 Synergies across performance metrics

Several emergent patterns among, and synergies between, modeled processes indicate that the prioritization of model components after spatial structure may depend on the question at hand. The two experiments illustrated the systematic influence of natural mortality on the magnitude and direction of estimation bias, regardless of spatial structure. The strength of this outcome was most pronounced for recruitment. This adds nuance to the finding of Thorson et al. (2015a), who suggested that spatial structure is necessary to appropriately quantify density dependence: in fact, only getting spatial structure correct leaves the model vulnerable to the influence of mis-specified natural mortality, which will universally dictate the direction of bias in recruitment estimates. Finally, both experiments led to the surprising finding that solely correcting steepness, selectivity or asymptotic length (with or without correct spatial structure) did not lead to systematic improvements in the estimation of recruitment. This result suggests that these parameters must be

correct in addition to other key informative processes (natural mortality, asymptotic length) before a positive effect is realized.

Two management quantities were selected, population spawning biomass and recruitment (a measure of stock productivity) because they capture different aspects of quantities on which management decisions could be based. Spawning biomass is a measure of the mature stock present each year and depends on the number of fish born and the rate at which they are removed at size, either due to natural causes or the fishery. It is recommended that these results be used to underscore the development of spatially-correct IPMs, and encourage research regarding the natural mortality of exploited populations, which is an active area of research even for well-studied species (Brodziak et al., 2011; Piner et al., 2011). Enabling accurate estimation of natural mortality is particularly important if IPM developers wish to accurately capture recent productivity. Finally, findings from Experiment 2: Space Last indicate that parameters that directly inform the stock's productivity and size-at-capture (steepness, asymptotic length and fishery selectivity) are best corrected in concert; though correcting the growth parameter (asymptotic length) led to the greatest accuracy improvements of the three.

1.5.3 Priorities for model builders

We recognize that "correcting spatial structure" is a large undertaking, and not nearly as straightforward as, for example, improving the precision of estimates pertaining to growth. Identifying spatial structure may involve, at minimum, the use of mark-recapture, genetic, and spatially explicit survey data (Cadrin et al., 2019, Cadrin, 2020). One possibility is that individual fisheries could evaluate whether the gains in relative error identified here warrant the expense associated with this endeavor. This sort of decision analysis must be made on a per-species basis and likely predicated on the value (social, economic, or otherwise) of the stock (or population) being assessed. For example, researchers could define the fiscal value to the fishery and market of a given animal biomass, and approximate the financial gain of reducing error in their specific assessment by getting a better sense of the population's spatial structure. However, for many species that are not exploited, or low-value artisanal fisheries, it is possible that the cost of doing so is much greater than the present annual value. In this case, researchers are encouraged to explore other modeling methods that approximate spatial structure, such as spatial delay-difference models (Thorson et al., 2015a). Alternatively, the population's value may be articulated in terms of social

or ecological terms, and such research carried out from the conservation perspective, has been done for marine mammals (Johnson et al., 2019; Punt and Donovan, 2007).

Similarly, the ability to identify natural mortality (or survivorship) "in some other way than by subtraction" (Larkin, 1977) has been a persistent challenge for population modelers. Our work emphasizes that the directional bias of recruitment estimates can be predicted by whether *M* was assumed to be smaller or greater than its true value, and additionally illustrated that this bias is only corrected when both spatial structure and natural mortality are correct. For that reason, managers may wish to forestall improving *M* estimation in favor of investigating spatial structure, as tools exist for addressing the latter.

1.5.4 Future directions

We summarized our results quite straightforwardly (calculation and averaging of relative error between estimates from estimation methods and true values from operating models). The analytical framework is kept simple to ease interpretation for scientists from a variety of fields and to highlight the most readily obvious trends. However, the results produced in this work would be amenable to exploration using more sophisticated optimization or machine learning algorithms, which could be employed to identify patterns among mis-specified models less apparent than in the present analysis. This is a direction the author intends to explore in future work.

Nothing can take the place of a good sensitivity analysis. Integrated population models, particularly those used for management purposes, must necessarily investigate the impacts of input parameter uncertainty on derived quantities. An emerging approach in fisheries assessment, which borrows from climatological modeling, is the "ensemble" approach. While a simple sensitivity analysis will capture model outcomes across a range of input parameter values, the ensemble is a collection of models representing the spectrum of uncertainty in many aspects, which could include spatial structure, observation model, likelihood components, and data inclusion (Ianelli et al., 2016). Our EMs fit to data from the same OM replicate can be considered as candidate models and could be evaluated against one another using traditional model-selection techniques (such as AIC). This approach could be extended to include spatial EMs with alternative assumptions about movement (which is often difficult to estimate), or a proxy thereof such as time-varying selectivity. It has been noted that spatial models may struggle to reliably distinguish movement from

selectivity (Sampson, 2014); in an ensemble context, the change in relative error in EMs from Experiment 1: Space First where selectivity was corrected second-to-last could be weighed against other models in those “strings” to argue for or against the assumption of logistic selectivity from the outset. Examinations of the tradeoff between the loss in flexibility by fixing asymptotic selectivity in the context of spatial structure and other mis-specifications are warranted.

1.6 Tables

Table 1.1: Key operating model assumptions (central column), which represent the true population dynamics. Simulated data included catch, survey index, and length-composition information. Annual recruitment deviations and fishing mortality values were drawn from the relevant distribution for each year of the simulation. When parameters are mis-specified they are fixed at the value specified in the rightmost column; A ditto mark (") indicates the same parameterization as in the correct EM (the rightmost column). Adapted from Lee et al. (2019).

	Process	**Letter code when mis-specified in EM**	**Value in OM**	**Treatment in EM when mis-specified**	**Treatment in EM when correctly specified**
Growth					
	Maximum Age		15 years	"	Fixed (15)
	Length at age 1 (L_1, cm)		12	"	Estimated with bounds [2, 16]
	CV of length at age 0		0.1	"	Estimated with bounds [0.03, 0.25]
	Asymptotic length (L_∞ ,cm)	L	53.4	42.5 or 61.3	Estimated with bounds [40, 70]
	CV of length at age ∞		0.1	"	Estimated with bounds [0.03, 0.25]
	Growth coefficient (K; yr^{-1})		0.25	"	Estimated with bounds [0.2, 0.4]
Movement					
	Fraction of fish moving from area 1 to area 2 at age 0 (logistic function increases with age; Figure 1.1d)	S	0.2	Not present when spatial structure mis-specified	Fixed (0.2)
	Fraction of fish moving from area 1 to area 2 at age 7 and older	S	0.999	Not present when spatial structure mis-specified	Fixed (0.999)
Beverton-Holt stock-recruitment relationship					
	Log unfished recruitment ln(R_0)		9.62	"	Estimated with bounds [13, 17]

	Standard deviation for recruitment in log-space (σ_R)		1.4	"	Fixed (1.4)
	Annual recruitment deviations in log space ("1970"–"2016")		Temporal process follows Normal (0, σ_R)	"	Estimated annually with a N(0,1.4) penalty
	Spawner-recruit steepness *h*	H	0.86 (fixed)	Fixed at 0.49 or 0.67	Fixed (0.86)
Mortality & Reproduction					
	Natural morality (yr^{-1})	M	0.2 (fixed) for all ages	Fixed at 0.12 or 0.33	Fixed (0.2)
	Annual apical fishing mortality (F, yr^{-1}) for each fleet ("1980"–"2016")		Increasing function ("1980"-"2000"); generated with mean of *F* equal to *M* ("2001"-"2016")	"	Estimated annually by solving the Baranov catch equation
	Proportion maturity at age		0.2 at age 2, 0.25 at age 3, 0.4 at age 4, 0.5 at age 5, and 1 at age 6 and older	"	
Selectivity Patterns					
	Fishery	X	Time-invariant length-based asymptotic pattern. Length-at inflection and width for 95% selection were fixed at 30 cm and 8, respectively. All ages available in the area are selected by the fleet, which occurs only in the adult area.	Est. dome shaped for length-based availability	Est. asymptotic for length-based availability
	Survey(s)		Time-invariant length-based asymptotic pattern. Length-at inflection and width for 95% selection were fixed at 8 cm and 8, respectively. All ages present are selected by the surveys in each area.	"	Est. asymptotic for length-based availability

Spatial Structure					
	Number of areas	S	2	“	All aspects of spatial structure mis-specified when invoked
	Number of fishery fleets	S	1 (adult area)	“	
	Number of survey indices	S	2 (1 per area)	“	
	Number of length composition data sets	S	3 (fishery fleet and two surveys)	“	

1.7 Figures

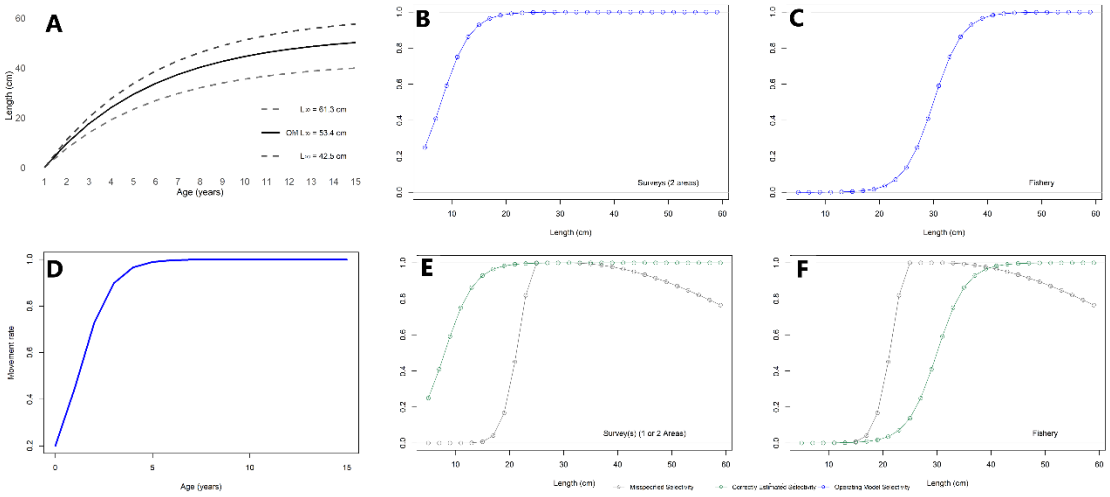


Figure 1.1: Top left: Length-at-age in the operating and estimation methods. Each stochastic OM replicate randomly assigns either the higher or lower growth regime to be implemented in all mis-specified estimation methods for that replicate. Bottom left: Movement from area 1 to area 2 by age in the operating model; in mis-specified estimation methods, spatial structure is ignored and thus movement is not applicable. Center and right columns illustrate selectivity-at-length in the OM (blue lines and points) and EM (green and grey lines and points). The grey lines indicate example estimates of dome-shaped selectivity for the EM; the green lines represent an EM with the correct selectivity form.

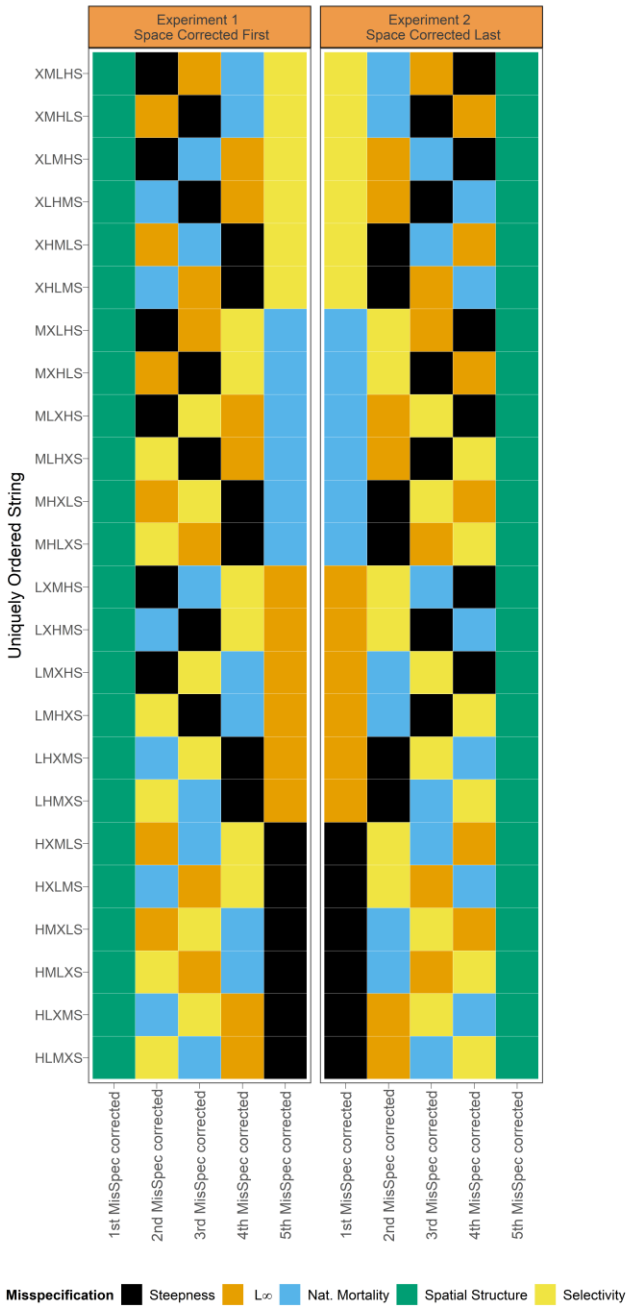


Figure 1.2: Schematic of the 24 ordered mis-specifications for the two experiments. Each row is a uniquely ordered string of five mis-specifications. Colored tiles represent an individual mis-specified estimation method; the x axis represents the order in which mis-specifications are corrected. Spatial structure (green tiles) is always corrected first in Experiment 1: Space First, and last in Experiment 2: Space Last. The study design therefore enables multiple replicates with a given mis-specification corrected in the same position (e.g. EMs in strings 13-18 for Experiment 2: Space Last all have asymptotic length corrected first). For each experiment, all six models for each of the 24 rows were fit to 50 OM replicates, before filtration.

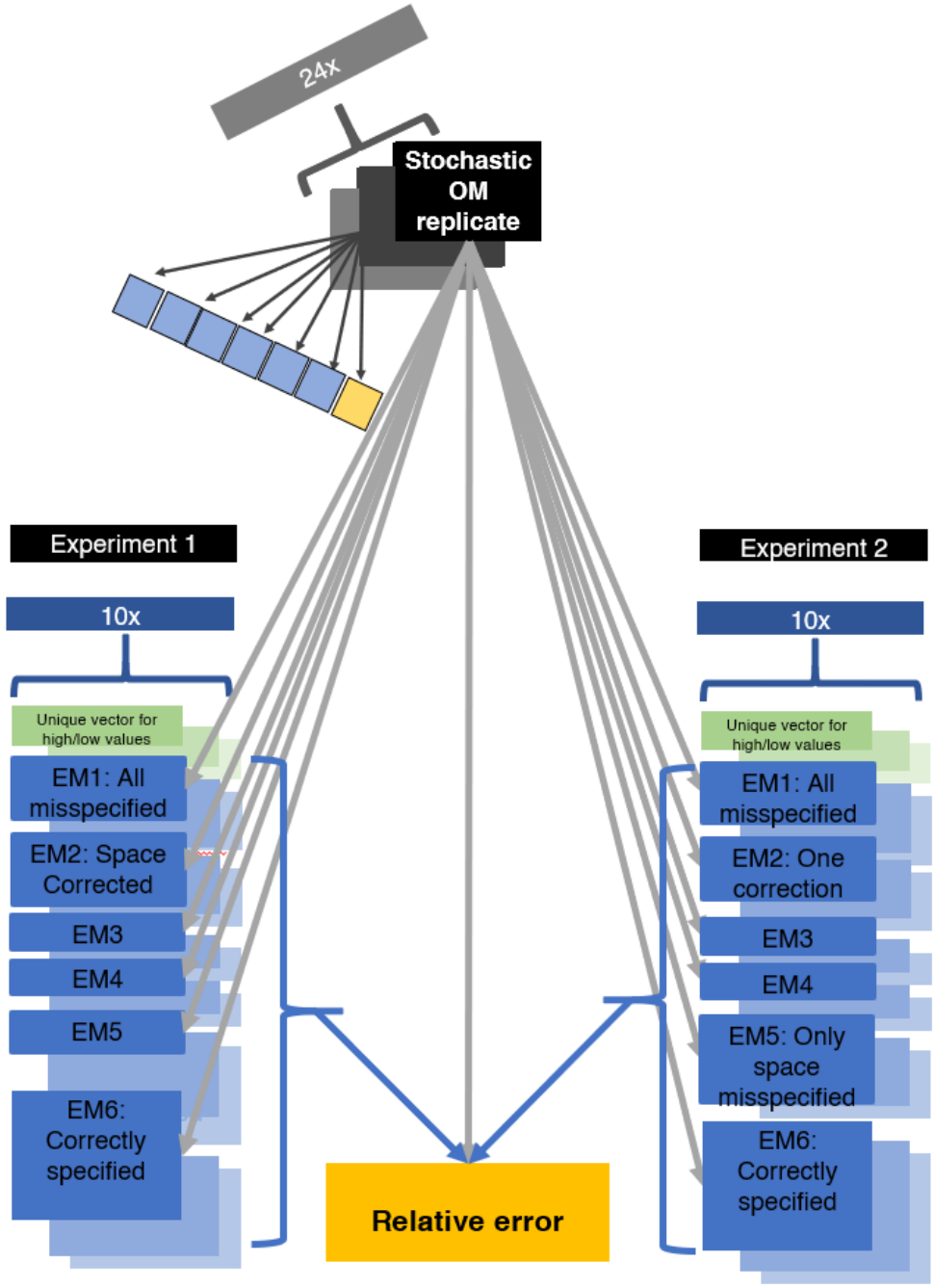


Figure 1.3: Schematic of experimental design. The data from each stochastic replicate of the operating model (greyscale boxes) are analyzed using six EMs (blue boxes) with varying degrees of mis-specification. For Experiment 2, the fifth EM contains only spatial structure mis-specified, and the sixth is correctly specified; Experiment 1 followed the same schematic, except that spatial structure was the first mis-specification corrected. The relative error is calculated by comparing the estimated population and reproduction quantities from each EM with the OM replicate associated with that set of estimation methods.

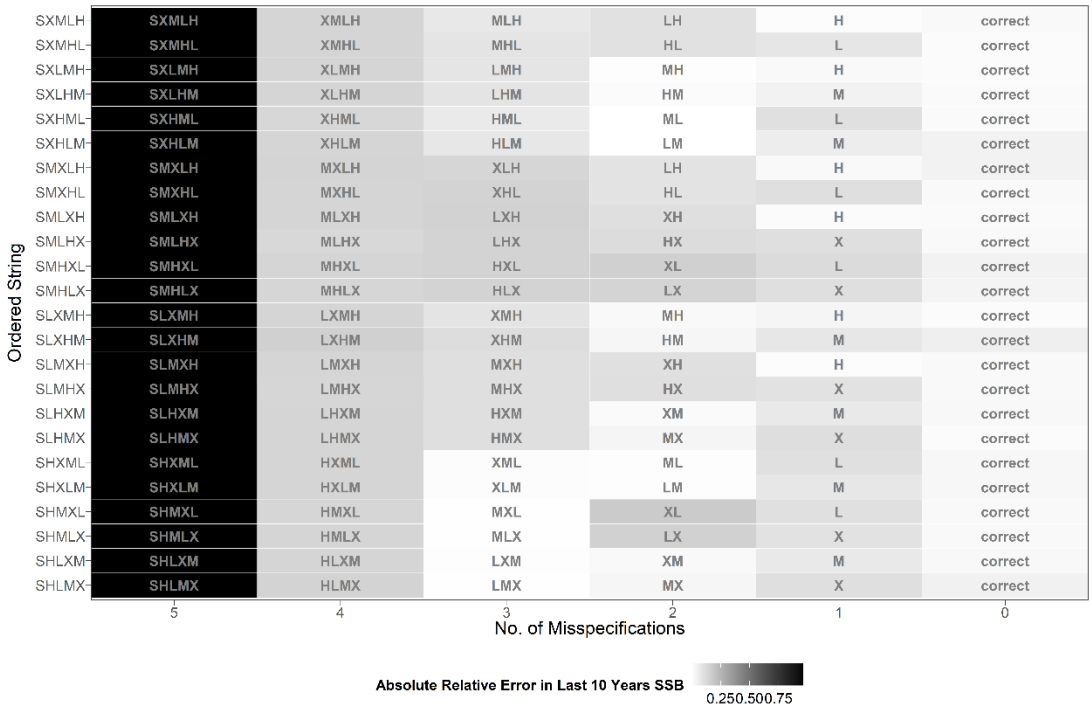


Figure 1.4: Mean absolute relative error for the last 10 years of spawning stock biomass (colors) for EMs averaged across replicates from Experiment 1: Space First. The x axis is sorted by the number of mis-specifications present, with the correctly-specified EM on the far right. Letters in each string indicate the order in which mis-specification correction proceeded for the EM concerned.

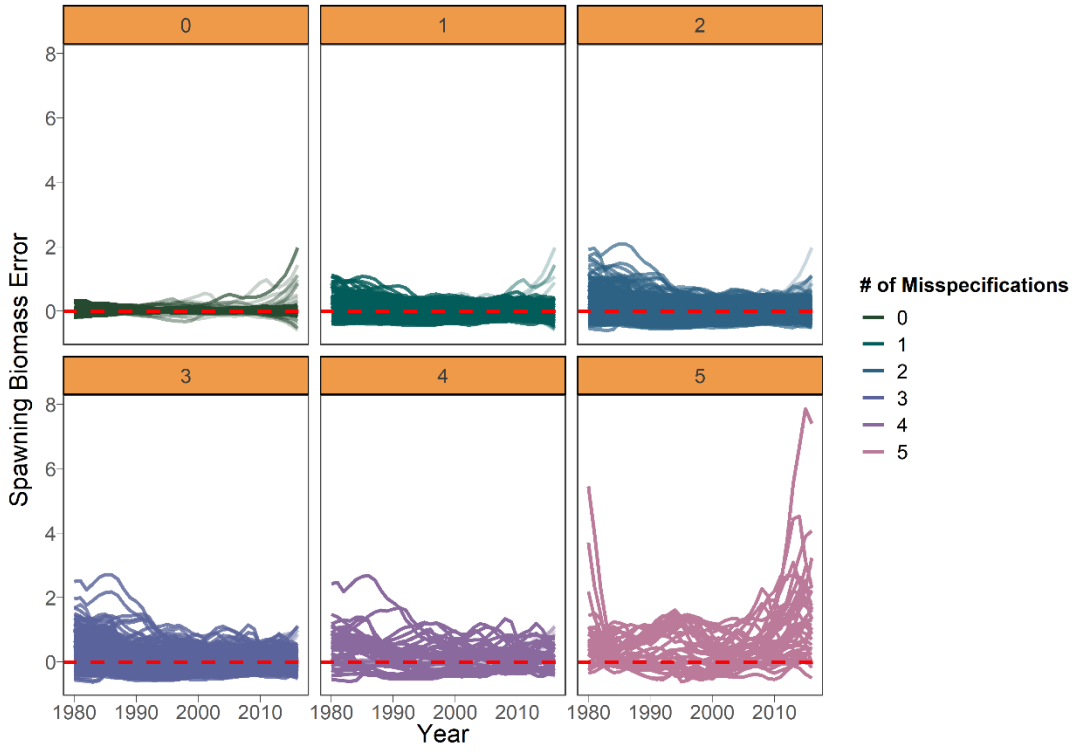


Figure 1.5: Time series of relative error in spawning biomass for Experiment 1: Space First (results for Experiment 2 are shown in Figure A.8). Panels and line colors correspond to the number of mis-specifications in the model. All models with four or fewer mis-specifications have spatial structure correctly specified.

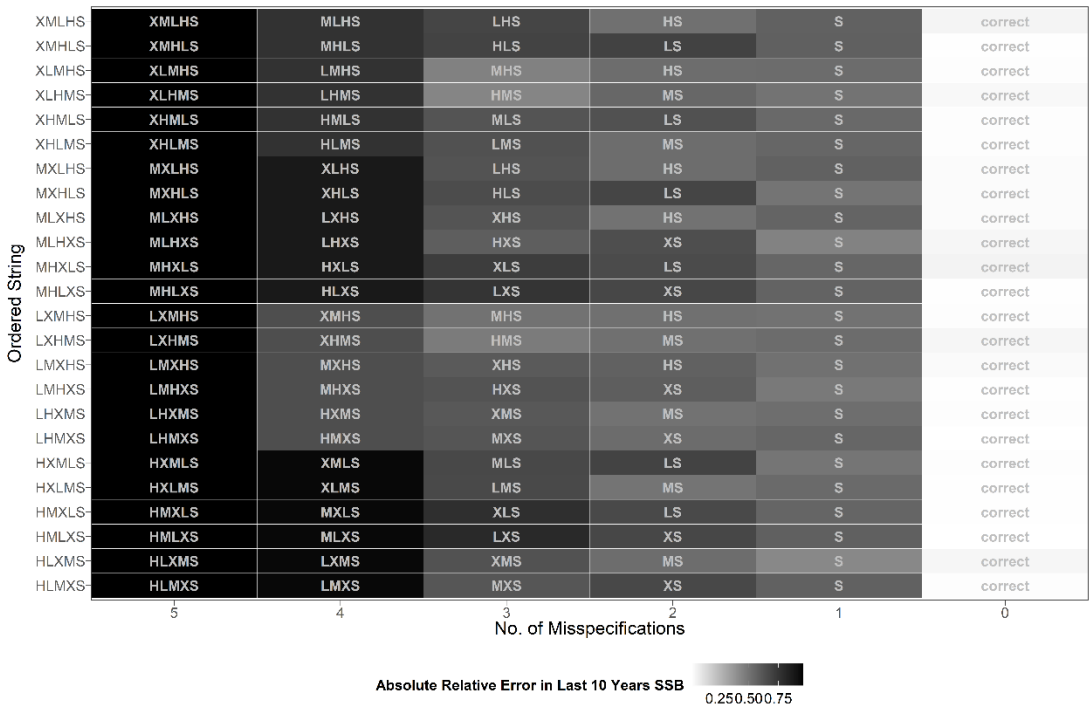


Figure 1.6: As for Figure 1.4, but the results shown are for Experiment 2.

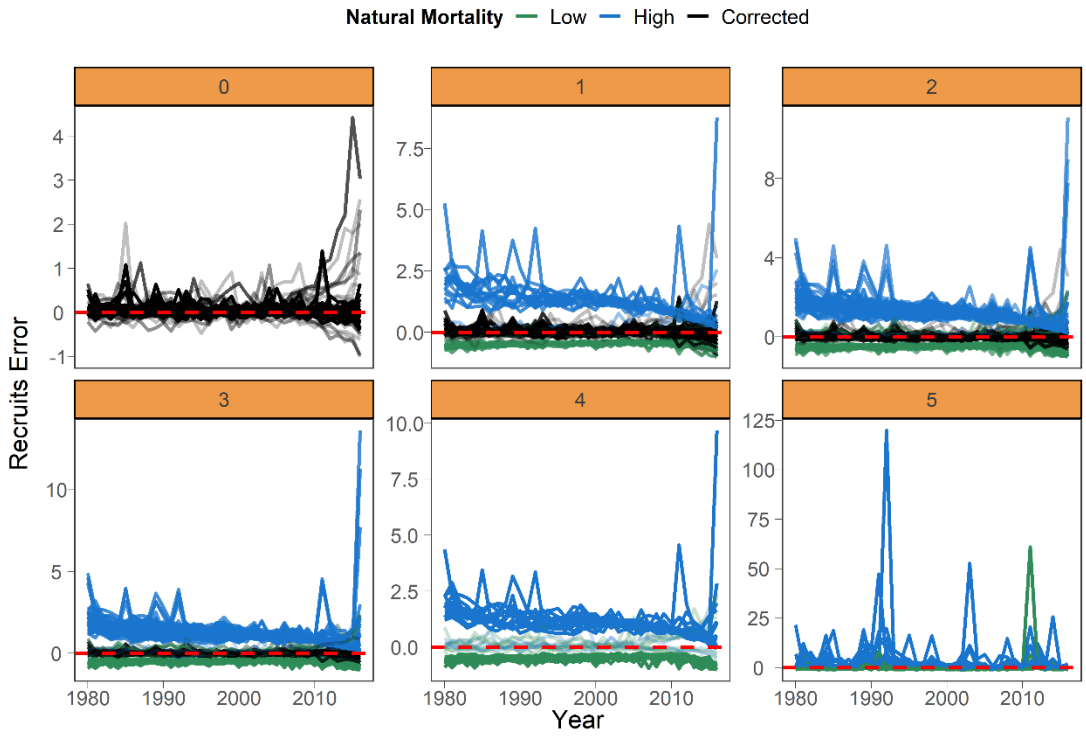


Figure 1.7: Time series of relative error in recruitment for Experiment 1: Space First. Panels and line colors correspond to the number of mis-specifications present in the model. The line colors indicating whether natural mortality (*M*) was mis-specified high (blue) low (green) or correctly (black) in the EM. Time series of relative error in recruitment for Experiment 1: Space First. Panels correspond to the number of mis-specifications in the model.

Chapter 2. OCEANOGRAPHIC FEATURES DELINEATE GROWTH ZONATION IN NORTHEAST PACIFIC SABLEFISH

2.1 Abstract

Renewed interest in the estimation of spatial and temporal variation in fish traits, such as body size, is a result of computing advances and the development of spatially explicit management frameworks. However, many attempts to quantify spatial structure or the distribution of traits utilize *a priori* approaches, which involve pre-designated geographic regions and thus cannot detect unanticipated spatial patterns. We developed a new, model-based method that uses the first derivative of the spatial smoothing term of a generalized additive model to identify spatial zones of variation in fish length-at-age. We use simulation testing to evaluate the method across a variety of synthetic, stratified age and length datasets, and then apply it to survey data for Northeast Pacific sablefish (*Anoplopoma fimbria*). Simulation testing illustrates the robustness of the method across a variety of scenarios related to spatially or temporally stratified length-at-age data, including strict boundaries, overlapping zones and changes at the extreme of the range. The approach is more sensitive than a pre-existing detection method (Rodionov, 2004). Results indicate that length-at-age for Northeast Pacific sablefish increases with latitude, which is consistent with previous work from the western United States. Model-detected spatial breakpoints corresponded to major oceanographic features, including the northern end of the Southern California Bight and the bifurcation of the North Pacific Current. This method has the potential to improve detection of large-scale patterns in fish growth, and aid in the development of spatiotemporally structured population dynamics models to inform ecosystem-based fisheries management, and has since been used for exploratory purposes in other western groundfish assessment and management applications (such as Pacific lingcod *Ophiodon elongatus*)

2.2 Introduction

There is no consensus on how to model region-specific growth patterns in assessment or population dynamics models. Fish somatic growth rates are typically modelled using the von Bertalanffy growth function (VBGF, Bertalanffy (1957)) or an alternative functional form, with parameters estimated using model-fitting procedures. The spatial resolution of the resultant estimates is necessarily predicated on the aggregation of the data, which is often defined by survey

stratification, political or management boundaries, and/or changes in sampling gear, not necessarily the ecology of the population (McGarvey and Fowler (2002); Williams et al. (2012)). For example, assessments of Alaskan sablefish stocks estimated separate VBGF parameters for two periods of survey data based on the *a priori* hypothesis that changes in survey gear type would affect estimates of fish growth from survey data (Echave et al. (2012); Hanselman et al. (2017); McDevvitt (1990)), and imposed a time block between which estimates of the growth curve parameters were quite similar in the stock assessment (Table 1). More sophisticated approaches that utilize hierarchical Bayesian methods to estimate latitudinal and regional effects on length- or weight-at-age require a design matrix of dimensions dictated by pre-supposed zones (e.g., Adams et al. (2018)). Such approaches are useful within a management context with rigid spatial boundaries, but do not represent the underlying growth process explicitly, and preclude the discovery of spatially-structured trends in fish size that do not match current management boundaries.

Existing methods to quantify spatial variation in somatic growth pose a trade-off. On one hand, researchers may impose a priori beliefs about spatial variation in stock traits or generate purely descriptive models of trait 'gradients' across regions or time periods, without a clear way to identify significant break points within them (King et al., 2001). This presents a challenge when developing population dynamics models that accurately represent the structure of managed stocks. An alternative tool is a model-based method that identifies break points in fish size-at-age, which can then be used to aggregate data and estimate parameters related to somatic growth. The significance of these breaks can be evaluated by comparing overlap in growth parameter estimates and tested against or among pre-specified breaks of interest (i.e., an area with a known ecosystem regime). To meet this need we present a new method, which uses the first derivative of smooth functions (splines) from a generalized additive model (GAM) to detect change points in spatially- and temporally-structured fisheries growth data that minimizes the use of pre-supposed stratifications in a simple, rapid computational framework. The method does not require the specification of multiple error structures nor the construction of spatial meshes, which can be computationally expensive when large (Thorson, 2019). The analysis of first derivatives of regression splines in GAMs for change-point analysis has been recently used in terrestrial paleoecology (Simpson, 2018) and geophysics (Beck et al., 2018). The underlying assumption is that the rate of change (the first derivative) of a given predictor is an appropriate measure of the

direction and magnitude of the predictor-response relationship. The spline itself may be highly non-linear, but predictor values at which the slope of the spline is largely positive or negative are taken to denote where the response variable is changing the most. Our GAM-based method has the potential to improve detection of large-scale patterns in fish growth, and aid in the development of spatially-structured population dynamics models.

We use simulation to test the robustness of the method using synthetic length-at-age data of varied complexity, and present a case study application to Northeast Pacific sablefish (*Anoplopoma fimbria*). Sablefish are a highly mobile, long-lived, and valuable groundfish that have high movement rates (10 – 88% annual movement probabilities across Alaska, with a mean great-circle distance of 191 km in a single year; Hanselman et al. (2014)) and range from Southern California to the Bering Sea. Concurrent population declines across the entire range over the past few decades have increased concern about the status of sablefish, and interest in identifying the causes of the downward trend. Sablefish stock assessment and management occur independently within political boundaries, namely Alaska (AK), British Columbia (BC), and the US West Coast in the California Current (CC), assuming that these are closed stocks. However, recent work has shown that there is little genetic evidence for population differentiation in sablefish across the northeast Pacific (Jasonowicz et al. (2016)), although there is evidence for differences in growth rate and size-at-maturity throughout the range (McDevitt, 1990). This suggests that the current delineation of assessment and management regions may be incongruent with the stock's actual spatial structure and underscores the potential value of developing a population dynamics model that represents the heterogeneity of sablefish growth throughout their range.

We developed a data-model-based method that would simultaneously identify spatiotemporal zones between which fish length-at-age varies and illustrate correlations between growth and spatiotemporal covariates (such as an increase with latitude). A method to identify such patterns in important population traits can help researchers determine whether current management scales are appropriate given the dynamics present in the population. Because these dynamics are potentially environmentally linked, such a method can also uncover whether spatiotemporal patterns in investigated traits correspond to major environmental features (such as ocean currents) or forcings (such as climactic oscillations), which can help inform the implementation of ecosystem-based fisheries management.

2.3 Methods

2.3.1 Method summary

The method fits a GAM to the vector of observed lengths of fish of a single age as the response variable, predicted by separate smoothers at knots t for year, latitude, and longitude, using the mgcv package (Wood, 2011) in R (R Development Core Team, 2016), i.e.

$$g\big(\boldsymbol{E}(\boldsymbol{X})\big) = \beta_0 + f(y_t) + f(s_t) + f(k_t) + \epsilon_t \qquad (2.1)$$

where $\boldsymbol{E}(\boldsymbol{X})$ represents the expected mean of fish length, g is an invertible, monotonic link function (in this case, the natural logarithm) that enables mapping from the response scale to the scale of the linear predictor, and the additive effects of latitude (s_t), longitude (k_t) and year (y_t), which are smoothed using a thin plate regression spline *f*. ϵ_t is a residual error term, assumed to be normally distributed. The effects of latitude, longitude and year on expected length-at-age are estimated as separate smoothers. To simplify the analysis, we fit the GAM to data for a single age-class and sex at once (e.g., age six for the simulated datasets), thus precluding the need to control for age or sex. Using fish of only a single selected age from all regions also minimizes the concern of differing age-based survey selectivities between management regions.

The first derivatives of the linear predictor with respect to latitude, longitude and year are evaluated to identify areas or periods (breakpoints) between which there is evidence for changes in fish length-at-age. The equations below provide an example using latitude s_t, but the process is repeated for each smoother. The finite differences method (as in Simpson, 2018) approximates the first derivative of the trend from the fitted GAM. For instance, the vector of derivatives G for latitude is produced via the following:

$$g(\boldsymbol{G_t}) = \frac{g(s_t + \alpha) - g(s_t)}{\alpha} \qquad (2.2)$$

where $g(s_t)$ is a vector of predicted fish lengths at latitudes and α = 0.001 in this analysis, with other effects (year, longitude) held constant. Therefore, the numerators of the elements of $\boldsymbol{G}$ are predicted lengths at two adjacent latitudes, separated by interval α, which is necessarily small. The standard error of the derivative estimates are computed as:

$$SE_t = \sqrt{\boldsymbol{G_t V}} \tag{2.3}$$

where V is the variance for the current spline; the square root provides the standard error for each derivative estimate of that predictor. These steps are repeated across the range of explored years and longitudes. All simulated datasets (see Simulation testing) were fit using a link function g with smoothing functions f for both spatial covariates as well as for year. For each parameter, we identify at which predictor value (e.g., latitude) the maximum absolute value of the first derivative is obtained; this is rounded to the nearest integer (e.g. a value between 22.5 and 23.4 would be rounded to 23) and defined as the "breakpoint" if its 95% confidence interval (generated using the standard error estimates for the derivative) does not include zero (see Figures 2.1 and 2.2, which illustrate the raw data, smoothers and first derivatives thereof for two synthetic datasets). The rounding step was implemented to ease comparison in the simulation study; we did not wish to treat a breakpoint estimate as incorrect if it differed by less than half of one degree (approximately 55 kilometers) from the true breakpoint. The raw length and age data (including all ages of fish) are then re-aggregated based on the identified breakpoints. For each of these new aggregated data sets, the parameters of the VGBF (Equation (2.4)); L_∞ - asymptotic length [cm], k - the rate at which asymptotic length is approached [cm/yr] and t_0 - the estimated age at length zero in years) are estimated using maximum likelihood, assuming that errors are normally distributed with zero mean and standard deviation σ). This study performed estimation using Template Model Builder (Kristensen et al., 2016).

$$\bar{L_a} = L_\infty \times (1 - \exp\big(-k(a - t_0)\big) + \varepsilon; \qquad \varepsilon \sim \mathrm{N}(0, \sigma^2) \tag{2.4}$$

2.3.2 Simulation testing

2.3.2.1 Outline and design

We conducted a simulation study to evaluate the performance of the proposed GAM-based method, based on datasets generated using an individual-based model (IBM, see Appendix B for full details). The IBM is capable of simulating individual characteristics by following the life history processes (survival and growth) of individual fish, with reproduction governed by a generalized stock-recruitment relationship to produce new individuals. An IBM was used to capture these key processes to simulate data similar in form to what would be included in a fishery stock assessment, which is difficult to do analytically or using age/size aggregated models. We

simulate spatial variation by generating length-at-age datasets under different growth 'Regimes' (defined as distinct L_1 and/or L_2 values, leading to varied L_∞) and assign latitudes and longitudes to fish grown under each regime. The IBM implements the VBGF using Schnute's (1981) formulation, which requires *k*, L_1, and L_2, with L_∞ computed as:

$$L_\infty = L_1 + \frac{L_2 - L_1}{1 - exp\big(-k(a_2 - a_1)\big)} \qquad (2.5)$$

where L_1, L_2 represent the expected lengths of fish at ages a_1, a_2, (3 and 30 years, respectively) and *k* is the growth coefficient. Each annual increment for every individual fish is subject to lognormal error. We considered five growth scenarios consisting of two growth "Regimes" with either completely distinct spatial or temporal ranges, or spatial ranges with some overlap. We designed our growth regimes to mimic the level of variation in L_1 and L_2 present in the sablefish dataset, which was as high as 26%. In our synthetic population for regime 1 L_1 = 10 cm, L_2 = 70 cm and k = 0.30 yr^{-1}; regime 2 was designed using L_1 and L_2 parameters 20% higher than regime 1 (L_1 = 12 cm, and L_2 = 84 cm, k = 0.30 yr^{-1}). Expected growth curves for the simulated regimes are shown in Figure B.1.

The simulated spatial extent ranges from 0° to 50° in latitude and longitude. The five simulation scenarios (Table 2.2) were designed to represent a variety of possibilities for spatial growth variation, with one scenario including a temporal regime change in growth. To simulate spatial zones, locations of fish grown under a certain regime were sampled from a uniform distribution with boundaries defined by the spatio-temporal scenario at hand (Figure 2.3). All fish in scenario 1 (no spatial or temporal variation) were grown under regime 1 and sampled (uniformly) over latitude and longitude between 0° to 50°. In scenario 2, fish were grown in two regimes, and fish grown under regime 1 were between 0° and 25° (latitude and longitude) while fish grown under regime 2 had coordinates sampled between 25° to 50°. The same approach was applied for scenario 3, except that fish grown under regime 2 were sampled from 20° to 50°, thus creating an overlap zone between 20° and 25°. All simulated fish in scenario 4, had latitudes sampled from 0° to 50°. Fish simulated under regime 1 were assigned longitudes sampled randomly from 0° to 48° and fish simulated under regime 2 have longitudes sampled randomly from 48° to 50°, forming a vertical "band" of larger fish in higher longitudes. The final simulation scenario (5) involved temporal changes in growth, with a change from growth regime 1 to regime 2 in year 50. This

meant that the growth increment generally increased for individuals whose lifespan covers this breakpoint, although note that the GAM is fit to fish of a fixed age. Fish locations for the temporal break scenario are sampled identically to the scenario without spatial variation.

One hundred replicate datasets were generated for each scenario, which averaged 530 age-six fish per dataset (a sensitivity analysis was performed reducing the sample size by 25% or 50%). For all runs, the initial values for the parameters were t_0 = 0.1 yrs, σ = 1.1, with L_∞ = 150cm and k = 0.1yr^{-1}. The estimation procedure also calculated the predicted length at the endpoints of the estimated growth curve (Equation (2.5); the length at pre-specified minimum (L_1) and maximum (L_2) ages, which were 3 and 30 years in the simulation studies). These values and their standard errors were used in the evaluation of the method as L_∞ and k are typically negatively correlated.

2.3.3 Performance metrics

We considered two performance metrics: 1) the proportion of simulations in which the correct spatial and/or temporal breakpoints were detected - we tabulated the number of times a breakpoint found using a GAM fit to a dataset matched the true latitude, longitude, and year; and 2) the coverage probabilities (determined by the 95% confidence intervals) for L_1 and L_2 . For all but the scenario with overlapping ranges (scenario 3), we only considered the GAM analysis to have correctly identified the true breakpoint only if it was an exact match. The 'true' dataset for scenario 3 contained fish grown under regimes 1 and 2 in a shared region between 20° and 25° latitude and longitude, so the detected breakpoint was counted as an accurate match if it fell within this range. For each scenario, after aggregating each of the 100 simulated datasets into the GAM-designated spatiotemporal strata and estimating the growth curve, we determined whether the 95% confidence intervals of the estimated fish lengths at ages zero and fifteen (our a_1 and a_2) contained the true L_1 and L_2 values. For example, fish generated under regime 1 and occupying latitudes and longitudes between 0° and 25° may have been re-aggregated via the GAM analysis into a de facto 'region' ranging from 0° to 24° for an "early" period of years 1 through 37; the parameters of the VBGF were estimated on this per-strata basis, and the terminal lengths of the estimated curve compared to those from which they were generated, in this case, regime 1. Fits from the complementary de facto 'region' ranging from 24° to 50°, and/or a "late" period, would be compared to whichever regime generated the majority of fish therein. An estimated endpoint from a GAM-defined region was considered a match if the 95% confidence interval for it contained the true value of L_1 or L_2 .

To facilitate comparison between the proposed GAM-based method and an extant approach, we applied the sequential t-test analysis of regime shifts (STARS, Rodionov, 2004) using length-at-age for age 6 to our simulated datasets for both spatial and temporal changes. The STARS method was originally developed to detect climate regime shifts in time-series data, and was noted for its sensitivity to changes towards the end of a series. The method examines the sequential differences in the value of a t-distributed variable, and determines whether subsequent measurements (at the next year or latitude, for example) exceed the expected range. We used a minimum regime 'length' of five, meaning detected shifts between latitudes, longitudes or years must persist for at least five consecutive units, and the default p-value cutoff of 0.05. We believe this captures the timescale of regime shifts of interest to ecologists, and a significance cutoff frequently used in such analyses. From the STARS analysis of each dataset, we selected the breakpoint(s) with the largest positive "regime shift index", which represents a cumulative sum of the normalized anomalies. This is qualitatively similar to the "largest first derivative" metric used in the proposed GAM-based method and, as in that case, was applied regardless of where the breakpoint was detected. We implemented the same steps, whereby the detected spatial and/or temporal breakpoint(s) were used to re-aggregate and estimate growth parameters, and the proportion of accuracy and coverage probabilities for L_1, and L_2 tabulated.

2.3.4 Application to Northeast Pacific Sablefish

We obtained fishery-independent length and age data from the Bering Sea, Aleutian Islands, and Gulf of Alaska Sablefish Longline Survey (Rutecki et al., 2016) and the U.S. West Coast Groundfish Bottom Trawl Survey (Northwest Fisheries Science Center, 2019) conducted annually by the Alaska Fisheries Science Center and the Northwest Fisheries Science Center, respectively. We also obtained length and age records from the Canadian Department of Fisheries and Oceans (Wyeth et al., 2005); see Table 2.1 for a summary of survey data used in the application. Data from each management region included measured length, sex, age, and the starting latitude and longitude, which determined the survey station. Due to computational constraints, and to avoid disproportionate influence of more heavily-sampled areas on breakpoint estimates, we randomly subsampled 15,000 total records from each of the three management regions. The subsampling was random with respect to latitude, longitude, age and sex, using the sample_n function from the package dplyr (Wickham et al., 2019). We applied the method to identify spatial and temporal

breakpoints for each sex separately at several key ages: age 4 (before the length-at-50%-maturity for both males and females in all management regions), age 6 (after the length-at-50%-maturity for both males and females in all management regions) and age 30, roughly the length at which sablefish are expected to obtain their maximum length (Johnson et al., 2015). Our sampling method produced a data set with an average of 1,315 age 4, 1,283 age 6, and 65 age 30 sablefish of each sex from each management region. Growth model fitting was performed using all available data from each of the three management regions (see Table B.3 for sample sizes). In constructing the GAM, we investigated the use of an AR1 temporal structure for the residuals with lags of 1 to 3 years, but these models did not improve AICc over the initial model (without autoregressive structure).

We re-aggregated all data to match the breakpoints that appeared in the GAM analysis for key ages, as well as an ecosystem-based breakpoint at 145^0W. We selected this breakpoint based on work by Waite and Mueter (2013) who used cluster analysis to delineate unique zones of chlorophyll-a variability, which has been shown to be influential in the sablefish recruitment process (Shotwell et al., 2014) but by definition such an effect is not detectable in our analysis that only examines fish larger and/or older than recruits. The North Pacific Fishery Management Council uses 145^0W, which includes a cluster of several seamounts in the Gulf of Alaksa, to delineate a groundfish slope habitat conservation area (Siddon and Zador, 2018). We employed a stepwise exploration of whether estimates of L_∞ were significantly different between detected regions using the method and generated from this ecosystem break using the entire, non-sub-sampled dataset. Asymptotic length was used to ease comparison between estimated values and those used in the current assessments. This involved first aggregating and estimating the VBGF for ten unique spatiotemporal strata for each sex, defined by the one temporal and three spatial breakpoints found among the key ages selected for analysis using the GAM in addition to the break at the aforementioned ecosystem feature. To account for length-based selectivity, which is implemented only for the British Columbia data, we applied a penalty to the likelihood function as follows:

$$L_{D|\theta} = \Pi_i S_{L_i} \frac{1}{\sqrt{2\pi}\sigma a_i} e^{-(L_i-\hat{L}_i)/2\langle\sigma a_i\rangle^2} \Big/ \int_{-\infty}^{\infty} S_l \frac{1}{\sqrt{2\pi}\sigma a_i} e^{-(\hat{L}_i-l)/2\langle\sigma a_i\rangle^2} \qquad (2.6)$$

where L_i is the observed length at a given age a_i, $\hat{L}_i$ is the corresponding estimate based on VBGF parameters , S is a logistic selectivity function with parameter L_{50}, the length at which 50% of individuals (male or female) are fully selected, set to 52.976 cm (Samuel Johnson, SFU, pers. comm.)

$$S_l = \frac{1}{1 + exp(L_{50} - L)} \qquad (2.7)$$

As length-based selectivity is assumed constant in both the California Current and Alaskan assessments, S_L is set to 1.0 when fitting data points from those regions. We then examined whether the 95% confidence intervals for L_∞ overlapped for any temporally-split datasets from the same region (e.g., region 1 female sablefish data before and during 2010 and after 2010). If they did, we pooled the data for that region and sex for all years. In the second step, we examined if spatially-adjacent regions (from any time period) for the same sex had 95% confidence intervals for L_∞ that overlapped, and combined regions for which this was the case on a by-sex basis. This stepwise approach reduces unnecessary partitioning of the data into spatiotemporal strata that do not ultimately result in different estimates of L_∞, and allowed us to examine whether any of our detected breakpoints or the *post hoc* ecosystem split was informative regarding growth estimates. Once the most parsimonious structure was identified through this method, we generated predicted lengths-at-age for the entire dataset.

2.4 Results

2.4.1 Simulation Study

The simulation study demonstrated that the first-derivative GAM-based method is able to detect both spatial and temporal breakpoints correctly in the majority of scenarios, with the exception of the scenario where the spatial break occurred near the edge of the simulated spatial extent at 48° longitude, where it only detected the break location correctly in 15% of simulations. Figure 2.4 displays the coverage probabilities for the 95% confidence intervals and proportion of simulations wherein the correct breakpoint was detected perfectly or with a "relaxed" criteria (within 2 degrees, roughly 220 km, or 2 years), demonstrating the success rate of the method across a variety of simulations. Figure B.3 and B.4 present histograms of detected breaks for each scenario.

For all scenarios, the method achieved the highest coverage probabilities for the length-at-age 0 (L_1, 48%-97% coverage for three scenarios and 27% in the scenario with overlap). Coverage probabilities for length-at-age 15 (L_2) were slightly lower (43% - 74% for three scenarios and 16% in the scenario with overlap). In terms of spatial breakpoint detection, there was not a qualitatively strong difference in the method's ability to correctly detect latitudinal vs. longitudinal breakpoints across scenarios. The GAM-based method correctly detected the lack of a breakpoint in 86% of simulations without breaks; there was no discernable pattern to the spurious spatial breakpoints identified in the remaining simulations. The method did less well at detecting breakpoints for scenario 4 (a "true" spatial break at 48°), assigning the break between 45° and 50° longitude in 100% of simulations; similarly, for the scenario with a single breakpoint at 25°, the GAM-based method was 100% accurate when the criteria were relaxed to include breaks from 24° to 26°. Relaxing the criteria in this manner increased the method's accuracy to over 90% for all scenarios except one (Figure 2.4c). We computed the mean absolute error in both L_1 and L_2 estimates across scenarios and found the maximum error to be 1.84 cm for L_1 and 6.98 cm L_2, both obtained in scenario 1. Finally, we did not find the method's accuracy sensitive to either halving or reducing the sample size by 25% (Table B.2).

2.4.2 Comparison to STARS Method

The STARS method (Figure B.2) was inferior to the proposed GAM-based method at detecting spatial or temporal break points for all simulated scenarios, with a slight exception for the break at edge case (scenario 4). For all other scenarios, the STARS method performed up to 90% worse than the proposed GAM-based method at detecting latitude and longitude breaks, and 20% worse at detecting year breaks. It also performed worse in terms of the coverage probability of L_1 (63% vs 67% for the GAM-based method) and L_2 (18% vs 52%), and did slightly better than the proposed method in detecting the break-at-edge, though only at 31% (vs 11%).

2.4.3 Application to NE Pacific Sablefish

The latitude smoother suggested a generally increasing cline in length-at-age with latitude, with a significant breakpoint around 50°N (approximately the northern end of Vancouver Island, Canada) detected when the GAM was fit for age four and six sablefish (Figures 2.5c, 2.6c; Figures B.6, B.7, B.9). North of this breakpoint, female L_2 estimates were consistently larger than 70 cm,

whereas they averaged 65 cm south of it. Both age six and age 30 female sablefish identified a breakpoint at 36°N (approximately Monterey, CA, USA). Both males and females obtained the lowest estimated L_2 south of this breakpoint, at 55 cm for males and 60 cm for females. In all GAM-detected regions, L_∞ was higher for female sablefish than males, and the resultant L_2 differed between regions within sexes by up to 26%. The temporal smoother did not exhibit a strong one-way trend, and was flat for age-30 fish of both sexes, although it did detect a break in 2009-2010 for both sexes of age 4 and 6 sablefish. Parameter estimation at this temporal stratification generated 95% confidence intervals for L_∞ that overlapped for males within all regions and for females in region 5 (Figure B.14). The number of spatiotemporal strata was reduced to 14 after combining years of data for region-sex combinations where overlap was found in the second phase. Once re-aggregated and re-estimated, we did not find overlapping confidence intervals for L_∞ for any adjacent regions, so this set of specifications (five spatial regions for both sexes, and a temporal break for females in regions 1 through 4) was retained as our final spatiotemporal stratification. The stratification consists of three regions bounded on their western border by a break at 130°W; from south to north, these regions (labeled 1, 2 and 3 on Figure 2.7) are defined by latitudes 36°N and 50°N. They correspond generally to Monterey, CA and the northern tip of Vancouver Island, BC. Region 4 is the area between 130°W and the ecosystem break at 145°W (roughly Cordova, AK). Datapoints collected to the west of the ecosystem break are assigned to region 5.

2.5 Discussion

Empirical work has suggested that somatic growth in fishes follows ecosystem gradients rather than management boundaries (Pörtner and Knust, 2007; Taylor et al., 2018). The ongoing emphasis on ecosystem-based fisheries management calls for the analysis of fish stocks (ideally in a multi-species context, but also as single species) at meaningful spatial scales, across which changes can be detected. Our goal was to investigate the performance of a method to improve detection of large-scale patterns in fish growth and apply it to length-at-age data from the Northeast Pacific sablefish. Our method determined that the current management scale (three political breaks at national boundaries) is incongruent with the underlying pattern of variation in sablefish growth. We discerned that the spatial variation in sablefish growth corresponds well with major oceanographic features, principally the splitting of two major ocean features and the edge of a

highly productive zone. Below, we discuss the results of the simulation study and provide further guidance on how researchers could apply our proposed method to new datasets. We then discuss the results found during the application to northeast Pacific sablefish, with respect to ecosystem concerns.

2.5.1 Implications of Simulation Results

Our GAM-based method indicated tradeoffs between the accuracy of breakpoint detection and resultant coverage probabilities in the estimated growth curve, as well as large differences in the coverage probabilities of fish length at younger versus older ages. We find it encouraging that the approach could correctly detect breakpoints for the scenario with overlapping ranges, which is likely more like real-world fish populations than the singular, immediate breakpoints considered in other scenarios. However, the assigned 'zonation' of these populations necessarily combined fish with contrasting growth curves into a single dataset for estimation and resulted in a loss in accuracy (coverage probability) for the endpoints of the growth curve. Alternate GAM-based methods, such as the clustering approach applied in Winton et al. (2014), have also demonstrated that detecting spatial structure through a spatially explicit process can reveal distinct sub-areas in fish traits (e.g., mortality). That study also found that models did not necessarily require explicit ecosystem data (like temperature) to perform as well as models with only spatial information.

We suggest that our method be used as a tool to guide the identification of general zones between which growth could vary, and not take detected breakpoints as the absolute truth. Importantly, suggestions of spatial breakpoints produced by the method should necessarily be considered in the context of the ecosystem, and prior knowledge of how the fishery at hand responds to features (e.g., temperature, depth) that vary with latitude and/or longitude. Absent an ecosystem-wide analysis, strong directional trends in any generalized additive term (such as the positive trend with latitude observed here) or a breakpoint at the edge of the study area can be indicative of a change somewhere in the margins and extend the reach of future survey designs.

The method performed best for both performance metrics for the scenario in which growth regimes 1 and 2 overlapped in space (which had the advantage of being 'matched' whenever the detected breakpoint fell within the range of overlap, 20° to 25°). The most commonly detected breakpoint in latitude and longitude for that scenario, before rounding, was the midpoint of this range (22.5°),

likely an artifact of the penalization function within the GAM, which seeks to minimize curvature on either side of a given knot (i.e., the breakpoint). This penalization function controls the degree of smoothness on the spline and can lead to fitting overly-complex models when unchecked (Wood, 2011). Since the purpose of this analysis was diagnostic (the detection of where the spline is changing the most), we were able to avoid undue influence from this parameter by a) selecting only the value corresponding to the maximum first derivative and b) that had confidence intervals not containing zero, which are common in highly curved splines. We also chose to use only the maximum absolute value of the derivative to avoid splitting the spatio-temporal surface into many small zones, which may have led to problems of small sample size, or ultimately be unrealistic to implement in a population dynamics model for the stock.

We detected spurious spatial or temporal breaks in ~10% of simulations for which no breakpoints were present. However, some erroneous detection can be expected considering the inherent noise in our datasets, and that there is no minimum threshold for breakpoint detection; a single, small derivative among many zeros that did not have a confidence interval containing zero could be 'picked'. This observation partially motivated the two-phase procedure employed for the sablefish application, so it is likely that such erroneous detection would be reduced if overlapping growth estimates were disregarded (our simulation analysis investigated the accuracy of the first stage). We evaluated if an autoregressive structure improved our simulation models as length-at-age can be time-dependent, but it did not; this may not be the case for other fisheries.

In addition, we did not simulate nor consider error or bias in the aging (i.e., otolith reading) process (Cope and Punt, 2008), which would potentially introduce uncertainty in breakpoint detection. Based on aging workshops conducted for sablefish, we consider aging results used in the case study to be roughly comparable among regions (Fenske et al., 2019). With these caveats in mind, we envision (and demonstrate) using the method as a tool to identify general regions and periods of change in fish length-at-age, which will necessarily be evaluated against pre-existing knowledge of the fish population and its ecosystem.

Neither the GAM-based nor the STARS approach is appropriate for extrapolation (prediction beyond the range of covariates, or outside of the ecosystem, used in model fitting), particularly because they use indirect variables such as latitude which may have nonlinear or inverted relationships with fish physiology in other ecosystems (Austin, 2002). It is likely there are

thresholds in, or types of, spatiotemporal growth variation that will be poorly detected by most methods, which we see as a promising area for future research.

2.5.2 Implications of detected breakpoints for northeast Pacific Sablefish

Our evaluation of size-at-age for northeast Pacific sablefish was directly motivated by the notion that sablefish growth may vary at a scale that differs from present management boundaries. For northeast Pacific sablefish, we applied the method to each sex separately at a set of key biological ages and determined that sablefish length-at-age differs most significantly across five regions, whose boundaries can be defined by major oceanographic features (the Southern California Bight, and the bifurcation of the North Pacific Current) as well as a known ecosystem boundary in the Gulf of Alaska. It is evident from this and previous work (Echave et al., 2012; Gertseva et al., 2017; McDevvitt, 1990) that there is some level of variation in sablefish growth, whether in the growth rates themselves or the spatiotemporal scale at which variation in growth occurs. Previous work with sablefish data has utilized an *a priori* method, wherein length and age data were aggregated into pre-hypothesized spatial zones and fitted VBGF curves were compared using Akaike's Information Criterion. This 'information-theoretic' (Guthery et al., 2003) method is fairly straightforward computationally, and has been implemented separately for the California Current (Gertseva et al., 2017) and Alaska federal sablefish fisheries (Echave et al., 2012; McDevvitt, 1990). The California Current analysis identified a statistically significant break in VBGF parameters for sablefish at approximately $36°\text{N}$, between Point Conception and Monterey, CA, with additional evidence for an increasing cline in L_∞ with increasing latitude and a general increase in estimated L_∞ and L_2 for more northerly regions. These results mirror the trend in the latitudinal smoother (Figures 2.5 and 2.6) and the breakpoint detected at $36°\text{N}$ (Figure 2.7), which is incidentally a management sub-boundary used by the US Pacific Fishery Management Council (2013). That work also found an increase in k estimates for areas sampled south of the Vancouver area (ca. $49°\text{N}$), which was posited to be the result of samples coming from the "southern end of a faster-growing northern stock", a suggestion supported by our findings of another breakpoint at $50°\text{N}$. Preliminary analyses of sablefish tagged in Alaska suggest that the British Columbia management region exports fish into the California Current and Gulf of Alaska, a diffusion pattern that could potentially taper off with decreasing latitude; the distance between Vancouver, B.C. and Monterey, C.A. is approximately three times the mean great-circle movement distance for

sablefish determined by Hanselman et al. (2014), which is a measure of the shortest possible distance traveled between tagged and recovered animals. Gertseva et al. (2017) described how sablefish have been shown to be highly mobile, with ontogenetic movements off the coastal shelf; such combined, complex life patterns could yield higher growth rates in northern latitudes that interact with a more generalized shelf-slope pattern of ontogenetic movement observed in groundfish overall.

There are several noteworthy trends in the stratified growth estimates (Figure 2.8) that warrant future research. Firstly, the post hoc incorporation of a spatial break at 145°W based on ecosystem data was not ruled out during the significance testing of L_∞. This supports the notion that environmental features may result in variations in growth, and that the proposed GAM-based method is amenable to improvements based on the incorporation of climate or ecosystem knowledge. Additionally, both latitudinal breakpoints are loosely associated with significant oceanographic features, namely start of the southern California Bight at Point Conception (~34°N) and the bifurcation of the North Pacific Current, which splits into the Alaska and California currents as it approaches the west coast of North America. The breakpoint at 36°N is slightly north of the start of the bight, but also characterized by dynamic, mostly southward floor in the nearshore environment. The formal location of the North Pacific bifurcation varies, but is generally centered off the coast of British Columbia (Cummins and Freeland, 2007) (Figure 2.7). In common with the ecosystem split identified in the Gulf of Alaska, these oceanographic features lead to distinct zones of productivity (Kim et al., 2009; Mackas et al., 2001) that could influence resource availability and subsequent growth.

The temporal break in year 2010 was conserved (supported by significantly different L_∞ estimates) only for female fish, and more so in the southerly latitudes (such as regions 1 through 4, which are mostly comprised of California Current data), and exist along a steeper north-south cline. We note, however, that the procedure used to eliminate 'overlapping' L_∞ estimates concerned only statistical differences in values (and are therefore sensitive to sample sizes). The biological significance of these values would need to be investigated in the context of fecundity and length-weight differences between regions.

Preliminary analyses of sablefish movement rates from tagging data from Alaska (as analyzed in (Hanselman et al., 2014)) indicate that male sablefish seem to move more frequently to and from

sea mounts, which are situated within the GAM-defined regions identified here. There are several possibilities for why female sablefish seem to exhibit finer spatiotemporal structure in growth. Empirical work in Canada (Mason et al., 1983) that examined early life history of fishery-caught coastal sablefish observed a slight cline in mean fork length with increasing latitude, although the sex ratio within the study was biased towards females. That study suggested that selectivity for female sablefish may be higher due to higher congregating or feeding activity, in addition to the fact that females grow larger and are likely preferentially targeted in the commercial fishery in British Columbia, which is also true for the fixed-gear fisheries in the California Current (Kapur et al., 2021b). This could render females more sensitive to changes in fisher behavior, such as the implementation of catch shares off the US west coast in 2011. Expanding the method to allow for detection of multiple spatial and/or temporal breaks at once may enable further investigation of this phenomenon, although it may lead to the creation of spurious regions with insignificant difference in growth parameters, as observed in the first phase of the case study.

A plausible scenario that would generate our observed results could be that changes in fisher behavior or climate during the last ~10 years caused female sablefish to move northward in greater numbers, or simply experience size-based truncations in regions to the east of 145^0W due to fishing pressure. Each of these phenomena would have an inverse effect on resultant size-at-age, with fish entering the northern ecosystem tending to grow larger and high, persistent fishing pressure in any region leading to truncations in terminal size. Because we only detected slight declines size-at-age between time periods for female sablefish, it is possible that either fishery-related effects simply have not lasted long enough to be strongly evident, or such effects are being counteracted by more fish entering ecosystems favorable to higher terminal sizes. A closer examination of sex-related movement would be useful towards this understanding.

Consideration of temporal variation in sablefish growth is further complicated by the exploitation history of the fishery, which has steadily moved north- and west-ward in the California Current and Alaska over the last several decades, encountering ‘larger’ fish with subsequent expansion (Pacific Fisheries Management Council, 2013). This suggests that differences in mean length across the region could be attributable to different degrees, durations, or patterns of fishing pressure (Hilborn and Minte-Vera, 2008), interacting with inherent growth variation to produce such spatiotemporal patterns. A principal conclusion of Stawitz et al. (2015) was that the form of

sablefish growth variation differs among ecosystems, wherein the California Current is a more climactically variable ecosystem. Such ecosystem-driven trends may be diluted when analyzing the data as a composite, as in our study. Notably, our temporal smoother did not produce a distinct annual or cyclic trend. Methods that consider the space and time components co-dependently (as in vectorized auto-regressive spatiotemporal models (Thorson, 2019)) may strengthen the ability to disentangle such trends, and also to consider covarying spatial effects (e.g., near- and offshore).

2.6 Tables

Table 2.1: Overview of survey methods, data available and most recent VBGF parameters used for sablefish in stock assessments.
*Time-blocked VBGF parameters for Alaska Federal assessment 1996-2018
**Time-blocked VBGF parameters for Alaska Federal assessment from 1960-1995 (Hanselman et al., 2017). ^The BC assessment fixes length at age-1 to 32.5cm.

Region	**Survey Method**	**Sample size used in this analysis to fit GAM**		**VBGF parameters from recent stock assessments**					
		M	**F**	**L_∞ (cm)**		**k (years $^{-1}$)**		**t_0 (years)**	
				M	**F**	**M**	**F**	**M**	**F**
West Coast of US (Johnson et al., 2015)	Trawl on chartered commercial fishing vessels	7,778	7,222	57	64	0.41	0.32	0 (fixed)	0 (fixed)
British Columbia	Stratified trap survey	6,912	8,088	68.99	72.00	0.29	0.25	^	^
Alaska Federal (Hanselman et al., 2017)	Longline on chartered commercial fishing vessels	6,818	8,182	*67.8 ‡65.3	*80.2 ‡75.6	*0.29 ‡0.28	*0.22 ‡0.21	**‡2.27	**‡1.95

Table 2.2: Summary of simulation scenarios used to test the proposed GAM-based method given various extents of spatial growth variation, and a single temporal scenario.

Scenario Number	Scenario Description	Stratification
1	No spatial breaks	Latitude and Longitude ~ U[0,50], all fish under regime 1
2	Single, spatial break in middle of range, with no overlap	Latitude and Longitude ~ U[0,25] under regime 1; Latitude and Longitude ~ U[25,50] under regime 2
3	Some overlap between regions	Latitude and Longitude ~ U[0,25] under regime 1; Latitude and Longitude ~ U[20,50] under regime 2
4	Single spatial break at edge of range with no overlap	Latitude ~ U[0,50] for regimes 1 and 2; Longitude ~ U[0,48] for regime 1 Longitude ~ U[48,50] for regime 2
5	Single temporal break at year 50 (of 100); no spatial variability	Latitude and Longitude ~ U[0,50], all fish under regime 1 from years 0 to 49 and regime 2 thereafter

Table 2.3: Description of final spatiotemporal regions, and the sex-specific growth parameters estimated in the analysis. The Region column corresponds to regions depicted in Figure 2.7, with "early" period being observations before or during 2010, where applicable. Parameter estimates are those used to plot fitted curves in Figure 2.8. *Age 0.5 yrs was used to report L_1 estimates, except for values from Regions 3 and 4 for which L1 corresponds to lengths at age 4.

Region	Sex	Period	Sample size used to fit GAM	Estimated VBGF Parameters			Corresponding estimated endpoints of growth curve	
				L_∞ (cm)	K (years^{-1})	t_0 (years)	L_1 (cm)	L_2 (cm)
1	Female	Early	616	60.44	0.29	-2.15	32.21	60.43
1	Female	Late	699	62.86	0.16	-4.31	34.22	62.63
1	Male	All years	1,314	55.11	0.28	-2.59	32.08	55.11
2	Female	Early	4,913	69.14	0.22	-0.96	19.22	69.08
2	Female	Late	3,356	67.91	0.19	-1.96	24.84	67.73
2	Male	All years	8,871	59.04	0.21	-2.34	26.83	58.98
3	Female	All years	1,640	70.15	1.29	2.41	61.09*	70.15
3	Male	All Years	1,328	69.21	1.18	2.32	59.72*	69.21
4	Female	All years	6,384	60.26	2.12	3.54	37.56*	60.26
4	Male	All years	3,671	74.66	0.66	1.93	55.49*	74.66
5	Female	All years	5,884	74.62	0.39	1.14	50.37	74.62
5	Male	All years	4,607	63.94	0.58	0.52	55.4	63.94

2.7 Figures

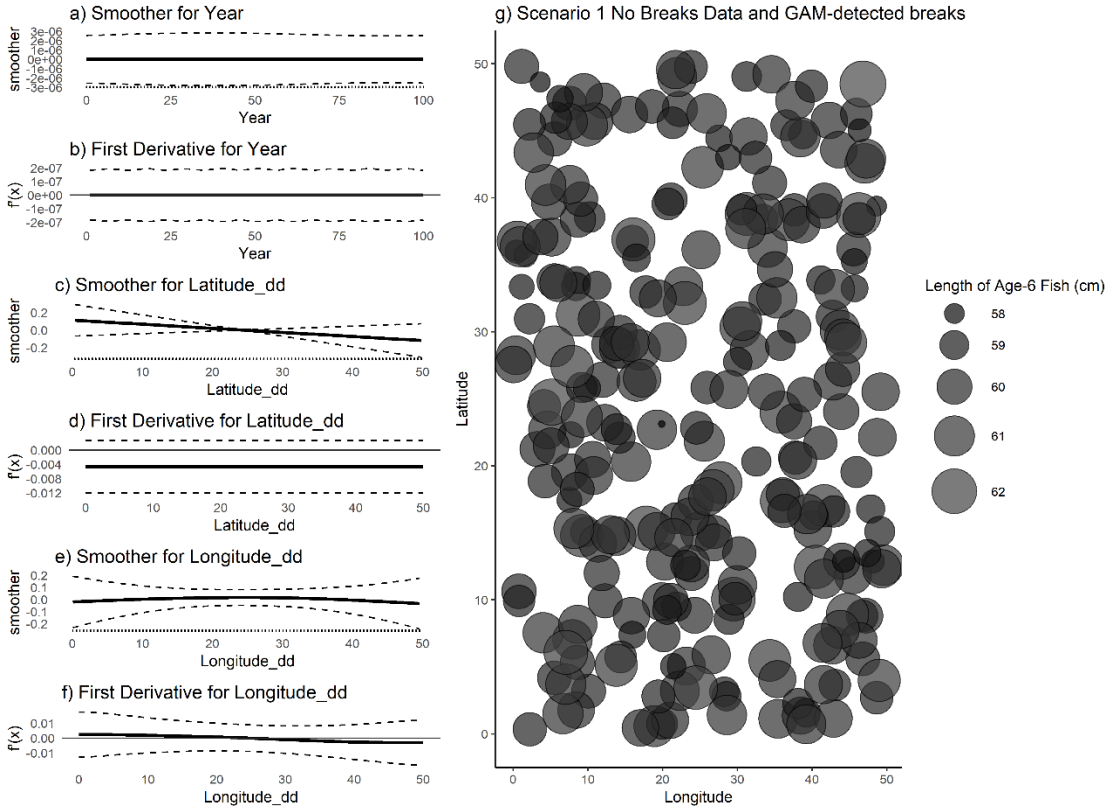


Figure 2.1: (a,c,e) raw value of GAM smoothers for Year, Latitude and Longitude; (b,d,f) mean (black line) and 95% CI (black dashed lines) of first derivative of the smoothers; (g) map of age-6 fish for a single simulated dataset with no designated spatial or temporal breaks. No break points were detected by the GAM.

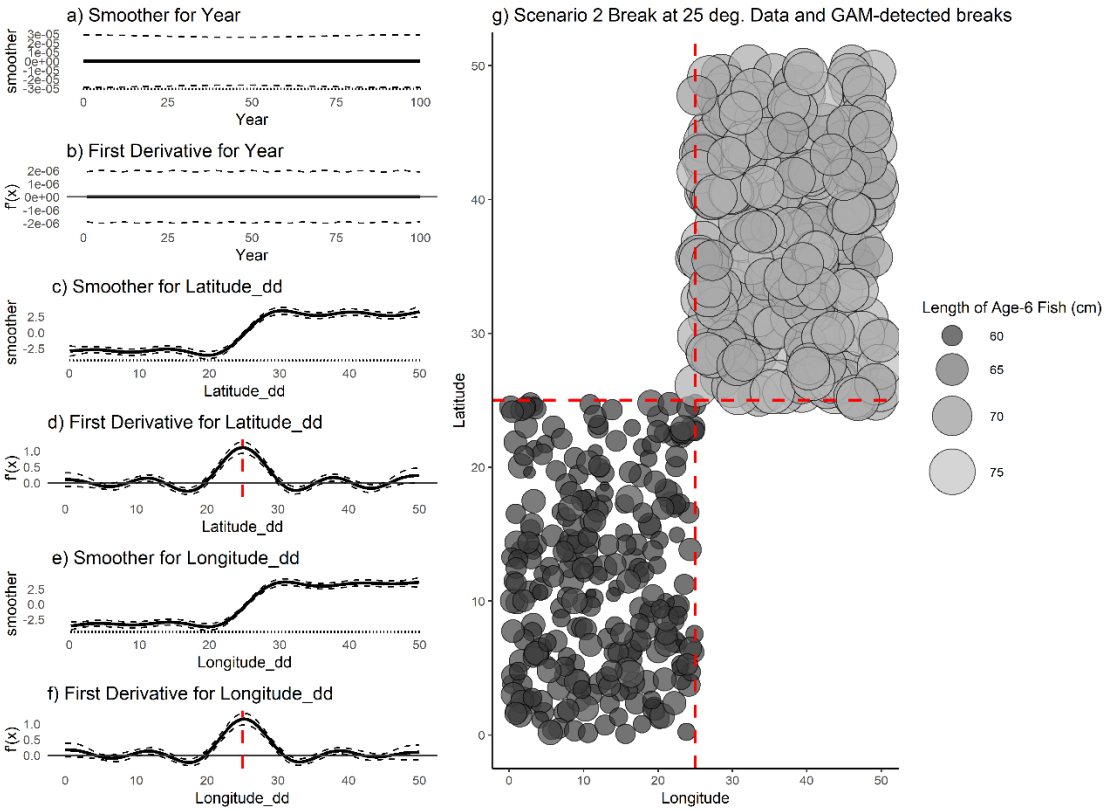


*Figure 2.2: (*a,c,e) raw value of smoothers (fitted regression splines) for year, latitude, and longitude; (b,d,f) mean (black line) and 95% CI (black dashed lines) of the first derivatives of the smoothers; (g) map of age-6 fish for a single simulated dataset with a single, symmetrical break at 25° latitude and longitude. Dashed red lines indicate detected break points, which are the maximum value obtained for this data set and do not have a confidence interval that contains zero.

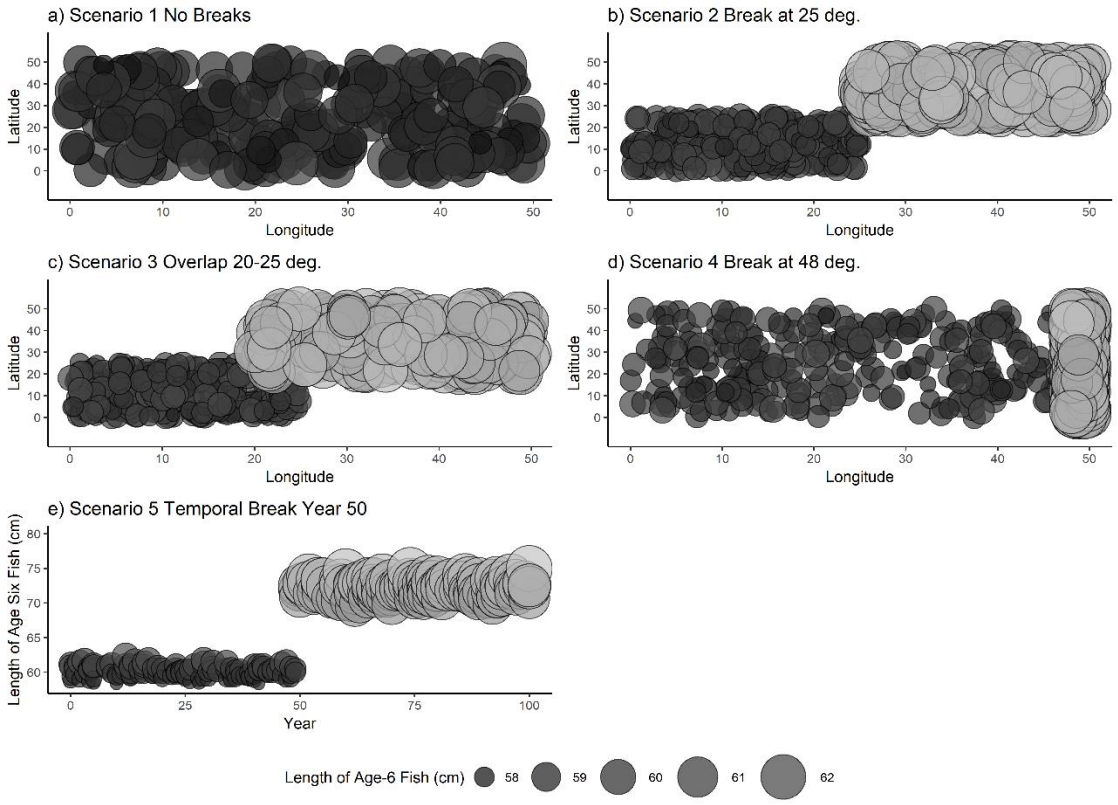


Figure 2.3: Example dataset for each of the scenarios in Table 2.2. For each of the five scenarios, points represent the length and location of a single simulated fish at age six. Fish locations (latitudes and longitudes) were sampled from a uniform distribution of the boundaries indicated in Table 2.2.

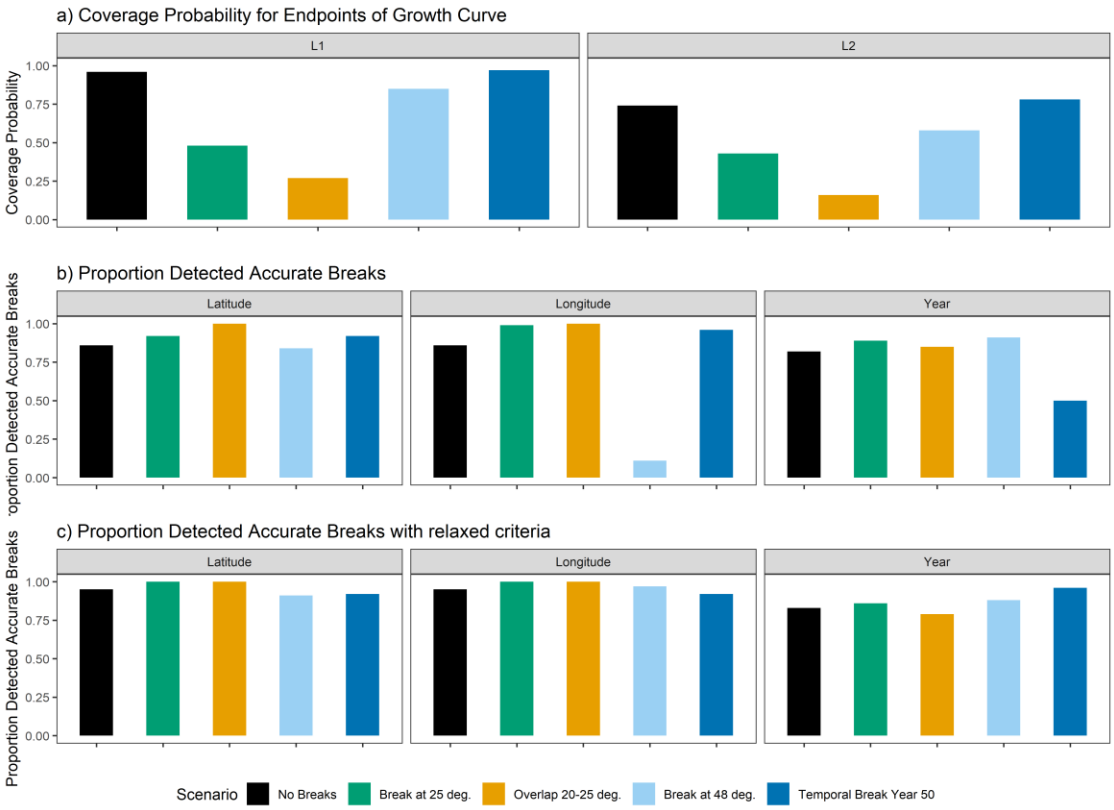


Figure 2.4: a) coverage probabilities for the endpoints of the growth curve, L_1 (left) and L_2 (right) ;b) proportion of 100 simulations for each spatial scenario wherein the correct latitudinal breaks (left), or longitudinal breaks (center) or temporal break (right) were detected, c) the same as b) but with the criteria for a 'match' relaxed to include breakpoints within two degrees or years of the truth.

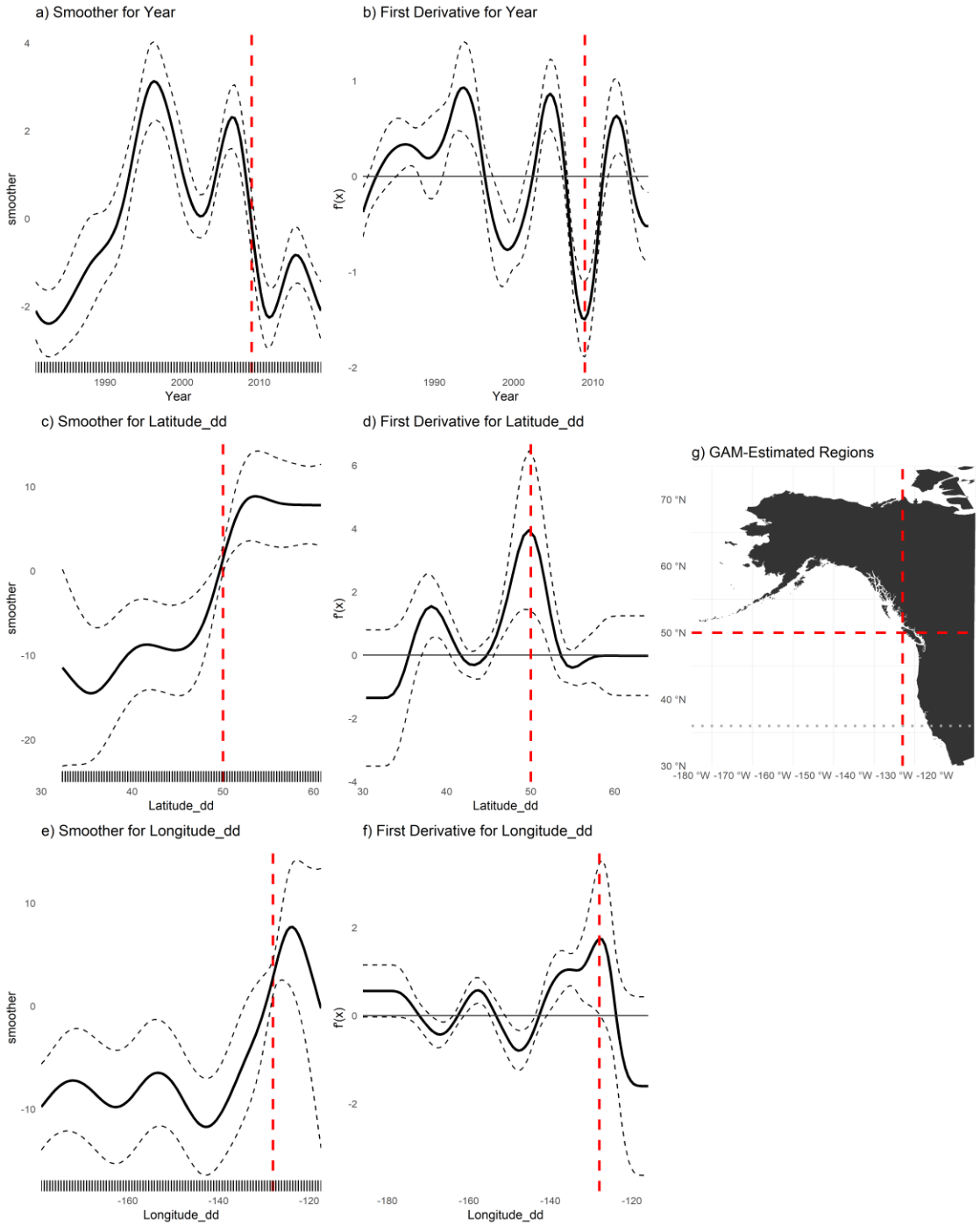


Figure 2.5: (a,c,e) Plots of smoothers (fitted regression splines) for year, latitude, and longitude, and first derivatives thereof for female age four sablefish (b,d,f). On a-f, vertical dashed lines indicate latitudes, longitudes or years that correspond to the highest first derivative and had a confidence interval that did not include zero. g) map with model-detected breakpoints (red dashed lines) and breakpoints detected for other ages (grey dotted line).

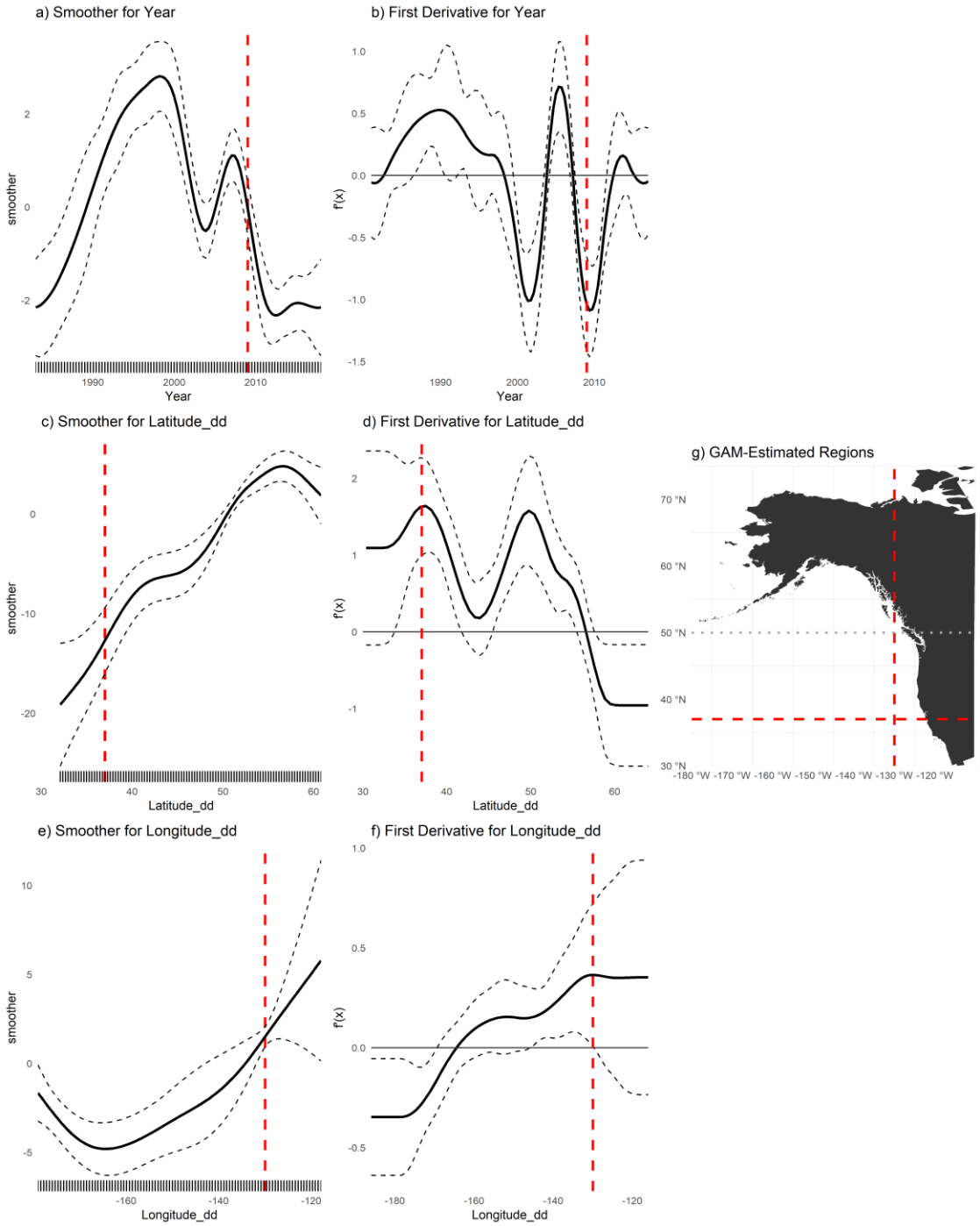


*Figure 2.6: (*a,c,e) Plots of smoothers (fitted regression splines) for Year, Latitude, and Longitude, and first derivatives thereof for female age six sablefish (b,d,f). On a-f, vertical dashed lines indicate latitudes, longitudes or years that corresponded to the highest first derivative and had a confidence interval that did not include zero. g) map with model-detected breakpoints (red dashed lines) and breakpoints detected for other ages (grey dotted line

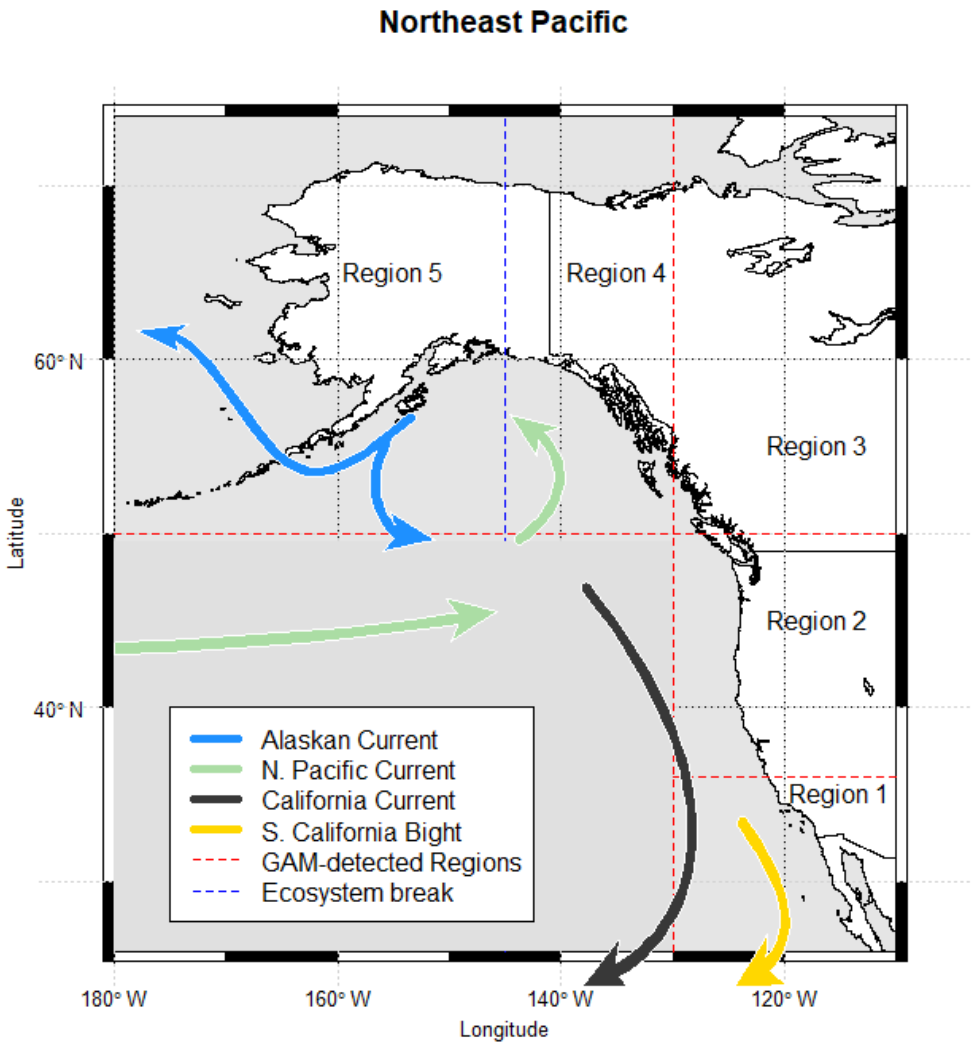


Figure 2.7: Method-detected breakpoints (red dashed lines) and an ecosystem-based break (blue dashed lines) used to delineate growth regions for sablefish. Map made in R using current data from: https://data.amerigeoss.org/en_AU/dataset/major-ocean-currents-arrowpolys-30m-85

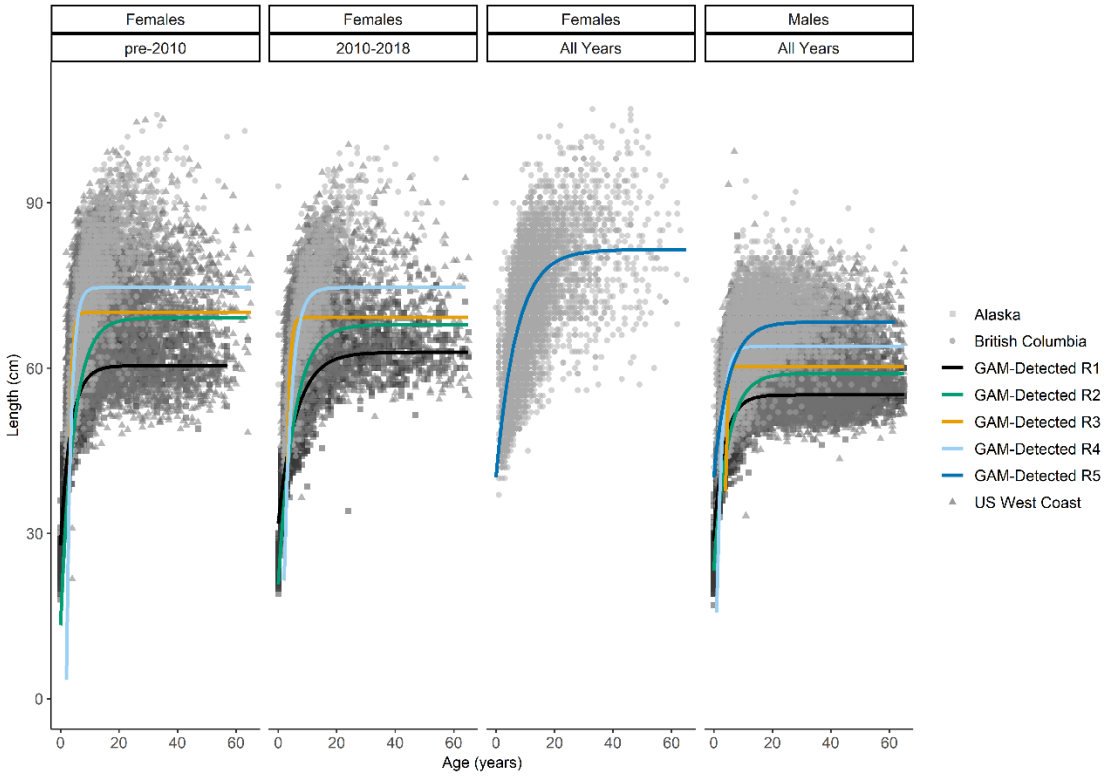


Figure 2.8: Fits of von Bertalanffy growth function (colored lines) to data at the final spatiotemporal aggregation (panels). Points are raw survey data with color and shape corresponding to their source.

Chapter 3. A TRANSBOUNDARY OPERATING MODEL FOR NORTHEAST PACIFIC SABLEFISH

3.1 Abstract

Sablefish (*Anoplopoma fimbria*) are a long-lived, valuable groundfish that have high movement rates, and range from Southern California to the Bering Sea. Synchronous sablefish population trends, including a general decline across the entire range during the past few decades, have increased concern about the status of the population. Traditionally, sablefish stock assessment and management has occurred independently at regional scales (Alaska, British Columbia, and the US West Coast), assuming these contain closed, independent stocks. However, recent genetic work has shown that northeast Pacific sablefish are not genetically distinct between these traditional management regions, although there is evidence for differences in growth rate and size-at-maturity across the range. This suggests that the current delineation of assessment and management regions may be incongruent with the stock structure. An operating model (OM) is developed to represent the dynamics of sablefish throughout the northeast Pacific for a hypothesized spatial structure, which includes four demographic stocks and six sub-areas, with movement among them. This OM is conditioned using observed fisheries landings, standardized indices of relative abundance, discard ratios, and age-composition data from each of the three management regions from 1960-2019. The OM is able to obtain satisfactory fits to most data sources and similar trajectories of spawning stock biomass as the regional assessments given the ability to estimate a time-series of stock-specific recruitment deviations, the allocation of recruitment among sub-areas, unfished recruitment, and a subset of fishery selectivity parameters. Sensitivity analyses suggest that movement rates (which are not estimated within the OM) are especially important to the maintenance of biomass for British Columbia, which is suspected to export at least one-third of adult biomass to adjacent areas on an annual basis. This inter-regional dependency underscores the importance of considering spatial stock structure in sablefish management, at the regional scale or otherwise. The OM can be used to generate data for inclusion in a management strategy evaluation of this important resource. It can also be adapted to explore other questions related to the spatial scale of sablefish dynamics or modified to produce similar analyses for other species.

3.2 Introduction

Sablefish (*Anoplopoma fimbria*) supports a valuable fishery in the northeast Pacific, but the estimation of population dynamics for management purposes is complicated by their wide distribution across management boundaries and ocean ecosystems. Sablefish are a long-lived groundfish that exhibit high movement rates and range from Southern California to the Bering Sea. Synchronous population trends, including a general decline across the entire range during the past two decades, have increased concern about the status of the sablefish population. Traditionally, sablefish stock assessment and management have occurred independently at regional scales (Alaska, British Columbia, US West Coast), assuming these contain closed stocks. However, recent genetic work has shown that northeast Pacific sablefish are not genetically distinct between these traditional management regions (Jasonowicz et al., 2016). This suggests that the current delineation of assessment and management regions may be incongruent with the stock structure.

Evidence for sablefish stock structure arises from differences in demographic rates throughout the northeast Pacific; sablefish growth and maturity schedules correlate strongly with oceanographic currents (*Chapter 2*; Williams (In Prep)), while the strength of recruitment in given years appears to correlate with sea-surface height more strongly in the northern California Current than in the southern (Tolimieri et al., 2018). Although tagging studies have revealed that adult sablefish can readily migrate the entire extent of the northeast Pacific on an annual basis (Hanselman et al., 2014), it appears that ecological gradients (namely in temperature) as well as the bifurcation of the northeast Pacific Current north and south of British Columbia contribute to roughly four demographic zones of sablefish growth (Chapter 2). There is a need to develop a spatially-structured population dynamics model for northeast Pacific sablefish to evaluate how spatial structure (including movement and demographic variation) may contribute to regional biomass trends and, ultimately, management performance given that the presence of various growth morphs in an assessed population can lead to reduced precision in management quantities (Punt et al., 2017), and the evidence suggesting sablefish biomass is highly mixed among management regions,.

This chapter presents the population dynamics equations, data inputs and treatment, and likelihood components included in such a model for northeast Pacific sablefish. The methods describe the

generalized model framework (which could theoretically be applied to any spatially-structured population with similar data), and then the parameterization and specific steps taken to condition the operating model (OM) using historical (1960-2019) data for sablefish. *Appendix C* provides further background material on 1) the conversion of movement rates-at-length to movement rates-at-age; 2) the development of novel spatiotemporal indices of relative sablefish abundance for the US West Coast and Alaska, and 3) a suite of sensitivity analyses used to evaluate major uncertainties related to using the operating model as the basis for a MSE, namely: movement rates, the inclusion of age-composition data, and the treatment of recruitment deviations. The objective of this chapter is to determine whether satisfactory fits to input data and similar trends in biomass to recent, regional assessments can be recovered from a spatially-structured statistical catch-at-age model, which would validate the possibility that sablefish indeed exhibit spatial structure in the northeast Pacific. The data types and/or model processes that the model struggles to fit or estimate, or those to which it is particularly sensitive, inform tradeoffs in the transition to a spatially-structured representation of this species and others like it. Finally, the terminal parameter values from the conditioned base OM, along with the correlation matrix of recruitment deviations among spatial areas, can be used to generate data for use in the management strategy evaluation in Chapter 4.

3.3 Methods

3.3.1 Model Overview

The operating model is meant to represent a hypothesis about the spatial structure of the sablefish population in the northeast Pacific, which is suspected to have geographically smaller demographic units (Kapur et al., 2020) and greater connectivity (Hanselman et al., 2014) than the prevailing management paradigm. The spatial structure of the OM must enable the calculation of population dynamics (e.g., spawning stock biomass, numbers-at-age) at the management level (m), while accounting for spatial variation in demography identified by life history analyses. In the OM, six modeled spatial areas are the union of demographic regions, or "stocks", and three political management regions (Figure 3.1). Sablefish management throughout the northeast Pacific benefits from regular age-, size- and sex- structured data collection, including fishery-independent survey programs in all three regions (Figure 3.2). The operating model is a two-sex, age-structured, spatially-structured population dynamics model, with an annual timestep from 1960-2019 (the

"historical model period"), and intra-year dynamics accounted for using three "seasons". Ages range from age-0 to a plus-group age A (in whole years), and length zero through a plus-group length l (in cm); numbers- and biomass-at-age greater than age A or numbers- and biomass-at-length greater than L collapse into these terminal groups, after which growth and selectivity are assumed constant. Fish may move among the spatial areas. The OM is coded in Template Model Builder (Kristensen et al., 2016), and many elements are similar in structure to the widely-used statistical catch-at-age modeling software platform Stock Synthesis (Methot and Wetzel, 2013). Spatial areas, also termed "sub-areas", are the smallest unit at which population dynamics are tracked, and each spatial area contains animals from only one stock, and is part of only one management region. Stocks are sets of spatial areas with the same demographic parameters, including those for growth and maturity. In equations, sub-areas are indexed using the letters i and j, stocks using the letter k, and management regions using the letter m. Density-dependence (recruitment following a Beverton-Holt stock-recruitment relationship) occurs at the stock level. Fish that move to a new stock assume the demographic characteristics of that stock and contribute to the spawning biomass for that stock. The proportion of recruits assigned to each stock's constituent spatial area(s) is defined by an annual parameter. Movement rates at age are specified among spatial areas based on the results of a tag-recapture study. The following sections describe the equations used to represent the population and fishery dynamics, and to condition the OM[2]. The first sections describe the general model structure, and the later sections indicate the specific parameterization used for northeast Pacific sablefish.

3.3.2 Population Dynamics (Numbers at Age)

The OM operates on a yearly timestep, assuming an equal sex ratio of age-0 recruits within each sub-area (Equation 3.1). The calculation of numbers-at-age in 1960 $\left(N^{i}_{B0,\gamma,a}\right)$ is described in Section 3.3.8. There are three "seasons" within each year to enable the tracking of natural mortality, fishery exploitation, and the movement of fish among areas (which may have different maturity and growth schedules). The "beginning" season corresponds to the start of the year, when the numbers-at-age from the end of the previous year are incremented by one and recruits for that

[2] "Conditioning" is the process of specifying the values of parameters of an operating model – it is analogous to fitting the operating model to the available data but not identical as several parameter values are pre-specified based on auxiliary analyses.

year are assigned to age zero. The mid-year numbers-at-age are defined once fish grow in length and move among sub-areas and the entire population experiences half of the annual natural and fishing mortality; this is the population available to survey sampling. The second half of fishery-specific exploitation, plus half of annual mortality, is then applied to produce the "end" of the year numbers-at-age.

The end-year numbers-at-age after movement and all mortality inform the spawning stock biomass (SSB) and recruitment for the subsequent year. For ages 1 through A, the number of individuals in sub-area i of sex γ and age a at the start of year $y+1$ is the sum of individuals in sub-area i that have survived and did not emigrate the year prior, or those that have survived and immigrated into sub-area i from another sub-area j. The plus group in sub-area i is comprised of individuals that are at least of age A and remain in sub-area i, plus individuals that are of at least age A and have moved to sub-area i from sub-area j ($i \in j$). The basic dynamics of fish of age a and sex γ during year y in sub- area i within the modeled period are given by:

$$N^i_{B_{y,\gamma,a}} = \begin{cases} 0.5R^i_y & if\ a = 0 \\ N^i_{E_{y-1,\gamma,a-1}} & if\ 1 \le a < A \\ N^i_{E_{y-1,\gamma,A-1}} + N^i_{E_{y-1,\gamma,A}} & if\ a = A \end{cases} \tag{3.1a}$$

$$N^i_{\mathrm{M}_{y,\gamma,a}} = \begin{cases} N^i_{\mathrm{B}_{y,\gamma,0}} e^{-M^k/2} & \text{if } a = 0 \\ (1 - \sum_{i \neq j} \boldsymbol{X}^{i,j}_a) N^i_{\mathrm{B}_{y,\gamma,a}} e^{-M^k/2} \left(1 - \sum_f \phi^{if} s^f_{\mathrm{B}_{y,\gamma,a}} F^{1f}_y\right) \dots & if\ 1 \le a \le A \\ + \sum_{j \neq i} \boldsymbol{X}^{j,i}_a N^j_{\mathrm{B}_{y,\gamma,a}} e^{-M^k/2} \left(1 - \sum_f \phi^{jf} s^f_{\mathrm{B}_{y,\gamma,a}} F^{1f}_y\right) \end{cases} \tag{3.1b}$$

$$N^i_{E_{y,\gamma,a}} = (1 - \sum_f \phi^{if} s^f_{M_{y,\gamma,a}} F^{2f}_y) N^i_{M_{y,\gamma,a}} e^{-M^k/2} \qquad for\ 0 \le a \le A \tag{3.1c}$$

where $N^i_{B|M|E_{y,\gamma,a}}$ is the number of animals of age a and sex γ in sub-area i at the beginning B, middle M, or end E of year y; $\mathbf{X}^{i,j}_a$ is the age-specific matrix of movement probabilities from sub-area i to sub-area j; M^k is the stock-specific rate of natural mortality (sub-area i is nested within stock k); $s^f_{B|M|E_{y,\gamma,a}}$ is year-, sex-, age- and fleet-specific fishery selectivity at the beginning B, middle M, or end E of year y; F^f_y is the discretized fishing mortality rate on fully-recruited animals by fishery f during year y; a superscript of 1 corresponds to the first half of the year, and a superscript of 2 to the second half (see Section 3.3.5.3 for more details on how total mortality is

calculated); ϕ^{if} is 1 if fishery f occurs in sub- area i, and 0 otherwise; R_y^i is the number of recruits to sub-area i at the start of year y (see Section 3.3.4); and A is the plus-group age.

3.3.3 Growth

3.3.3.1 Length-at-age

Length-at-age is sub-area- and sex-specific based on expected growth increments from a von Bertalanffy growth function (Bertalanffy, 1938). One challenge in modeling size-at-age where growth varies spatially is that models can sometimes predict a decline in length as an individual moves from one habitat to another. Empirical observation and theory indicate that length is generally preserved despite changes in somatic growth rate (Shelton and Mangel, 2012; Stawitz et al., 2015). The growth model therefore accounts for larger fish moving into a sub-area of mostly smaller fish by tracking the mean size-at-age by sub-area, as a function of the numbers-at-age of fish with distinct growth schedules, and by modeling future fish growth in each sub-area using an incremental approach, preventing fish from shrinking upon arrival in a new sub-area. However, because fish can move among sub-areas with different growth patterns throughout their lifetime, there is a possibility that the mean length-at-age within a sub-area could decline due to a large influx of smaller-at-age fish (see below). The growth equations are indexed by sub-area i although growth depends on stock (k) not sub-area (i). The mean length-at-age a at the beginning (B) of year 0 (1960) for sex γ in sub-area i is given by:

$$L^i_{\mathrm{B}_{0,\gamma,a}} = L^i_{\infty_{1,\gamma}} + (L^i_{1,1,\gamma} - L^i_{\infty_{1,\gamma}}) \left(e^{-0.5\kappa^i_{1,\gamma}(a-4)}\right) \quad (3.2)$$

where $L^i_{1,1,\gamma}$ is the sub-area- and sex-specific length (cm) at age 4 in 1960 (growth between ages zero and 4 is assumed to be linear). This value is always lower than L^i_{∞}; $L^i_{\infty,1,\gamma}$ is the sub-area- and sex-specific asymptotic length (cm) for sub- area i in 1960; and $\kappa^i_{1,\gamma}$ is the sub-area and sex-specific growth rate (cm yr^{-1}) for sub- area i during 1960.

The mean mid-year (M) length-at-age a for animals of sex γ in sub-area i during year y is given by incrementing half of annual growth immediately before movement. All individuals in a sub-area are subject to the growth pattern of the stock associated with that sub-area. This enables the modeling of ecosystem-based effects on the growth process, where a fish born in a southerly (slow-

growing) region may grow to a greater size than expected based on its birth location if it moves to a northerly (fast-growing) region, and vice-versa.

To account for movement from one stock to another, the mid-year mean size of fish at age in a recipient sub-area becomes the weighted average of fish length-at-age remaining in the sub-area and the length-at-age of fish that underwent growth just before entering the sub-area from a different sub-area and stock. The growth pattern of recruits is not tracked throughout their lives. This calculation therefore applies to all ages.

$$L^{i}_{\mathrm{M}_{y,\gamma,a}} = \frac{X^{i,i}_{a} N^{i}_{\mathrm{B}_{y,\gamma,a}} \left[L^{i}_{\mathrm{B}_{y,\gamma,a}} + (L^{i}_{\infty_{y,\gamma}} - L^{i}_{\mathrm{B}_{y,\gamma,a}}) \left(1 - e^{-0.5\kappa^{i}_{y,\gamma}} \right) \right] + \sum_{i \neq j} X^{j,i}_{a} N^{j}_{\mathrm{B}_{y,\gamma,a}} \left[L^{j}_{\mathrm{B}_{y,\gamma,a}} + (L^{j}_{\infty_{y,\gamma}} - L^{j}_{\mathrm{B}_{y,\gamma,a}}) \left(1 - e^{-0.5\kappa^{j}_{y,\gamma}} \right) \right]}{X^{i,i}_{a} N^{i}_{\mathrm{B}_{y,\gamma,a}} + \sum_{j \neq \mathrm{i}} X^{j,i}_{a} N^{j}_{\mathrm{B}_{y,\gamma,a}}} \quad (3.3)$$

In common with mid-year length-at-age, end-year length-at-age is defined by the completion of the second half of annual growth:

$$L^{i}_{E_{y,\gamma,a}} = L^{i}_{M_{y,\gamma,a}} + \left(L^{i}_{\infty_{y,\gamma}} - L^{i}_{M_{y,\gamma,a}} \right) \left(1 - e^{-0.5\kappa^{i}_{y,\gamma}} \right) \qquad (3.4)$$

For $a < A$, end-year lengths are carried over to the next age at the start of the next year. For the plus group (age A) the mean size $y + 1$ is calculated as the weighted average of fish ageing into and/or remaining in the plus-group within the sub-area. This allows the size of fish in the plus group to reflect changes in numbers due to exploitation and/or natural mortality (Methot and Wetzel, 2013).

$$L^{i}_{B_{y+1,\gamma,A}} = \begin{cases} L^{i}_{E_{y,\gamma,a-1}} & if\ a < A \\ \dfrac{N^{i}_{E_{y,\gamma,A-1}} L^{i}_{E_{y,\gamma,A-1}} + N^{i}_{E_{y,\gamma,\mathrm{A}}} L^{i}_{E_{y,\gamma,a}}}{N^{i}_{E_{y,\gamma,A-1}} + N^{i}_{E_{y,\gamma,A}}} & if\ a = A \end{cases} \qquad (3.5)$$

3.3.3.2 Converting from lengths to ages

Certain model processes require translating between age and length. In a spatially unstructured model this calculation is readily conducted based on the generating equations, but is complex to conduct directly when there is movement among regions that have different growth rates. This difficulty is resolved via a matrix of mean length-at-age for each year, sex and sub-area. The matrix is calculated by conducting initial model runs of the OM given the input growth parameters and putative movement rates (which are set to values from external analysis and not estimated within

the OM). These simulated lengths-at-age a for fish are rounded to integers $L^i_{B_{y,\gamma,a}}$ and $L^i_{M_{y,\gamma,a}}$ (beginning-of-year and mid-year lengths, respectively). This provides consistency with the length bin structure of the model, with centimeter level precision as in the regional assessments (and also reduces computing time). These values are used to look up the relevant length-based value (retention rate, selectivity) for a fish of age a, sex γ in sub-area i at the relevant part of year y and are preferable to a deterministic length-at-age key, as this method accounts for changes in length due to movement and exploitation. Movement rates-at-age are defined by the expected length-at-age and uncertainty thereof, calculated externally to the OM (see *Appendix C* for more details).

3.3.3.3 Weight-at-age

Body weight is calculated using length-at-age at any part of the year:

$$w^i_{B|M|E_{y,\gamma,a}} = \alpha^i_\gamma \left(L^i_{B|M|E_{y,\gamma,a}}\right)^{\beta^i_\gamma} \qquad (3.6)$$

where α^i_γ and β^i_γ are the sub-area- and sex-specific constants of the allometric length-weight equation (typically shared within stocks); and $L^i_{y,\gamma,a}$ is the sub-area-, year-, and sex-specific mean length corresponding to age a at the beginning, middle, or end of the year (see previous section).

3.3.4 Reproduction or recruitment

Recruitment follows a stock-specific Beverton-Holt (Beverton and Holt, 1957) stock-recruitment curve with annual deviations. Density-dependence is assumed to occur at the level of the stock and is determined by the biomass (converted from numbers, as described above) of mature females, which may be the sum over two or more sub-areas, at the end of the previous year. The stock-recruitment relationship assumes that recruitment deviations occur at the stock level (and are thus equivalent among sub-areas within a stock).

$$R^k_{y+1} = \frac{4h^k R^k_0 S^k_y}{S^k_0(1-h^k) + S^k_y(5h^k - 1)} e^{-0.5\sigma^2_R + \tilde{R}^k_y} \qquad (3.7)$$

where h^k is the steepness of the stock-recruitment curve for stock k (expected proportion of R^k_0 at $0.2S^k_0$); R^k_0 is the unfished ("virgin") recruitment for stock k; $\tilde{R}^k_y$ are random annual recruitment deviations specific to stock k (but perhaps correlated among stocks) and assumed to be normally

distributed with mean zero and standard deviation σ_R; and S_y^k is the biomass of mature females in stock k at the end of year y. There is never more than one stock (k) in a sub-area (i):

$$S_y^k = \sum_{i,a} \Lambda^{ik} N^i_{B_{y,\gamma=female,a}} w^k_{B_{y,\gamma=female,a}} E_a^k \qquad (3.8)$$

$\mathbf{\Lambda}^{ik}$ is a matrix indicating whether sub-area i is nested within stock k; E_a^k is the proportion of females in stock k of age a that have reached maturity by age a; and S_0^k is the unfished female spawning biomass of stock k (Equation 3.25).

Recruits are proportionally allocated from stock k to sub-area i. The resulting annual recruitment spawned within each sub-area is used in Equation 3.1.

$$R_y^i = \tau_y^{ik} R_y^k \qquad (3.9)$$

where τ_y^{ik} is the proportion of recruits to stock k that recruit to sub-area i during year y. Spatial autocorrelation is anticipated in the number of recruits 1) in sub-areas within the same stock and 2) among stocks linked by movement, or those that share data inputs.

3.3.5 Fisheries

A "fishery" or "fishery fleet" is a discrete fishing operation. There are several fleets f in each management region m, and there may be more than one fishery operating in each sub-area. All management regions have at least one fishery and one survey and have age-composition data from one or more fisheries (Figure 3.2). Temporal coverage varies by management region, with catch records extending back to the early 1900s for Alaska and the US West Coast (Table 3.1), although the modeled period within the OM starts in 1960.

3.3.5.1 Catches

The catch used to condition the OM are aggregated at the fishery level f, which are nested within management regions m and a fishery may hence occur in more than one sub-area i.

3.3.5.2 Retention Curves

Ω is a logistic function defining the probability of animals of length l and sex γ being retained by fishery f during year y ($1 - \Omega$ is the probability of being discarded):

$$\Omega^{f}_{y,\gamma,l} = \beta_3^{y,f,\gamma,l}\left(1 + exp(-\left(l - \left(\frac{\beta_1^{y,f,\gamma,l} + \beta_4^{y,f,\gamma,l}}{\beta_2^{y,f,\gamma,l}}\right)\right)\right)^{-1} \qquad (3.10)$$

where $\beta_1^{y,f,\gamma,l}$ is the length at the inflection point of the logistic retention function; $\beta_2^{y,f,\gamma,l}$ is the slope at the point of inflection; $\beta_3^{y,f,\gamma,l}$ is the asymptotic fraction retained; and $\beta_4^{y,f,\gamma,l}$ is 0 for females and an arithmetic offset for males. Because the model requires the use of discards-at-age, $\Omega^{f}_{y,\gamma,l}$ is passed through an age-length key to extract the applicable retention value by age (see Section 3.3.3.2), so the values for Ω will vary at the beginning, middle and end of the year as the mean length-at-age of fish in each sub-area changes subject to growth and movement.

3.3.5.3 Exploitation rate

The bi-annual discretized fishing mortality for each fishery $F_y^{1/2f}$ is determined using Pope's approximation. This is the ratio of the fishery's landed dead catch to the total vulnerable, retained biomass available to fishery f at either the beginning or midpoint of year y. In the case where fishery f fishes in more than one sub-area, the total exploitable biomass is the biomass from all sub-areas where fishery f operates. Note that F_y^{1f} is assumed to occur before movement and natural mortality.

$$F_y^{1f} = \frac{\frac{C^{f}_{obs,y}}{2}}{\sum_{i,\gamma,a} \phi^{if} w^{f}_{B_{y,\gamma,a}} s^{f}_{B_{y,\gamma,a}} \Omega^{f}_{B_{y,\gamma,a}} N^{i}_{B_{y,\gamma,a}} e^{-M^{k}/4}} \qquad (3.11)$$

$$F_y^{2f} = \frac{\frac{C^{f}_{obs,y}}{2}}{\sum_{i,\gamma,a} \phi^{if} w^{f}_{M_{y,\gamma,a}} s^{f}_{M_{y,\gamma,a}} \Omega^{f}_{M_{y,\gamma,a}} N^{i}_{M_{y,\gamma,a}} e^{-M^{k}/4}} \qquad (3.12)$$

where $\Omega^{f}_{B|M_{y,\gamma,a}}$ is the length-based retention curve at the beginning or middle of the year, converted to age-based retention using the age-length lookup key (see Section 3.3.3.2); and $C^{f}_{obs,y}$ is the observed catch for fleet f during year y.

3.3.5.4 Retained Catch

The bi-annual predicted catches $C_{pred,y}^{1/2f}$ are calculated by applying the Fs for the appropriate time step to the relevant numbers-at-age. The sum of these two quantities is the total annual catch for fishery f during year y. Predicted discarded catch is calculated identically, except that $(1 - \Omega_{y,\gamma,a}^{f})$ is used instead of $\Omega_{y,\gamma,a}^{f}$.

$$C_{pred,y}^{1f} = \sum_{i,\gamma,a} \phi^{if}\, s_{B_{y,\gamma,a}}^{f} F_{y}^{1f} w_{B_{y,\gamma,a}}^{f} \Omega_{B_{y,\gamma,a}}^{f} N_{B_{y,\gamma,a}}^{i} e^{-M^{k}/4} \qquad (3.13)$$

$$C_{pred,y}^{2f} = \sum_{i,\gamma,a} \phi^{if}\, s_{M_{y,\gamma,a}}^{f} F_{y}^{2f} w_{M_{y,\gamma,a}}^{f} \Omega_{M_{y,\gamma,a}}^{f} N_{M_{y,\gamma,a}}^{i} e^{-M^{k}/4} \qquad (3.14)$$

The annual retained catch for each fleet in each management region $C_{y}^{m,f}$ is obtained by summing the catches during the first and second halves of the year by fleet and sub-area for the sub-areas in which the fleet concerned operates. No fleet fishes in more than one management region.

$$C_{y}^{m,f} = \sum_{f \in m} (C_{pred,y}^{2f} + C_{pred,y}^{1f}) \qquad (3.15)$$

3.3.5.5 Discard Ratios

Predicted discarded catch by weight $\widetilde{D}$ is converted to predicted discard ratios by dividing the total annual discarded catch by the total landed (retained + discarded) catch for fleet f in year y.

$$\widetilde{D}_{pred,y}^{1f} = \sum_{i,\gamma,a} \phi^{if} s_{B_{y,\gamma,a}}^{f} F_{y}^{1f} w_{B_{y,\gamma,a}}^{f} \left(1 - \Omega_{B_{y,\gamma,a}}^{f}\right) N_{B_{y,\gamma,a}}^{i}\, e^{-M^{k}/4} \qquad (3.16)$$

$$\widetilde{D}_{pred,y}^{2f} = \sum_{i,\gamma,a} \phi^{if} s_{M_{y,\gamma,a}}^{f} F_{y}^{2f} w_{M_{y,\gamma,a}}^{f} \left(1 - \Omega_{M_{y,\gamma,a}}^{f}\right) N_{M_{y,\gamma,a}}^{i}\, e^{-M^{k}/4} \qquad (3.17)$$

$$D_{pred,y}^{f} = \frac{\widetilde{D}_{pred,y}^{1f} + \widetilde{D}_{pred,y}^{2f}}{\widetilde{D}_{pred,y}^{1f} + \widetilde{D}_{pred,y}^{2f} + C_{pred,y}^{1f} + C_{pred,y}^{2f}} \qquad (3.18)$$

3.3.6 Selectivity (fisheries and surveys)

Selectivity can be age- or length-specific, and vary by sex and through time. Selectivity curves can be fully-selected (1.0 for all ages/lengths) or follow an asymptotic, power function, double-normal, or "dome-shaped" selectivity pattern, the latter described by either a normal or gamma distribution. The equations below use the "a notation", but l is used for the British Columbia fisheries (see Section 3.3.3.2 for the conversion between lengths and ages).

Asymptotic selectivity is defined as follows:

$$s_{\gamma,a}^{f} = \left(1 + e^{-\left(a - a_{50,\gamma}^{f}\right)/\delta_{\gamma}^{f}}\right)^{-1} \qquad (3.19)$$

where $a_{50,\gamma}^{f}$ is the age-at-50%-selectivity; and δ_{γ}^{f} is a parameter of the asymptotic curve. The inverse power function is defined by a parameter δ_{γ}^{f}:

$$s_{\gamma,a}^{f} = \frac{a^{\rho_{\gamma}^{f}}}{max\left(a^{\rho_{\gamma}^{f}}\right)} \qquad (3.20)$$

Dome-shaped selectivity is defined either by specifying a fleet- and sex-specific mean μ_{γ}^{f} and standard deviation σ_{γ}^{f}:

$$s_{\gamma,a}^{f} = e^{-0.5(a-\mu_{\gamma}^{f})/{\sigma_{\gamma}^{f}}^{2}} \qquad (3.21)$$

or the shape (α_{γ}^{f}) and rate (β_{γ}^{f}) parameters of the gamma distribution:

$$s_{\gamma,a}^{f} = a^{\left(\alpha_{\gamma}^{f}-1\right)}e^{-a/\beta_{\gamma}^{f}} \qquad (3.22)$$

$$s_{\gamma,a}^{f} = s_{a,\gamma}^{f}/max(s_{a,\gamma}^{f}) \qquad (3.23)$$

The Supplementary Material in Methot and Wetzel (2013) provide details on the parameterization of the double-normal functional form. The OM only estimates the ascending and descending limb parameters for each sex for fisheries that use the double normal functional form (to reduce the number of estimable parameters).

3.3.7 Equilibrium abundance

The unfished age and sex distribution by sub-area is calculated at the stock level and partitioned into sub-area. Matrices **A**, **X**, and **S** are each $n \times A$ squares.

$$\widetilde{\boldsymbol{N}} = (\boldsymbol{I} - \boldsymbol{AXS})^{-\mathbf{1}} 0.5 R_0^k \tau^{ik} \qquad (\mathbf{3.24})$$

where I is the identity matrix; S is a matrix of stock-specific (annual) natural survivorship; X is a matrix of proportional movement-at-age among sub-areas; A is a matrix indicating the incrementation (transition) of fish between age-classes, which is 1 for each one-year increment; and $\boldsymbol{\tau}$ is a matrix specifying the distribution of recruits in stock k to sub- area i; if time-varying, the first year values for $\boldsymbol{\tau}$ are applied.

Unfished equilibrium spawning stock biomass is defined as:

$$\widetilde{\boldsymbol{N}} = (\boldsymbol{I} - \boldsymbol{AXS})^{-\mathbf{1}} 0.5 R_0^k \tau^{ik} \qquad (3.25)$$

3.3.8 Initial conditions

The model is initialized using the equilibrium distribution, penalized recruitment deviations, and an initial estimated fleet-specific rate F_{init}^f to account for the fact that fishing occurs in at least one sub-area before the model start year; selectivity patterns during the start year apply during this initial calculation.

$$N_{init,\gamma,a}^i = \begin{cases} \widetilde{N}_{\gamma,a}^i \left(1 - \sum_f \phi^{if} s_{\mathrm{B}_{y,\gamma,a}}^f F_{init}^f\right) e^{-M^k} e^{-0.5\sigma_R^2 + \tilde{R}_{init}^k} & if\ a < A \\ \dfrac{\widetilde{N}_{\gamma,A-1}^i \left(1 - \sum_f \phi^{if} s_{\mathrm{B}_{y,\gamma,A-1}}^f F_{init}^f\right) e^{-M^k}}{1 - \left(1 - \sum_f \phi^{if} s_{\mathrm{B}_{y,\gamma,A}}^f F_{init}^f\right) e^{-M^k}} e^{-0.5\sigma_R^2 + \tilde{R}_{init}^k} & if\ a = A \end{cases} \qquad (3.26)$$

where $\tilde{R}_{init}^k$ is the stock-specific initial recruitment deviate.

3.3.9 Data Generation

3.3.9.1 Surveys

Surveys occur at the midpoint of the year. The operation of some surveys spans more than one sub-area, in which case the predicted biomass related to survey f in year y ($B^f_{Pred,y}$) is computed as the sum across sub-areas.

$$B^f_{Pred,y} = q^f_y \sum_a \sum_i \sum_\gamma \phi^{if} w^f_{\mathrm{M}_{y,\gamma,a}} N^i_{\mathrm{M}_{y,\gamma,a}} s^f_{\mathrm{M}_{y,\gamma,a}} \qquad (3.27)$$

where q^f_y is the catchability coefficient for survey f during year y; ϕ^{if} is a matrix defining whether survey f operates in sub- area i; and $s^f_{M_{y,\gamma,a}}$ is the selectivity for survey f at age a and sex $\boldsymbol{\gamma}$ at the midpoint of year y.

3.3.9.2 Age compositions and ageing error

The expected proportion of observed numbers-at-age must account for imprecision and bias in ageing due to otolith reads occurring at different labs and/or by different readers. The ageing error matrix converts true ages a into expected ages $\tilde{a}$, the expected age value at age a and sex γ at the midpoint of year y (when samples are obtained). First, a distribution of expected ages (which may differ from true age) is defined as:

$$P^f_{\gamma,a,\tilde{a}} = \begin{cases} \Phi(\theta_2, a, \sigma^f_{a\,\gamma}) & if\ \tilde{a} = 1 \\ \Phi(\theta_{a+1}, a, \sigma^f_{a\,\gamma}) - \Phi(\theta_a, a, \sigma^f_{a\,\gamma}) & if\ 1 < \tilde{a} < A \\ 1 - \Phi(A, a, \sigma^f_{A\,\gamma}) & if\ \tilde{a} = A \end{cases} \qquad (3.28)$$

where $P^f_{\gamma,a,\tilde{a}}$ is the probability of an animal of sex γ and true age $\tilde{a}$ is observed as age $\tilde{a}$ from data collected by fleet f (i.e., the ageing error matrix); $\tilde{a}$ is the estimated age a incremented to mid-year values by adding 0.5; Φ is the cumulative normal density function; θ_a is the lower limit of age bin a; and $\sigma^f_{a\,\gamma}$ is the fishery-, age- and sex-specific standard deviation of observed ages.

The expected catch-at-age for fishery (Equation 3.29) or survey (Equation 3.30) f during year y (expressed as proportions) is calculated by multiplying the midyear catch numbers-at-age by the ageing error matrix and dividing by the total numbers-at-age vulnerable to fishery f:

$$\rho_{y,\gamma,a}^{f=fishery} = \frac{\sum_i \phi^{if} P_{\gamma,a}^{f} s_{\mathrm{M}_{y,\gamma,a}}^{f} F_y^{2f} \left(1 - \Omega_{\mathrm{M}_{y,\gamma,a}}^{f}\right) N_{\mathrm{M}_{y,\gamma,a}}^{i}}{\sum_i \sum_a \phi^{if} P_{\gamma,a}^{f} s_{\mathrm{M}_{y,\gamma,a}}^{f} F_y^{2f} \left(1 - \Omega_{\mathrm{M}_{y,\gamma,a}}^{f}\right) N_{\mathrm{M}_{y,\gamma,a}}^{i}} \qquad (3.29)$$

$$\rho_{y,\gamma,a}^{f=survey} = \frac{\sum_i \phi^{if} P_{\gamma,a}^{f} s_{\mathrm{M}_{y,\gamma,a}}^{f} N_{\mathrm{M}_{y,\gamma,a}}^{i}}{\sum_i \sum_a \phi^{if} P_{\gamma,a}^{f} s_{\mathrm{M}_{y,\gamma,a}}^{f} N_{\mathrm{M}_{y,\gamma,a}}^{i}} \qquad (3.30)$$

3.3.10 Operating Model development for transboundary ablefish

The following sections describe the data sources and parameterization of the OM described above for the northeast Pacific Sablefish population. The biological stocks for the sablefish application are defined by distinct demographic regimes, with growth described by *Chapter 2*, and movement among sub-areas by Rogers et al. (In Prep). The model period is from 1960-2019. The plus-group age A for the OM is 71 years. The plus-group length bin is 81cm. Input parameters, estimation bounds, and the data used when conditioning the OM are shown in Tables 3.1 and 3.2.

3.3.10.1Demographic parameters

3.3.10.1.1 Spatial Structure

The northeast Pacific sablefish population represented in this analysis occupies the exclusive economic zones of the United States and Canada from the Mexico-California border to the Aleutian Islands (Figure 3.1). The second chapter of this dissertation identified five spatial areas between which sablefish length-at-age differs significantly. After the completion of that work, it was decided that the OM should contain four major demographic regions, especially since "Region 3" from that analysis (representing a small area off the British Columbia coastline north of the US-Canada border and east of 135°W) would have had very little area-specific data and present estimation difficulties. Instead, the detected breakpoint at 50°N was maintained, such that the British Columbia coastline is split into two sub-areas. Input growth parameters were re-estimated at the appropriate spatial strata using the same data inputs and methods described in *Chapter 2* (Figure 3.3).

The demographic regions (stocks, k) do not coincide with political boundaries, so the smallest spatial unit of the OM is defined by the six sub-areas (i) that represent the intersection of the four stocks and three management regions (Figure 3.1). To aid in translation between these nested

spatial structures, and to ensure that fisheries and surveys operate in the correct regions, presence-absence matrices were defined to represent the intersection of each set (sub-areas, stocks, and management regions). For example, the matrix $\mathbf{\Lambda}$ indicates whether sub-area i is nested within stock k (Figure 3.1):

$$\vec{\imath} \in \vec{k} = \begin{cases} & Stock\ 1 & Stock\ 2 & Stock\ 3 & Stock\ 4 \\ \text{A4} & 0 & 0 & 0 & 1 \\ \text{A3} & 0 & 0 & 1 & 0 \\ \text{B3} & 0 & 0 & 1 & 0 \\ \text{B2} & 0 & 1 & 0 & 0 \\ \text{C2} & 0 & 1 & 0 & 0 \\ \text{C1} & 1 & 0 & 0 & 0 \end{cases} \tag{3.31}$$

Similarly, the matrix $\mathbf{\phi}$ indicates whether fishery f operates in sub-area i (Equation 3.32):

$$\vec{\mathrm{f}} \in \vec{\imath} = \begin{cases} & C1 & C2 & B2 & B3 & A3 & A4 \\ AK\ Fixed\ Gear & 0 & 0 & 0 & 0 & 1 & 1 \\ AK\ Trawl & 0 & 0 & 0 & 0 & 1 & 1 \\ BC\ Longline & 0 & 0 & 1 & 1 & 0 & 0 \\ BC\ Trap & 0 & 0 & 1 & 1 & 0 & 0 \\ BC\ Trawl & 0 & 0 & 1 & 1 & 0 & 0 \\ US\ West\ Coast\ Fixed\ Gear & 1 & 1 & 0 & 0 & 0 & 0 \\ US\ West\ Coast\ Trawl & 1 & 1 & 0 & 0 & 0 & 0 \end{cases} \tag{3.32}$$

3.3.10.1.2 Growth

Growth in the OM follows a latitudinal cline whereby sablefish have a higher asymptotic length at more north-western locales (i.e., sub-areas A3 and A4), corresponding to major oceanographic features (Figures 2.7 and 3.3; Table 3.3). The OM also has a time block in growth for females for all but the most north-westerly sub-areas, following findings in Chapter 2 that the growth parameters for females in these sub-areas differs significantly before and after 2010.

3.3.10.1.3 Weight-at-Length

Weight-at-length is stock-specific, time-invariant, and shared between sexes. The allometric parameters used in Equation 3.6 for stock 1 (US West Coast) are taken from the recent regional stock assessment. For stock 4, two parameters (α and β, Equation 3.6) are estimated outside of the operating model using `optim()` in R (R Core Team, 2019) using weight-length data from the recent regional stock assessment for Alaska. Parameter values for stock 2 are the average of the regional assessment values for the US West Coast and British Columbia; for stock 3, the average of the

parameters from British Columbia and those from the fits to the Alaskan length-weight data are used. The resultant weight-length parameter values are given in Figure 3.4 and Table 3.4.

3.3.10.1.4 Maturity

Stock-specific maturity was estimated externally using both macroscopic and histological data for the four stocks (Williams, In Prep). The estimates are obtained using a generalized additive model that had covariates for depth, age, region, and sampling location, aggregated to the spatial domain of each stock as represented in the OM. This model was identified among other candidate models using AIC. The resultant maturity-at-age curves for each stock used in the OM are shown in Figure 3.5.

3.3.10.1.5 Recruitment

Only two parameters for τ^{ik} are estimated per year because τ_y^{ik} is a proportion, and two stocks (Stock 1 and Stock 4) are associated with only one sub-area. These parameters are the proportion of recruits from Stock 2 that is assigned to sub-area C2, and the proportion of recruits from Stock 3 that is assigned to sub-area B3. Then, $\tau_y^{A3,Stock3} = 1 - \tau_y^{B3,Stock3}$ and $\tau_y^{B2,Stock2} = 1 - \tau_y^{C2,Stock2}$ (Table 3.5).

3.3.10.1.6 Movement

Movement between areas is modeled using a matrix **X**, which represents the transition probabilities for a fish of age a in sub-area i at the time before the catch is removed and surveys are conducted (Figure 3.6). Movement among sub-areas was estimated outside the OM using an analysis of over 30 years of tag-recapture data (Appendix C.1). The analysis provided values for movement parameters among the six modeled sub-areas for sablefish larger than 400 mm (Rogers et al., In Prep), approximating the transition from immature to mature sablefish. Fish residing in sub-area i are subject to the same movement probabilities regardless of their natal sub-area. It was necessary to convert from size-based movement estimates to sex-, stock-, and age-specific movement rates, as described in *Appendix C.1*. Briefly, this involved calculating the dot-product of the distribution of length-at-age (which can vary by sex and sub-area) and the putative movement rates to calculate the expected movement rate by age, sex, and sub-area. Fish of both sexes and in all sub-areas are expected to be larger than 400 mm by age 2, at and after which movement probabilities are

specified for all sub-areas. Movement is assumed to be zero before age 1, and the upper limit of 800mm captures the asymptotic length for all sub-areas. The resultant movement matrices are presented in Figure 3.6. Sensitivity to the movement rates is explored in *Appendices C.1* and *C.3*.

3.3.10.2Fisheries and surveys for transboundary Sablefish

3.3.10.2.1 Fishery catches

There are at least two fishery-dependent sources of catch records for each management region. The hook and line and pot fisheries for the US West Coast and Alaska are combined into a single fixed-gear fishery for each management region; the recent regional assessments for the US West Coast also combined these fisheries (Haltuch et al., 2019; Kapur et al., 2021b). The length-based selectivity function is 1.0 for all fisheries and lengths for fisheries operating in the Alaskan and US West Coast management regions, because these assessments for these regions have historically not modeled length-based selectivity for any fleet. Fishery selectivity is mirrored for both sexes in British Columbia. Because preliminary model runs did not suggest fishery selectivities very different from those used in current regional assessments, all fisheries except for the US West Coast trawl and fixed-gear fisheries had length- or age-based selectivity set to the patterns used from the recent regional assessments.

Time blocks for US West Coast fishery retention and selectivity schedules were based on previous research with respect to influential management ‘milestones’, and the recent introduction of catch shares within the trawl fishery; these are identical to those used in the recent benchmark assessment (Haltuch et al., 2019). They include: (a) the rapid post-war fishery development, (b) the introduction discarding trip limit program implemented in 1982 for the trawl fleet and 1997 for the fixed-gear fishery, (c) a change in selectivity in 2003 resulting from large scale movements of all fleets in response to large spatial closures (Rockfish Conservation Areas), and (d) the implementation of the 2011 catch share program (Table 3.6). Selectivity for the Alaskan fixed-gear fishery is divided into the period before and after the introduction of the IFQ fishery program in 1995. The available fisheries and compositional data corresponding to each region are shown in Figure 3.2 and the catches used in the OM are shown in Figure 3.7.

3.3.10.2.2 Fishery discards

In Alaska, discard estimates are unavailable before 1993 (Goethel et al., 2021) yet always account for < 5% of total catch, so discards in the Alaskan fishery are not explicitly modeled in the OM and all fish are assumed to be retained. For other fisheries, discard mortality is assumed to be 100%, although these are often set to values less than 100% in some regional assessments.

3.3.10.2.3 Surveys

Survey input data were developed using the Vectorized Auto-regressive Spatio-Temporal model (VAST, Thorson, 2019), which enabled the inclusion of one or more regional surveys into sub-area-specific indices of relative abundance. For Alaska and the US West Coast, estimated indices of relative abundance are spatially stratified to align with the sub-areas in the OM (Figure 3.8; Table 3.8). The indices roughly mimic the trend and scale of individual indices used in separate assessments (Figure 3.9). The data inputs and methods used to standardize survey data within VAST are detailed in *Appendix C.2*.

Due to computational constraints, and the desire to use an index that does not vary greatly in trend from those used in the actual assessment each region, the OM used two design-based indices (in lieu of VAST estimates) for British Columbia. The rationale for this is that British Columbia stakeholders felt that a novel standardization using VAST did not capture the steep declines in relative abundance they have observed and preferred that the OM be conditioned using data that they considered reflected the state of knowledge for that region.

Survey selectivity is time-blocked for the US West Coast VAST surveys from 1960-1995 and 1996-2003 (due to a change in survey timing, whereby the survey started earlier during the year after 1995) and from 2003-2019, to account for the fact that the VAST standardization incorporates two distinct surveys (the Triennial and West Coast Bottom Trawl Survey, WCGBTS), which respectively operated before and after 2003. Selectivity for other surveys is not time blocked (Table 3.6). Initial explorations of allowing survey selectivity to be estimated did not result in functional forms very different from those used in the regional assessments, nor through time. Therefore, survey selectivity is set to the relevant values from the regional assessments for all survey fleets. Surveys from British Columbia use length-based selectivity, which are converted to

selectivities-at-age via a lookup table (see Section 3.3.3.2). Table 3.6 indicates the functional form used for each survey.

3.3.10.3Composition data

Age-composition data for surveys and/or commercial fisheries are compiled for each management region, including the Alaska and West Coast fixed-gear fisheries, the West Coast trawl fishery, the British Columbia trap fishery, and the two British Columbia survey fisheries for which abundance observations exist (Figure 3.8). All age-composition data are sex-specific except those for the Alaska fixed-gear fishery. Age frequencies are binned to years and span ages 1-30 for Alaska, 1-34 for British Columbia, and 0-50 for the US West Coast.

Input sample sizes for the age-composition data, by fleet, year, and sex, are set to those used in the regional assessments (Table 3.9). For Alaska, the input sample sizes are the total number of hauls per year for the fixed-gear fishery. For the US West Coast, input sample sizes for the fixed-gear and trawl fisheries are computed using an expansion procedure based on state-level fishery-dependent sampling (see Kapur et al. (2021b), for a brief description). Input sample sizes for British Columbia are the numbers of sampled fish from each survey or fishery fleet each year.

3.3.10.4Ageing error

The OM uses two matrices defining the precision and bias of age reads, which apply to any fishery or survey with age-composition data within each management region (Table 3.10). Ageing error matrices were computed by management region external to this analysis (see Section 3.3.11.1.5). Sablefish ages were derived from visually counting annual otolith rings using the 'break-and burn' method and can be difficult to age due to longevity, resulting in variable estimates of individual age across multiple age reads of an individual sample. The two components of ageing error include bias and precision and can be estimated using a hierarchical model (Punt et al., 2008) and data on multiple age reads across independent laboratories and staff within these laboratories. Bias is the expected difference between true age and average-read age, and precision is the variability around that average read age. The ageing-error method from Punt et al. (2008) estimates: (1) the true proportion-at-age in the sample and (2) the bias and precision for each set of age reads. These age reads could be from individual readers, although it is common to pool samples across age readers to represent laboratories. This method treats the 'true' age for each otolith as a random effect and

estimates the marginal likelihood of all other fixed effects while integrating across these random effects.

3.3.11 Conditioning the transboundary sablefish operating model

Conditioning refers to the estimation and/or specification the parameters of the operating model. As mentioned above, parameter values related to growth, movement and maturity are estimated externally and thus are not estimated by fitting the OM to the monitoring data (Table 3.2). The goal of the conditioning process is to find parameter vectors that lead to model outputs that roughly mimic the data and the derived quantities from the actual regional assessments (i.e., the general trend in spawning stock biomass). The OM includes novel datasets (survey indices, ageing error matrices), accounts for movement, and models key processes (such as survey catchability and density dependence) in a manner that differs from the recent regional assessments, so model results and the fits to available data are not expected be identical to those assessments. The likelihood functions used in the maximum likelihood framework are described below, and are ordered to match the description of the OM.

3.3.11.1Likelihood components

This section provides an overview of the contribution to the objective function by data type.

3.3.11.1.1 Recruitment Deviations

The recruitment deviations are penalized to be normally distributed and to sum to zero:

$$L_{Rec} = \frac{1}{2}\left(\sum_{y} \frac{\widetilde{R_y^2}}{\sigma_R^2} + ln(\sigma_R^2)\right) - \frac{\sum_y \tilde{R}_y^2}{0.01} \qquad (3.33)$$

The penalty on the annual partitioning of recruits within stocks comprised of two spatial areas, τ_y^{ik}, is based on a beta-distribution parameterized using survey observations:

$$L_\tau = \sum\nolimits_{yik} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \tau_y^{ik} \left(1-\tau_y^{ik}\right)^{\beta-1} \qquad (3.34)$$

where $\alpha = \left(\frac{(1-\mu)}{0.01} - \frac{1}{\mu}\right)\mu^2$, $\beta = \alpha\left(\frac{1}{\mu} - 1\right)$ and μ is the mean ratio of observed survey biomass in constituent areas i for the final 10 years in the historical period: $\left(\sum_{2010}^{2019}\sum_{i\in k}\frac{B^i_{Obs,y}}{B^k_{Obs,y}}\right)/10$. Ratios from the survey data suggest that most recruits move out of British Columbia, which is consistent with the movement estimates from the analysis of tag-recapture data. Specifically, for the six-area, four-stock EMs, the prior mean for recruitment apportioned northward from the North BC-Gulf of Alaska stock into the Gulf of Alaska was typically 80%; the prior mean for recruits apportioned southward from the South BC-North US West Coast stock into the Northern US West Coast was typically 90% (see *Appendix C* for figures and more details).

3.3.11.1.2 Catch and Discards

The discard ratios are assumed to be lognormally distributed, with a very small standard deviation to reflect high degrees of observer coverage across fleets (0.2), i.e.:

$$L_{Disc} = \frac{1}{2}\left(\sum_y\sum_f \frac{\left[ln(D^f_{obs,y}) - ln(D^f_{pred,y})\right]^2}{0.2^2}\right) \qquad (3.35)$$

where $D^f_{obs,y}$ is the observed discard ratio by year and fleet; and $D^f_{pred,y}$ is the OM-predicted discard ratio for year y and fleet f (see Equation 3.16).

3.3.11.1.3 Survey biomass indices

Estimates of relative abundance (biomass) from each survey are fit under the assumption that the indices for each year are lognormally distributed. The value for fleet-specific q is calculated analytically. Catchability is time-blocked for the West Coast VAST surveys from 1960-1995 (due to a change in survey timing because the survey started earlier in the year after 1995) and from 2003 onwards (due to a switch from the Triennial to the Bottom Trawl surveys, Table 3.6). Selectivity is estimated using the settings presented in Table 3.7.

$$\left(q^f_{y*}\right) = \frac{\sum_{y\in y*}\ell n(\frac{B^f_{Obs,y}}{B^f_{Pred,y}})/\left[\left(\sigma^f_y\right)^2 + 0.2^2\right]}{\sum_{y\in y*}1/\left[\left(\sigma^f_y\right)^2 + 0.2^2\right]} \qquad (3.36)$$

$$L_{surv} = \sum_{y \in y*} \sum_{f} \left(0.5 \ell n \left[\left(\sigma_y^f \right)^2 + 0.2^2 \right] + \frac{1}{2 \left[\left(\sigma_y^f \right)^2 + 0.2^2 \right]} \left(\ell n \left(B_{Obs,y}^f / B_{Pred,y}^f \right) - \ell n \left(q_{y*}^f \right) \right)^2 \right) \quad (3.37)$$

where $B_{Obs,y}^f$ is the observed index of relative abundance for survey f and year y; $B_{Pred,y}^f$ is the predicted biomass corresponding to the index for survey f during year y (see Equation 3.27); q_{y*}^f is the catchability coefficient for fleet f during the block of years defined by the set $y*$; and σ_y^f is the observation error for survey f and year y. An additional standard deviation in log space of 0.2 is added to provide additional flexibility in the survey-fitting routine (Francis, 2011).

Finally, a penalty is added to the likelihood for values of q_{y*}^f greater than 5, to prevent unreasonably small exploitable biomass in spatial areas, particularly those that are isolated by movement and demography:

$$Penalty_q = \begin{cases} 0 & if\ q_{y*}^f <= 5 \\ \sum_{y \in y*, f} 0.01 \left(q_{y*}^f - 5 \right)^2 & if\ q_{y*}^f > 5 \end{cases} \quad (3.38)$$

3.3.11.1.4 Compositional data

Age-composition data from selected surveys and fishing fleets are fit assuming the data are multinomially distributed. The contribution of the age-composition data for fleet f to the objective function is as follows:

$$L_\rho^f = -\sum_y \sum_a \sum_\gamma \log(\rho_{y,a,\gamma}^f) n_{y,\gamma}^f \widetilde{\rho_{y,a,\gamma}^f} \quad (3.39)$$

where $n_{y,\gamma}^f$ is the total number of samples for sex γ and year y in the data for fleet f; for all regions, these are set to the values from the regional assessments. No additional tuning of input sample sizes is conducted for the age-composition data. $\widetilde{\rho_{y,a,\gamma}^f}$ is the observed proportions-at-age for fleet f, year y, and sex γ; and $\rho_{y,a,\gamma}^f$ is the OM-predicted proportions-at-age for fleet f, year y, and sex γ.

Due to computational constraints in the conditioning of the operating model, as well as data conflicts identified during the regional assessment process for the US West Coast, neither length-composition nor conditional age-at-length data are included when conditioning the operating

model. This choice is appropriate given the values of the growth parameters are not estimated during the conditioning process, and the management strategy evaluation for which the OM is developed does not consider estimation methods that require length data.

3.3.11.1.5 Ageing error

Although sablefish lack a true age validation study, thousands of otoliths have been read within and between laboratories, including 15 individuals from tag-recapture studies in Alaska with known ages (i.e., no bias and perfect precision) and 24 thin sectioned samples also assumed to be aged without bias (but with imprecision). In addition to within-laboratory double reads, this analysis includes between-laboratory reads from the Northwest Fisheries Science Center (NWFSC), the Alaska Fisheries Science Center (AFSC), the Alaska Department of Fish and Game (ADFG), and the Department of Fisheries and Oceans Canada (DFO). Samples from individual readers within each laboratory are combined into two reads for each lab and are used to estimate a single vector of precision and bias across the age bins.

This analysis includes 20,717 individual otolith samples with 17,777, 16,569, and 6,871 age reads from the AFSC/ADFG combined, the NWFSC, and the DFO, respectively (Figure 3.10). Some samples had ages from all agencies, while others had only double reads from a single agency. The 24 thin sectioned samples from the ADFG (and read by all labs) are combined with the 15 known age fish from the AFSC tagging study and assumed to have known ages for this analysis as these are the best possible ages available for sablefish. In total, 39 fish are assumed known age fish, with four samples aged greater than 15 and a maximum age of 22. There is a lack of older fish in the known age samples. The minimum age from the known age fish is 3. All other data sets had a minimum age of 1. Maximum ages from the non-known age samples are 83, 91, and 94 from the AFSC, NWFSC, and DFO, respectively. Mean ages from the non-known age samples are 10, 10, and 14 from the AFSC, NWFSC, and DFO, respectively. Median ages from the non-known age samples are 7, 5, and 10 from the AFSC, NWFSC, and DFO, respectively.

Previous age error analyses that used only the AFSC “known-age” tagging study fish and treated them as unbiased led to unrealistic, non-linear estimates of ageing error for the NWFSC and the DFO data sets at older ages, due to the small known age data set. Given that the AFSC tagging study samples are insufficient to anchor the age-error analyses the thin sections samples are

combined with the tagging study samples and assumed to be known age samples for this analysis. The age estimation method is applied with a reference age of 4, maximum age of 100, and with a plus group of 50. The known age samples are assumed unbiased, and curvilinear bias using a 2-parameter Hollings-form relationship of bias with true age is estimated for the AFSC, the NWFSC, and the DFO samples. Curvilinear standard deviations based on a 3-parameter Hollings-form relationship of standard deviation with true age is estimated for the known age, AFSC, NWFSC, and DFO samples. Figure 3.11 shows estimates of ageing bias and precision for the known age samples and each agency.

3.3.12 Model projections

The conditioned OM can be used to project the populations forward in time for use in the Management Strategy Evaluation presented in *Chapter 4*. These simulated datasets should represent process and observation error relevant to northeast Pacific sablefish. Process error is represented via an annual recruitment deviate ($\tilde{R}^k$) for each stock, and a vector of future catches by fleet (Table 3.11).[3]

Stochastic datasets of survey abundance are generated, subject to random observation error simulated from the terminal year of the OM (age-composition data are not simulated for the projection period). The future catches are assumed to be known exactly; fishing mortality *F* is tuned within the projections to realize these catches from the expected vulnerable biomass.

3.3.13 Future recruitment deviations

The matrix of year- and stock- specific recruitment deviations $\tilde{\mathbf{R}}$ for the projection period is simulated externally to retain the observed spatial correlation from the historical period ($\tilde{R}^k_{1960-2019}$).

These correlations are highest, in an absolute sense, between adjacent stocks in the US and Canada, with a 57% correlation between stock 2 (the US West Coast and southern British Columbia), and stock 3 (northern British Columbia/the Gulf of Alaska) and a correlation of -57% between northern British Columbia/the eastern Gulf of Alaska and the western Gulf of Alaska/Bering Sea/Aleutian

[3] In the MSE analysis, the input vector of fleet-specific catches comes from a harvest control rule and might require the calculation of catch from fleet-specific *F*. Additionally, actual catches from 2020 are used in the MSE, although the OM terminates in 2019.

Islands (stocks 3 and 4, Figure 3.12). The positive correlation between stocks 2 and 3 is likely caused by the high level of mixing between the four areas composing these two stocks, and the fact that there is a single survey for British Columbia, which straddles both stocks. The negative correlation across the 145*W line (splitting the eastern Gulf of Alaska from the western Gulf of Alaska, Bering Sea and Aleutian Islands) is consistent with the ecological perception of those areas as separate, large marine ecosystems (Cleaver and Evans, 2019). There is a near-zero correlation (<1%) between stocks 1 and 2. This is consistent with findings by Tolimieri and Haltuch (2018), that suggest that recruitment dynamics in the southern part of the US West Coast (south of 36^0N) are distinct from the rest of the coast, and not as strongly linked to stockwide dynamics as recruitment in the northern regions; the recruitment paradigm in the northern region appears to be strongly linked to sea-level height (Tolimieri and Haltuch, 2023). The last synthesis of recruitment trends throughout the northeast Pacific, based upon recent assessments, also indicated US West Coast and British Columbia recruitment indices are more similar to one another than to Alaskan recruitments (Fenske et al., 2019). Recruitment indices for Alaska appear to lag those in the more southern regions by one to two years.

The simulated deviations are based on the Cholesky decomposition of the upper triangle of ($\sum$), where ($\sum$) is the matrix of Pearson correlation coefficients among k simulated vectors of uncorrelated, zero-centered gaussian random variables for each stock $N(0, \sigma_R)$ (upper triangle; Figure 3.12):

$$\tilde{R} = N(0, \sigma_R) chol(\Sigma) \qquad (3.40)$$

The resulting matrix of future recruitment deviations therefore accounts for both the overall variation in recruitment and the spatial autocorrelation among stocks during the historical model period (Figure 3.12).

3.3.14 Demography in projection years

Surveys in the projections occur at the same frequency as the historical surveys, such that the two US West Coast surveys, the two Alaskan surveys, and the British Columbia Stratified Random survey are each conducted annually. Of the two historical British Columbia surveys only the Stratified Random survey is simulated into the future, because the British Columbia Offshore Standardized survey ended in 2010. Survey abundances are assumed to be log-normally distributed

about their expected values (proportional to the biomass vulnerable to each fleet in the middle of the year) with survey observation error variance given by the fleet-specific observation error variance (σ^f, from the terminal year [2019] of the VAST index for survey f, Table 3.8):

$$B^f_{forecast} = q^f \sum_a \sum_i \sum_\gamma \phi^{if} w^f_{M_{y,\gamma,a}} N^i_{M_{y,\gamma,a}} s^f_{y,\gamma,a} \left[exp(\varepsilon^f_y - \frac{{\sigma^f}^2}{2}) \right]; \varepsilon^f_y \sim N(0, {\sigma^f}^2) \quad (3.41)$$

The simulation process is repeated 100 times to generate many datasets that account for both process and observational uncertainties.

3.4 Results

3.4.1 Fits to available data

The conditioned operating model fit the fleet-specific catches perfectly (Figure 3.13), and approximates the discarding ratios for the British Columbia longline and trawl fisheries and the US West Coast fixed-gear fisheries adequately (Figure 3.14). The OM did not fit the British Columbia trap discard fractions as well as the regional model, and overestimated the last 10 years of the US West Coast trawl discards (Figure 3.14).

Fits to the survey data were comparable to those from the regional assessments for all four VAST surveys (except for the terminal year in sub-area A4, Figure 3.15), and slightly better than the regional assessment for British Columbia surveys. Analytical q values were highest for all time blocks for the smallest stock/sub-area, C1 (between 2.95 and 6.28); q for other surveys ranges from 0.16-2.56 (Table 3.12). The input indices for British Columbia were identical to those used in the regional assessment, which also did not fit well to the steep decline in the offshore standardized survey from 2000-2010, although the OM fit the sharp increase biomass during the latter years of the stratified random survey slightly better than the regional assessment (Figure 3.15). OM predictions of survey biomass for sub-area A4 (Alaska) did not capture the steep increase in survey biomass observed between 2018-2019, and more closely followed the lower-variability trend observed in sub-area A3 (Figure 3.15). Overall, the OM fits to survey data were qualitatively similar to those achieved by the regional assessments.

Aggregated fits to age-composition data for the Alaskan fixed-gear fishery were roughly comparable to those from the regional assessment; males were slightly over- and under-fit for the US West Coast fixed- and trawl-gear fisheries, respectively (Figures 3.16 and 3.17). The age-

composition data from the British Columbia surveys were fit similarly to the regional assessment, except for the male plus-group for the offshore standardized survey, which was somewhat underfit (Figure 3.16, center-left panel). The fits to the age-composition data for the British Columbia trap fishery were adequate, except for an over-estimation of young (<5 year) individuals of both sexes (Figure 3.16, top-right panel). This over-estimation was persistent throughout the time series for British Columbia (Figure 3.16). The plus group for the age-composition data in Alaska and British Columbia was considerably smaller than that in the OM (at 30 and 35 years, respectively); this was predicted adequately for the Alaskan fixed-gear and both British Columbia surveys, but somewhat overfit for females in the British Columbia Trap Fishery and males in both British Columbia surveys, for which the observed proportion of individuals in the plus group was larger than any younger age. Similarly, the OM estimated a greater number of young (<5 year) individuals of both sexes for both US fisheries, although the over-estimation was much more pronounced for females in the trawl fleet since 2010, when the catch shares program started (Figure 3.16).

3.4.2 Operating model state variables

Fleet-specific values for F typically reached their maximum during the 1970s and 1990s for all fleets, except for the British Columbia longline and US West Coast fixed-gear fisheries, which increased from ~2005 onwards. F was never greater than $0.15 yr^{-1}$ for any fishery (Figure 3.18).

Fishery selectivity curves for the Alaskan or British Columbia fisheries were identical to those used in the regional assessments, while the selectivity values from the conditioned OM for the US West Coast fixed-gear fishery were less variable through time than in the regional assessment (Figure 3.19; Table 3.13). The US West Coast trawl fishery selectivity varied through time in both the regional assessment and the OM (Figure 3.19). Survey selectivity and retention curves were identical to those in the regional assessments for all surveys (Figure 3.20; Table 3.14) and fisheries (Figure 3.21; Table 3.15), respectively.

The recruitment deviations were at or above zero in the first model year for all stocks; resulting recruitment time series were within similar range as those estimated in the regional assessments (Figures 3.22 and 3.23). The log standard deviation in recruitment was estimated to be 0.71 (Table 3.16). Estimated recruitment in the conditioned OM was most similar to the recent regional

assessments for Alaska and the US West Coast, while the scale of British Columbia recruitment was estimated somewhat above that regional assessment for the data rich period (~2005-2019, Figure 3.2). Generally, the OM captured a similar trend in estimated recruitment for each management region, with a high period in Alaska during the 1980s and late 2010s, and high periods for the US West Coast during the 1960s, late 1990s and late 2010s, and a recent maximum between 2016-2018 in all regions.

Estimates of recruitment partitioning within stocks 2 and 3 indicated a preference for recruits to settle in sub-area C2 rather than B2, and in sub-area A3 rather than B3 (Table 3.17). From 2000 onwards, the OM values for τ suggested that roughly 90%-95% of recruits spawned in stock 3 (the US West Coast north of 36 degrees N and southern British Columbia) settle on the US West Coast (sub-area C2), whereas 75%-85%% of recruits spawned in stock 3 (northern British Columbia and the eastern Gulf of Alaska) settle in Alaska (sub-area A3) (Table 3.17).

The OM model scale in terms of unfished spawning stock biomass S_0 was similar to values estimated in recent regional assessments for the US West Coast and Alaska (Figure 3.24). S_0 values in the conditioned OM were 253,237 t, 114,468 t, and 168,072 t for Alaska, British Columbia and the US West Coast, respectively. The recent regional assessment estimates of S_0 were 295,351, 56,560 mt, and 168,875 mt, respectively. The SSB trend fell within the confidence interval for most of the period after the onset of survey and compositional data for Alaska and the US West Coast, whereas the trend for British Columbia matched that of the regional assessment though slightly higher in magnitude (Figure 3.24).

3.5 Discussion

This chapter presents a novel, transboundary age-structured model conditioned to historical observations of sablefish catch, discards, survey biomass and age compositions from 1960-2019. The model, and sensitivity analyses thereof, illustrate the plausibility of an interconnected transboundary sablefish stock that would result in the regional biomass trends estimated in the individual management models.

The US West Coast regional assessment has struggled to capture the discardratios for the trawl fishery without implementing short (2-year) time blocks in retention for recent years. Recent length-composition data (not included in the OM) suggest that that fishery has encountered a large

influx of small fish in recent years (Kapur et al., 2021b). The base OM used the same length-based retention curves as the regional assessments (Table 3.11), so discrepancies in fits to the discard data between the OM and regional assessments were likely a result of the different growth curves used here and the movement of individuals among regions.

Analytical q in the isolated sub-area south of 36°N as the highest of all fleets (between 2.95 and 6.28), suggesting that the mid-year biomass in that sub-area was too low to realize the survey observations estimated using VAST (wherein ~22% of survey biomass in the US West Coast is in sub-area C1, similar to the 26% assumed in the regional assessment, Kapur et al. (2021b)). This could be a result of the way movement and recruitment were specified in the OM; if movement is correctly specified, it's possible that the ~8% of individuals that migrate north from sub-area C1 to C2 (Figure 3.1) do so after survey sampling. It is conceivable that there is more diffusion of recruits and/or adults from north to south than the movement and spatial structure in the OM allows for. The number of returned tags from the US West Coast is smaller than the number of returns for British Columbia or Alaska, so it is possible that movement estimates are more accurate within the northerly regions than for the US West Coast. Finally, sub-area C2 might not operate as a standalone stock in terms of density dependence, although this would be inconsistent with findings by Tolimieri and Haltuch (2018) that suggest that recruitment dynamics in the southern part of the US West Coast (south of 36^0N) are distinct from the rest of the coast, and not as strongly linked to stockwide dynamics as recruitment in the northern regions; the recruitment paradigm in the northern region appears to be strongly linked to sea-level height (Tolimieri and Haltuch, 2023). Restructuring the OM was out of the scope of this study, but sensitivity analyses examined movement scenarios and the sharing of recruitment deviations among stocks (see *Appendix C.1*). Briefly, it does not appear that the range of uncertainty in movement rates developed for this analysis result in large differences in spawning stock biomass trajectories, but assuming no mixture among regions results in distinct (higher) perception of biomass for British Columbia (Figure C.12). The range-wide dynamics appear to be more heavily influenced by the recruitment trajectories for stocks 2 and 3, which is consistent with the fact that there is a large amount of data for these stocks, and fish from these stocks move to all sub-areas. Understanding the dynamics of recruitment, particularly for those stocks, would be helpful to confirm or update the density-dependence assumptions applied here.

Regional dynamics were generally reproduced given the same selectivity curves as used in the regional assessments. The US West Coast trawl survey is an exception to this, suggesting that the spatial dynamics within the OM require lower terminal selectivity values to reconcile the transfer of larger fish into this region. The estimated selectivity curves are not necessarily accurate reflections of fleet dynamics because the OM integrates multiple processes (namely movement and growth) across spatial areas, which span multiple fishing fleets. Waterhouse et. al. (2014) provide further discussion of how spatial processes can lead to misspecifications in selectivity curves.

The fits to the age-composition data were not expected to be as good those in as the regional assessments because of the large amount and variety of data integrated into the OM, and the fact that age-at-length in the OM differs from those in the regional models, is not estimated, and is subject to variation due to movement. It is particularly challenging to fit the data for young ages (<10 years), when growth and selectivity are changing most rapidly for most fisheries and surveys. As mentioned above, the recent regional assessment of the US West Coast (conducted in 2021, Kapur et al. (2021b)) detected a large influx of young fish in the Trawl fishery, which resulted in a near-tripling of the discarding rates between 2019 and 2021. The recent regional assessment in Alaska (conducted in 2021, Goethel et al. (2021)) also noted recent increases in the abundance of younger fish in deepwater strata in the longline survey, where they have not historically been caught, and were also able to reconcile these changes in availability via a selectivity time block. It is possible that the spatial dynamics in the OM (that terminates in 2019) are anticipating this influx, by allowing the large positive recruitment deviations from 2015/2016 in stocks 2 and 3 to enter the Alaskan and US West Coast fisheries. This phenomenon might also contribute to the slightly inflated estimates of discard fractions for the US West Coast fisheries in the recent time period. This suggests that spatially-structured estimation frameworks might advance the perception of changes in recruitment for sablefish, potentially avoiding foregone catches and/or allowing for management adaptation in advance of stock declines.

It is challenging to determine whether the estimates of recruitment partitioning among stocks represent actual ecological phenomena or are simply reflections of the fact that fishery yields and observed survey biomass (upon which the prior for τ_y^{ik} is based) are greater in sub-areas C2 and A3 than in their British Columbia counterparts; these parameters are inherently confounded with

movement. The directionality of τ_y^{ik} is consistent with the movement dynamics, in that roughly half of the adult population in British Columbia is exported to adjacent areas each year, Thus, these results suggest that recruits might follow a source-sink regime out of British Columbia, like the situation for older sablefish. Finally, in terms of model scale, it is difficult for spatial models to precisely capture the scale of the smallest component within the domain (D. Goethel, NOAA, pers. comm). It is likely that the degree of movement out of British Columbia specified in this model (Figure 3.5) requires a higher unfished biomass, as suggested by the OM estimates provided above (which is roughly double the estimate from the regional assessment model for British Columbia).

The stock structure defined in the OM follows from the analysis in Chapter 2, and the concept of the Bering Sea/Aleutian Islands (area west of 145°W) as a separate marine ecosystem from that of the Gulf of Alaska (Cleaver and Evans, 2019). The US West Coast is modeled consistent with recent empirical analyses indicating that the area south of Monterey Bay (36.6°N) exhibits distinct sablefish recruitment dynamics from the northern area (Tolimieri and Haltuch, 2023). Finally, the population off British Columbia is considered a highly connected component of the overall population, generally dependent on subsidies from Alaska to meet catch quotas and fit survey indices. It is clear the movement rates and the location and structure of density-dependence (recruitment) remain crucial uncertainties to evaluating the population status of long-lived, highly mobile species such as sablefish, echoing suggestions from simulation work Goethel et al. (2011) and applications to other groundfish species (Jacobsen et al., 2022; Mazur et al., 2023).

An important next step for this work would be to synthesize the drivers of recruitment dynamics for the sablefish population coastwide, to 1) confirm or modify the structural assumptions present in the operating model, and 2) allow for exploration of how oceanic conditions might lead to changes in the recruitment paradigm for sablefish, and effects on management performance. This analysis could inform the development of additional operating models and allow for investigation of the interaction of climate change and spatial uncertainty for high-value species such as sablefish.

3.6 Tables

Table 3.1: Input data included in operating model; italicized datasets are not used in this OM. *Original treatment of the data (i.e., in recent stock assessments used for management). **New index of relative abundance standardized using VAST, which combines survey(s) from this management region across space and time.

Region	Data Type	Description & handling notes	Reference
Alaska	Landings	2 fleets (early 1900s-present; typically truncated to 1970 onward): ●Fixed-gear (longline & pot) and trawl	Goethel et al., 2021; Hanselman et al., 2017
	Compositions	Fishery-dependent: ●Sex-aggregated ages (1999-present) from the fixed-gear fishery	
	Indices of abundance	Fishery-independent: ‡Alaska index of relative abundance for sub-areas A3 and A4 (1990-2019)	
British Columbia	Landings	3 fleets (1965-present): ●Commercial longline trap, longline hook, and trawl	
	Compositions	Fishery-dependent: ●Sex-specific ages from the trap fishery Fishery-independent: ●Sex-specific ages from the trap-based standardized survey (1991-2009) and trap-based stratified random survey (2003-present)	Fenske et al., 2019; DFO, 2020
	Indices of abundance	Fishery-independent: ●Standardized trap-based standardized survey CPUE (1991-2009); trap-based stratified random survey CPUE (2003-present)	Fenske et. al., 2019
US West Coast	Landings	2 Fleets (1990-present): ●Fixed gear (Hook & Line, and Pot), and trawl	Haltuch et. al., 2019; Kapur et. al., 2021
	Compositions	Fishery-dependent: ●Sex-specific ages from the fixed gear and trawl fleets (via commercial port sampling)	Haltuch et. al., 2019; Kapur et. al., 2021
	Indices of abundance	Fishery-independent: ‡California Current index of relative abundance for sub-areas C1 and C2 (1980-2019)	Haltuch et. al., 2019; Kapur et. al., 2021

Table 3.2: Overview of the parameters of the operating model. Acronyms referring to OM-specific regions are explained in-text and depicted in Figure 3.1. *In text, growth equations use sub-area indexing (i) to clarify the expected length-at-age as fish move among sub-areas from different stocks. In practice, sub-areas belonging to the same stock *k* share growth parameters patterns.

Symbol	Description	Operating model treatment	Estimated?
Model Structure & Notation			
i	Sub-areas		
n	Number of sub-areas	6	
k	Stocks (shared demography)	4	
m	Management regions m	3	
f	Fishing fleets	7	
f	Surveys (indices of abundance)	6	
f	Fishing fleets with composition data	4	
f	Surveys with composition data	3	
	Total number of parameters (excluding recruitment deviations and downscaling parameters)	710	403 estimated during OM conditioning
	Total number of recruitment deviations $\tilde{R}_y^k$	244	yes (and see sensitivities)
	Total number of recruitment downscaling parameters τ_y^{ik}	120	yes
Growth*			
M^k	Stock-specific natural mortality at age.	Fixed at values from assessments or means of constituent management areas: $M^{\text{Stock 1}} = 0.07\text{y}^{-1}$; $M^{\text{Stock 2}} =$	no

		0.08y^{-1}; $M^{\text{Stock 3}}$= 0.09 y^{-1}; $M^{\text{Stock 4}}$= 0.10y^{-1};	
$L^{i}_{\infty_{y,\gamma}}$	Asymptotic length (cm)	Sex-, stock- and year specific (Table 3.3)	no
$\kappa^{i}_{y,\gamma}$	Growth rate (cm yr^{-1})	Sex-, stock- and year specific (Table 3.3)	no
α^{k}	Coefficient of length-weight relationship (kg/cm)	Stock-specific (**Error! Reference source not found.**)	no
β^{k}	Allometric exponent of length-weight relationship	Stock-specific (**Error! Reference source not found.**)	no
Reproduction & Recruitment			
$\boldsymbol{\Lambda}$	Matrix indicating whether sub-area *i* is nested within stock *k*	Equation 3.31	
h^{k}	Steepness of the stock recruitment curve (expected proportion of R_0 at $0.2S_0$) for stock *k*	Stock-specific; fixed at 0.7 for all regions	no
E^{k}_{a}	Proportion of females in stock *k* that have reached maturity at age *a*	Figure 3.5	no
$\tilde{R}^{k}_{y}$	Annual recruitment deviations specific to stock *k*	Normally distributed with mean zero and standard deviation σ_R	yes
σ_R	Standard deviation of recruitment deviations	Estimated	yes [0.2, 2]
R^{k}_{0}	Unfished recruitment by stock	Starting values based on assessments	Yes: $R_0^{Stock\ 1}$[7,8]; $R_0^{Stock\ 2}$[9,10]; $R_0^{Stock\ 3}$[9,10]; $R_0^{Stock\ 4}$[6,9]
τ^{ik}_{y}	Proportion of recruits during year *y* spawned in	Table 3.17	Logit transformed [0,1]; constrained to sum to 1 within stock

	stock *k* allocated to sub-area *i; i ε k*		
Catches			
ϕ	Matrix indicating whether fleet *f* occurs in sub-area *i*		
$\beta_{1,2,3,4}^{y,f,\gamma,a/l}$	Age @ inflection point, slope @ inflection point, asymptotic selection, and male offset for logistic retention curve	From regional assessments; time blocks may apply (Tables 3.6 and 3.15)	no; retention Ω set to 1 for all ages for Alaska fishery fleets
$w_a^{k,f}$	Stock- and fleet-specific weight-at-age of captured fish	Table 3.4	no
Surveys			
σ_y^f	Survey observation error	Survey-specific, Figure 3.8 and Table 3.8	no (estimated externally)
q_y^f	Survey catchability coefficient	Estimated using an analytical formula for each survey with biomass; time blocks may apply (Table 3.6)	yes
Selectivity			
$a_{50,\gamma}^f$	Fleet- and sex-specific age-at-50%-selectivity, for logistic selectivity		yes
δ_γ^f	Fleet- and sex-specific slope parameter, for logistic selectivity		yes
ρ_γ^f	Fleet- and sex-specific exponent of the inverse power function		yes
μ_γ^f	Fleet- and sex-specific mean age or length, for dome-shaped normal selectivity		yes
σ_γ^f	Fleet- and sex-specific standard deviation, for		yes

	dome-shaped normal selectivity		
α_γ^f	Fleet- and sex-specific shape parameter, for dome-shaped gamma selectivity		yes
β_γ^f	Fleet- and sex-specific rate parameter, for dome-shaped gamma selectivity		yes
Double-normal selectivity curve parameters	Fleet- and sex-specific ascending limb, descending limb, and/or terminal parameter for double-normal selectivity	See the supplement to Methot and Wetzel (2013) for details.	Peak parameter estimated for both US West Coast fishery fleets; ascending limb estimated for US West Coast Fixed Gear
Age Compositions			
σ_A^f	Standard deviation at age in ageing error matrix	Management region, age, and sex-specific, see Table 3.10	no (estimated externally)
$P_{\gamma,a,\tilde{a}}^f$	Ageing error matrix	Management region, age, and sex-specific, see Table 3.10	no (estimated externally)
Movement			
$\boldsymbol{X}$	Sub-area age-based movement matrix	Figure 3.6	no (estimated externally)

Table 3.3: Growth parameters used in the operating model, updated following methods from Chapter 2. The standard deviation in length-at-age for each stock is not required for the OM as growth is pre-specified, but is implemented to calculate expected length-at-age distributions for the conversion of movement rates (see Appendix C.1).

Stock (subareas)	**Sex**	**Period**	L_∞^k **(Cm)**	κ_γ^k **(yr^{-1})**	σ_G^k
Stock 1 (C1)	Fem	<2010	60.30	0.29	4.98
Stock 1 (C1)	Fem	2010+	62.75	0.16	4.80
Stock 1 (C1)	Male	All years	55.07	0.28	3.89
Stock 2 (C2 & B2)	Fem	<2010	69.36	0.22	11.80
Stock 2 (C2 & B2)	Fem	2010+	68.08	0.18	9.68
Stock 2 (C2 & B2)	Male	All years	59.02	0.21	7.16
Stock 3 (B3 & A3)	Fem	<2010	73.51	0.76	8.43
Stock 3 (B3 & A3)	Fem	2010+	75.04	0.35	10.80
Stock 3 (B3 & A3)	Male	All years	65.66	0.34	9.90
Stock 4 (A4)	Fem	All years	81.43	0.14	6.57
Stock 4 (A4)	Male	All years	68.34	0.20	4.75

Table 3.4: Weight-at-length parameters for each stock.

Stock (sub-areas)	**α^k**	**β^k**
Stock 1 (C1)	3.34 x 10-6	3.27
Stock 2 (C2 & B2)	6.77 x 10-6	3.19
Stock 3 (B3 & A3)	5.90 x 10-6	3.10
Stock 4 (A4)	5.73 x 10-6	3.14

Table 3.5: Structure of matrix for τ^{ik}, the distribution of recruits during year *y* spawned in stock *k* to sub-area *i*; $i \in k$. These values are represented in equations with the symbol τ_y^{ik}.

	Destination sub-area					
Stock	**C1**	**C2**	**B2**	**B3**	**A3**	**A4**
1	1	0	0	0	0	0
2	0	$\boldsymbol{\tau}_y^{C2,2}$	1-$\boldsymbol{\tau}_y^{C2,2}$	0	0	0
3	0	0	0	$\boldsymbol{\tau}_y^{B3,3}$	1-$\boldsymbol{\tau}_y^{B3,3}$	0
4	0	0	0	0	0	1

Table 3.6: Start and end years and rationale for time-blocked quantities.

Fleet	Quantity	Time block(s)	Rationale
US West Coast fixed gear fishery	Fishery selectivity	1997-2002	Management trip limits
		2003-2010	Rockfish conservation area
		2011-2019	Catch shares
US West Coast fixed gear fishery	Retention	1960-1996	Post-war fishery development
		1997-2010	Management trip limits
		2011-2019	Catch shares
US West Coast trawl fishery	Fishery selectivity	1982-2002	Management trip limits
		2003-2010	Rockfish conservation area
		2011-2019	Catch shares
US West Coast trawl fishery	Retention	1982-2010	Management trip limits
		2011-2019	Catch shares
US West Coast Trawl survey sub-area C2	Survey selectivity and Catchability (q)	1996-2003	Change in survey timing
		2004-2019	Switch between surveys
US West Coast Trawl survey sub-area C1	Survey selectivity and Catchability (q)	1996-2003	Change in survey timing
		2004-2019	Switch between surveys
Alaska fixed gear fishery	Fishery selectivity	1995-2019	Introduction of IFQ fishery

Table 3.7: Surveys and fisheries, year range of available data, and form of selectivity curve used in the operating model.

Management Region	Fleet Name	Fleet Type(s)	Years	Selectivity form, if applicable	Note
Alaska	Fixed gear fishery	Comm. catches, age comps	1977-2019	Age-based logistic	
Alaska	Trawl fishery	Comm. catches, age comps	1970-2019	Age-based dome (normal)	
Alaska	Domestic Longline survey - Sub-area A4 and Domestic Longline survey - Sub-area A3	Standardized indices of relative abundance corresponding to sub-areas A4 and A3	1990-2019	Age-based logistic	Model-based standardization, stratified to match OM spatial structure
British Columbia	Longline fishery	Comm. catches, discards	1965-2019	Length-based dome (normal), sexes mirrored	
British Columbia	Trap fishery	Comm. catches, age comps	1973-2019, some age comps from 1965	Length-based dome (normal), sexes mirrored	
British Columbia	Trawl fishery	Comm. catches, discards	1965-2019	Length-based dome (gamma), sexes mirrored	
British Columbia	Stratified random survey	Standardized index of relative abundance, Age Comps	2003-2019 (intermittent)	Length-based logistic, sexes mirrored	Design-based standardization
British Columbia	Offshore standardized survey	Standardized index of relative abundance, Age Comps	1991-2009	Length-based logistic, sexes mirrored	Design-based standardization
US West Coast	Fixed gear fishery	Comm. catches, age comps, discards	1960-2019	Age-based double normal	
US West Coast	Trawl fishery	Comm. catches, age comps, discards	1960-2019	Age-based double normal	
US West Coast	Trawl survey – sub-area C1 and Trawl survey – sub-area C2	Standardized indices of relative abundance corresponding to sub-areas C1, C2	1980-2019	Age-based double normal	Model-based standardization, stratified to match OM spatial structure

Table 3.8: Estimated survey biomass (t) and standard error (in parentheses) used as input for the operating model. The values for Alaska and the US West Coast are are obtained via standardization using VAST as dobtained via standardization using VAST as described in Appendix C.2. The indices of relative abundance for British Columbia are calculated externally.

Year	VAST Survey sub-area A4	VAST Survey sub-area A3	BC Offshore Standardized Survey	BC Stratified Random Survevy	VAST Survey sub-area C2	VAST Survey sub-area C1
1980					104,814 (0.16)	38,834 (0.31)
1981						
1982						
1983					67,924 (0.12)	20,489 (0.29)
1984						
1985						
1986					81,802 (0.14)	33,207 (0.28)
1987						
1988						
1989					62,963 (0.13)	18,861 (0.25)
1990	320,761 (0.15)	91,599 (0.13)	20,018 (0.29)			
1991	294,572 (0.15)	104,761 (0.13)	19,336 (0.29)			
1992	284,817 (0.15)	106,660 (0.13)	25,574 (0.29)		81,414 (0.11)	14,880 (0.28)
1993	297,654 (0.14)	102,033 (0.13)	36,508 (0.29)			
1994	238,700 (0.14)	80,443 (0.13)	14,834 (0.29)			
1995	272,760 (0.14)	80,700 (0.12)	13,563 (0.29)		51,698 (0.1)	14,634 (0.22)
1996	307,753 (0.15)	68,713 (0.12)	11,257 (0.29)			
1997	298,304 (0.15)	77,302 (0.13)	7,722 (0.29)			
1998	233,180 (0.14)	65,903 (0.13)	12,039 (0.29)		69,996 (0.1)	15,815 (0.26)
1999	303,086 (0.15)	68,835 (0.13)	7,652 (0.29)			
2000	253,430 (0.15)	53,001 (0.12)	9,295 (0.29)			
2001	366,356 (0.15)	61,096 (0.13)	3,082 (0.29)		127,195 (0.09)	52,701 (0.23)
2002	314,571 (0.15)	58,384 (0.13)	8,393 (0.29)			
2003	313,012 (0.15)	58,155 (0.13)	28,652 (0.29)	28,371 (0.21)	90,812 (0.07)	32,727 (0.18)
2004	327,615 (0.15)	58,610 (0.13)	26,444 (0.29)	24,940 (0.21)	120,170 (0.07)	40,679 (0.15)
2005	272,051 (0.15)	56,226 (0.13)	19,430 (0.29)	23,831 (0.21)	76,021 (0.07)	37,167 (0.14)
2006	284,760 (0.15)	59,216 (0.13)	17,385 (0.29)	28,857 (0.21)	84,609 (0.08)	31,885 (0.14)
2007	260,341 (0.15)	64,219 (0.13)	10,347 (0.29)	20,610 (0.21)	64,279 (0.07)	38,418 (0.13)
2008	284,816 (0.15)	53,574 (0.13)	10,682 (0.29)	25,961 (0.21)	53,375 (0.07)	23,904 (0.14)
2009	237,528 (0.15)	60,231 (0.13)	7,085 (0.29)	18,121 (0.21)	47,619 (0.07)	25,363 (0.14)
2010	249,808 (0.15)	71,294 (0.13)	8,193 (0.29)	21,072 (0.21)	53,713 (0.07)	19,768 (0.14)
2011	323,139 (0.15)	67,361 (0.13)		19,807 (0.21)	62,348 (0.07)	22,086 (0.14)
2012	238,042 (0.15)	47,246 (0.13)		15,239 (0.21)	56,661 (0.07)	18,221 (0.15)
2013	157,453 (0.14)	43,205 (0.12)		19,796 (0.21)	55,644 (0.08)	27,188 (0.18)
2014	195,741 (0.15)	44,140 (0.13)		13,445 (0.21)	70,614 (0.08)	21,497 (0.15)

2015	154,031 (0.14)	50,596 (0.13)		22,642 (0.21)	71,304 (0.07)	22,152 (0.17)
2016	252,925 (0.15)	51,337 (0.13)		17,881 (0.21)	69,512 (0.07)	18,414 (0.15)
2017	282,693 (0.15)	59,601 (0.13)		28,866 (0.21)	104,078 (0.09)	24,136 (0.15)
2018	313,920 (0.15)	46,208 (0.13)		38,314 (0.21)	103,232 (0.07)	23,998 (0.16)
2019	447,201 (0.15)	77,448 (0.13)		41,964 (0.21)	91,377 (0.1)	25,944 (0.2)

Table 3.9: Input sample sizes for the age-composition data.

Year	Alaska Fixed gear fishery	BC Trap fishery	BC Offshore Standardized Survey	BC Stratified Random Survey	US West Coast Fixed gear fishery	US West Coast Trawl fishery
1982			275			
1983			580			
1984						
1985						
1986			24		71	190
1987			421		110	170
1988			385		44	120
1989			162		68	154
1990				188	52	170
1991				263	47	165
1992			298	462		15
1993			691	445	166	174
1994			434	514	145	137
1995			412	424	134	116
1996			350	244	79	91
1997			-1	46	112	111
1998			278	176	71	114
1999	1,141		322	133	115	142
2000	1,152		288	58	123	132
2001	1,003			62	94	126
2002	1,059		227	216	81	132
2003	1,185	318		121	114	151
2004	1,145	321	282	104	84	128
2005	1,164	316		128	154	146
2006	1,154	250		166	229	166
2007	1,115	328		160	179	168
2008	1,164	364		168	328	155
2009	1,126	432	218	149	354	118
2010	1,159	490	275		341	114
2011	1,190	556	191		351	110
2012	1,165	604	175		433	135
2013	1,157	508	216		358	148
2014	1,126	872	189		429	141
2015	1,179	460	314		571	126
2016	1,169	630	141		442	119
2017	1,190	582	79		323	114
2018	1,174	579			311	112
2019	1,140	582			250	94

Table 3.10: Expected age and precision of age reads for three management regions at selected ages.

	Expected Age (bias, $\tilde{a}$)			**Precision (σ_A^f)**		
Known age (years)	**AFSC**	**DFO**	**NWFSC**	**AFSC**	**DFO**	**NWFSC**
1	2.0	1.2	0.6	~0	~0	~0
20	18.1	22.5	18.0	3.2	2.9	3.3
50	44.3	63.0	56.6	9.2	5.1	8.3

Table 3.11: Description of projected data.

Description	Dimensions	Simulation Method	note
Forecasted recruitment deviations	Vector of length *k* for each of 21 years, where *k* is the number of stocks	Simulated from multivariate normal distribution, preserving spatio-temporal autocorrelation among deviations	This occurs once before forecasting begins.
Forecasted survey observations	Vector of length *f* for each of 21 years, where *f* is the number of survey fleets active in year y	Simulated with observation error from estimated vulnerable biomass, subject to catchability, plus additional variance (0.2) as in historical period	British Columbia standardized offshore survey discontinued in 2009 (BC stratified random survey continues)
Future catches	Vector of length *f* for each of 21 years, where *f* is the number of fishery fleets active in year y	NA; catches are assumed to be known exactly	

Table 3.12: Analytically-calculated survey catchability parameters (q).

Survey	**Block**	**Analytic *q***
AK LL Survey - Subarea A3	1960-2019	1.164
AK LL Survey - Subarea A4	1960-2019	2.376
BC Offshore Standardized Survey	1960-2019	0.160
BC Stratified Random Survey	1960-2019	0.402
WC Trawl Survey - Subarea C1	1960-1995	2.946
WC Trawl Survey - Subarea C1	1996-2003	4.898
WC Trawl Survey - Subarea C1	2004-2019	6.283
WC Trawl Survey - Subarea C2	1960-1995	0.944
WC Trawl Survey - Subarea C2	1996-2003	2.396
WC Trawl Survey - Subarea C2	2004-2019	2.558

Table 3.13: Fishery selectivity curve parameters. The interpretation of the parameters depends on the functional form (see Table 3.7).

Fishery Fleet	Block	p1_Fem	p2_Fem	p3_Fem	p4_Fem	p5_Fem	p6_Fem	p1_Mal	p2_Mal	p3_Mal	p4_Mal	p5_Mal	p6_Mal
AK Fixed Gear	1960-1994	1.335	1.946					1.335	1.946				
AK Fixed Gear	1985-2019	1.335	1.946					1.335	1.946				
AK Trawl	1960-2019	1.946	1.099					2.485	1.946				
BC Longline	1960-2019	3.652	1.698					3.652	1.698				
BC Trap	1960-2019	3.979	2.628					3.979	2.628				
BC Trawl	1960-2019	2.996	0.686					2.996	0.686				
US West Coast Fixed Gear	1960-1997	6.787	-4	2.129	3.982	-5	-1.5	4.871	-4	0.612	3.982	-5	-1.5
US West Coast Fixed Gear	1965-2002	4.181	-4	0.368	3.982	-5	-1.5	3.102	-4	-5.073	3.982	-5	-1.5
US West Coast Fixed Gear	1968-2010	4.601	-4	1.214	3.982	-5	-1.5	9	-4	3.201	3.982	-5	-1.5
US West Coast Fixed Gear	1969-2019	2.2	-4	-9.011	3.982	-5	-1.5	3.065	-4	-9.375	3.982	-5	-1.5
US West Coast Trawl	1960-1982	1	-4	-3.08	-9.031	-4.02	5	1	-4	-3.08	-8.932	-4.02	-5
US West Coast Trawl	1980-2002	1	-4	-3.08	4.081	-4.02	0.322	1	-4	-3.08	4.882	-4.02	-0.489
US West Coast Trawl	1968-2010	1	-4	-3.08	9.982	-4.02	4.916	1	-4	-3.08	6.474	-4.02	5
US West Coast Trawl	1969-2019	1	-4	-3.08	6.09	-4.02	-1.666	1	-4	-3.08	6.411	-4.02	-0.969

Table 3.14: Survey selectivity curve parameters. The interpretation of the parameters depends on the functional form (see Table 3.7).

Survey	Block	p1_Fem	p2_Fem	p3_Fem	p4_Fem	p5_Fem	p6_Fem	p1_Mal	p2_Mal	p3_Mal	p4_Mal	p5_Mal	p6_Mal
AK VAST Survey - Sub-area A3	1960-2019	1.335	2.079	14	20	26	32	1.335	2.079	70	74	78	82
AK VAST Survey - Sub-area A4	1960-2019	1.335	2.079	13	19	25	31	1.335	2.079	69	73	77	81
BC Offshore Standardized Survey	1960-2019	3.902	1.517	15	21	27	33	3.902	1.517	15	21	27	33
BC Stratified Random Survey	1960-2019	3.911	2.303	16	22	28	34	3.911	2.303	16	22	28	34
CC VAST Survey - Sub-area C1	1960-1995	1.645	-4	-5.117	-5.469	-3.688	-1.111	1.645	-4	-5.117	-5.469	-3.689	-1.111
CC VAST Survey - Sub-area C1	1968-2003	1.645	-4	-5.117	-5.469	-3.688	-1.111	1.645	-4	-5.117	-5.469	-3.688	-1.111
CC VAST Survey - Sub-area C1	1976-2019	1.645	-4	-5.117	-5.469	-3.688	-1.111	1.645	-4	-5.117	-5.469	-3.688	-1.111
CC VAST Survey - Sub-area C2	1960-1995	1.645	-4	-5.117	-5.469	-3.688	-1.111	1.645	-4	-5.117	-5.469	-3.689	-1.111
CC VAST Survey - Sub-area C2	1968-2003	1.645	-4	-5.117	-5.469	-3.688	-1.111	1.645	-4	-5.117	-5.469	-3.688	-1.111
CC VAST Survey - Sub-area C2	1976-2019	1.645	-4	-5.117	-5.469	-3.688	-1.111	1.645	-4	-5.117	-5.469	-3.688	-1.111

Table 3.15: Retention curve parameters. Retention is assumed to be 100% for Alaskan fisheries.

Fishery Fleet	Block	β_1^{Fem}	β_1^{Fem}	β_3^{Fem}	β_1^{Mal}	β_2^{Mal}	β_3^{Mal}	β_4^{Mal} (offset)
BC Longline	1960-2019	38.542	8.843	0.969	38.542	8.843	0.969	0
BC Trap	1960-2019	38.845	9.782	0.95	38.845	9.782	0.95	0
BC Trawl	1960-2019	37.785	8.843	1	37.785	8.843	1	0
US West Coast Fixed Gear	1960-1996	37.355	6.005	1	37.355	6.005	1	0
US West Coast Fixed Gear	1974-2010	41.721	6.005	1	41.721	6.005	1	0
US West Coast Fixed Gear	1969-2019	41	6.005	1	41	6.005	1	0
US West Coast Trawl	1960-1981	48.084	2.898	1	48.084	2.898	1	0
US West Coast Trawl	1989-2010	32.199	2.898	1	32.199	2.898	1	0
US West Coast Trawl	1969-2019	41	2.898	1	41	2.898	1	0

Table 3.16: Estimated parameters related to recruitment from final conditioned model.

Symbol	Description	Value
$\tilde{R}_y^k$	Recruitment deviations for stock *k*	See Figure 3.22 **Error! Reference source not found.**
σ_R	Log standard deviation of recruitment deviations	0.71
R_0^k	Log unfished recruitment by stock	lnR_0^1=7.99 lnR_0^2=9.49 lnR_0^3=9.84 $Rln\ {}_0^4$=8.27
τ_y^{ik}	Proportion of recruits spawned by stock *k* allocated to sub-area *i* during year *y*, where *i* is an element of *k*	See Table 3.17 **Error! Reference source not found.**

Table 3.17: Time series of τ^{ik} by stock in the operating model. This parameter is the proportion of recruitment spawned in a given stock *k* that is allocated to sub-area *i*. There are no τ^{ik} parameters for stocks 1 and 4 because they each pertain to only one sub-area.

	Stock 2		Stock 3	
Year	**To C2 (southbound)**	**To B2 (northbound)**	**To B3 (southbound)**	**To A3 (northbound)**
1960	0.219	0.781	0.486	0.514
1961	0.25	0.75	0.406	0.594
1962	0.864	0.136	0.568	0.432
1963	0.552	0.448	0.5	0.5
1964	0.398	0.602	0.309	0.691
1965	0.846	0.154	0.32	0.68
1966	0.712	0.288	0.321	0.679
1967	0.837	0.163	0.416	0.584
1968	0.579	0.421	0.475	0.525
1969	0.76	0.24	0.566	0.434
1970	0.535	0.465	0.53	0.47
1971	0.695	0.305	0.508	0.492
1972	0.831	0.169	0.587	0.413
1973	0.805	0.195	0.663	0.337
1974	0.786	0.214	0.41	0.59
1975	0.754	0.246	0.425	0.575
1976	0.836	0.164	0.392	0.608
1977	0.842	0.158	0.422	0.578
1978	0.927	0.073	0.166	0.834
1979	0.886	0.114	0.241	0.759
1980	0.92	0.0804	0.0564	0.944
1981	0.849	0.151	0.0912	0.909
1982	0.693	0.307	0.258	0.742
1983	0.852	0.148	0.174	0.826
1984	0.714	0.286	0.245	0.755
1985	0.766	0.234	0.249	0.751
1986	0.738	0.262	0.366	0.634
1987	0.661	0.339	0.411	0.589
1988	0.436	0.564	0.241	0.759
1989	0.172	0.828	0.235	0.765
1990	0.154	0.846	0.396	0.604
1991	0.714	0.286	0.484	0.516
1992	0.464	0.536	0.136	0.864
1993	0.657	0.343	0.208	0.792
1994	0.504	0.496	0.198	0.802
1995	0.582	0.418	0.394	0.606

1996	0.938	0.0622	0.145	0.855
1997	0.93	0.07	0.149	0.851
1998	0.901	0.0987	0.0878	0.912
1999	0.95	0.05	0.155	0.845
2000	0.898	0.102	0.177	0.823
2001	0.95	0.05	0.0707	0.929
2002	0.95	0.05	0.109	0.891
2003	0.95	0.05	0.051	0.949
2004	0.95	0.05	0.0525	0.947
2005	0.941	0.0595	0.0679	0.932
2006	0.95	0.05	0.05	0.95
2007	0.95	0.05	0.0812	0.919
2008	0.823	0.177	0.194	0.806
2009	0.95	0.05	0.243	0.757
2010	0.902	0.0979	0.256	0.744
2011	0.915	0.0846	0.187	0.813
2012	0.95	0.05	0.179	0.821
2013	0.715	0.285	0.297	0.703
2014	0.931	0.0692	0.0872	0.913
2015	0.799	0.201	0.219	0.781
2016	0.601	0.399	0.381	0.619
2017	0.933	0.0673	0.258	0.742
2018	0.93	0.0701	0.203	0.797
2019	0.935	0.065	0.204	0.796

3.7 Figures

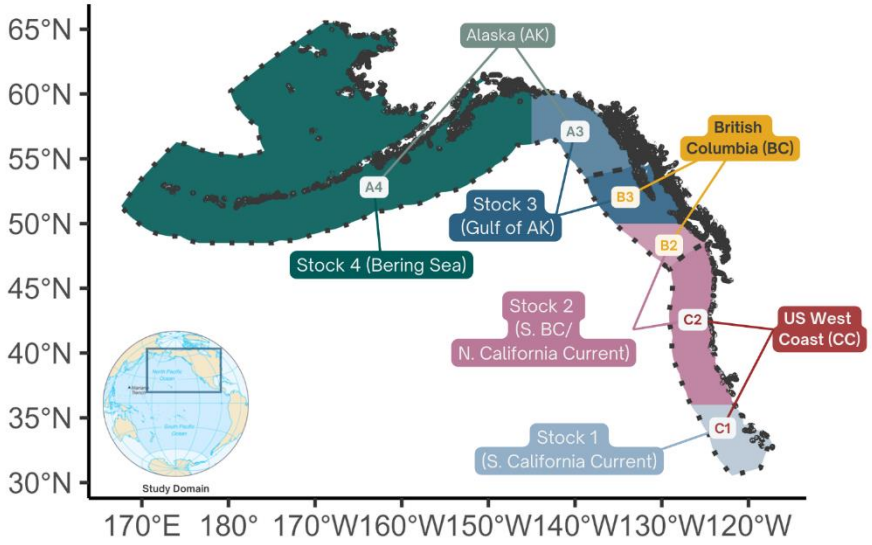


Figure 3.1: Schematic of OM spatial structure. There are six spatial areas (colored polygons), each nested within one of four biological "stocks". The three political management regions are outlined by black dashed lines. Sub-areas are referred to by the alphanumeric codes displayed on the map, with "A" referring to sub-areas within Alaska, "B" to those in British Columbia, and "C" to those off the US West Coast (California Current). Each sub-area corresponds to only one management region and one stock. Figure colors in this document correspond to those shown here.

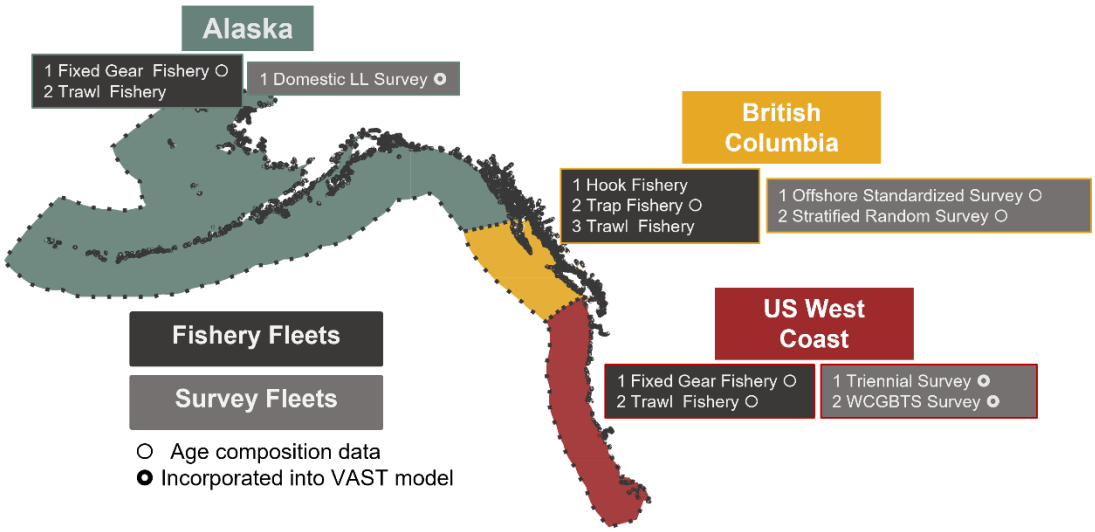


Figure 3.2: Schematic of surveys (light grey boxes), fisheries (dark grey boxes) and associated age-composition data (thin circles) used in the operating model for each of three management regions (colored boxes at top).

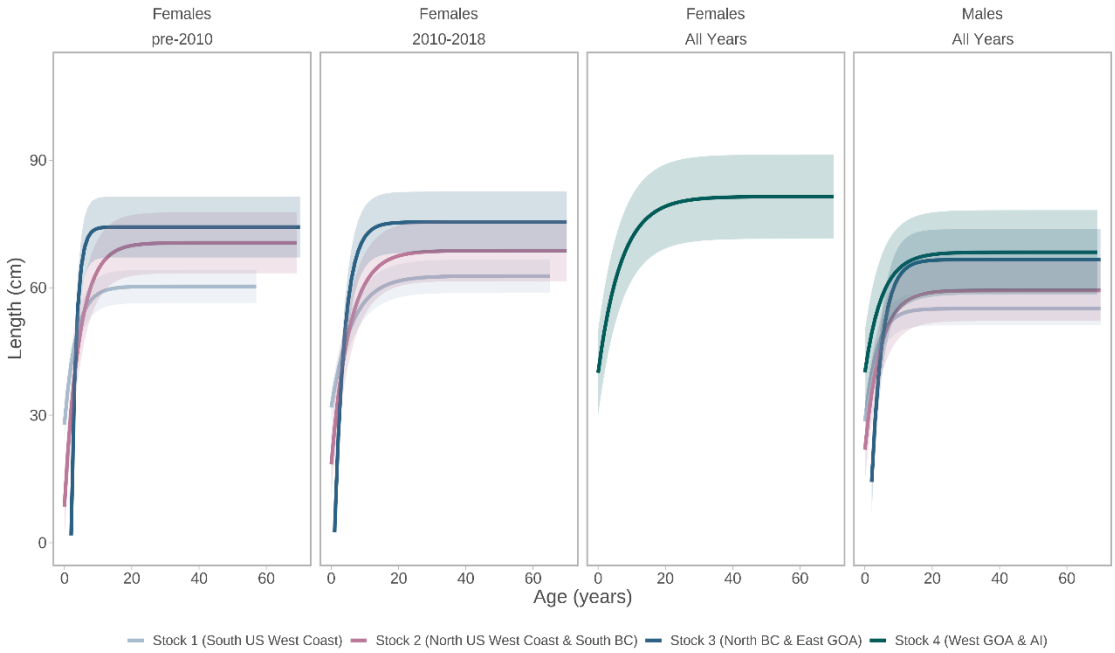


Figure 3.3: Length-at-age curves by stock, sex and time period.

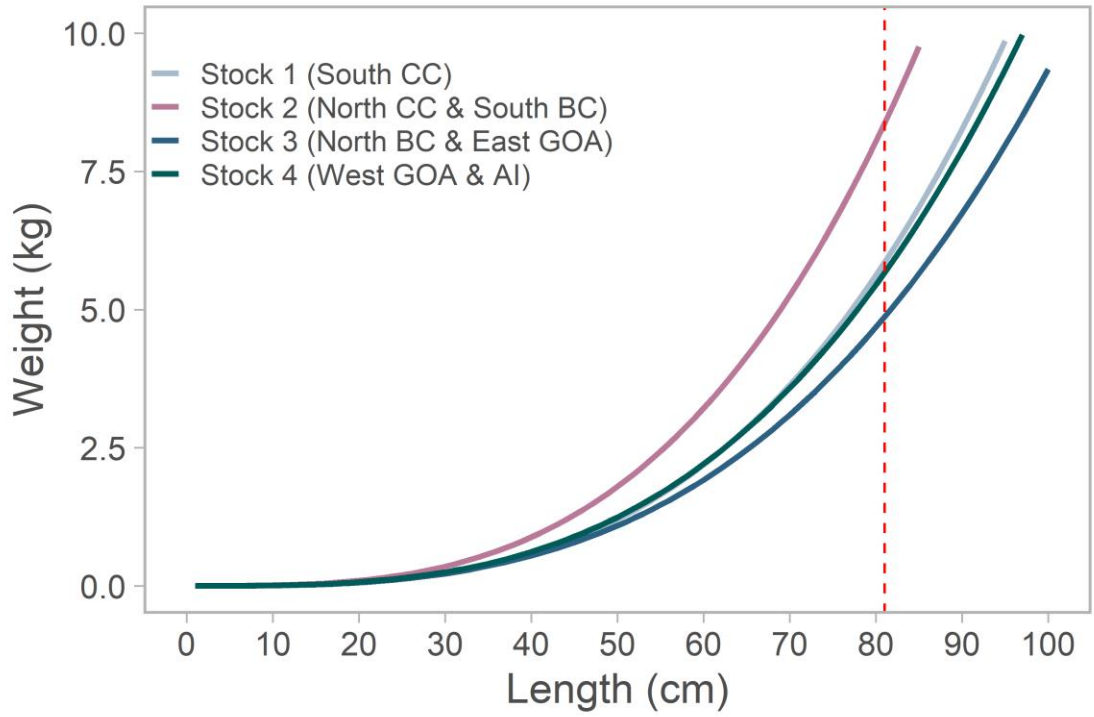


Figure 3.4: Weight-at-length curves by stock. The vertical line indicates the lower limit of maximum length bin in the operating model (81cm).

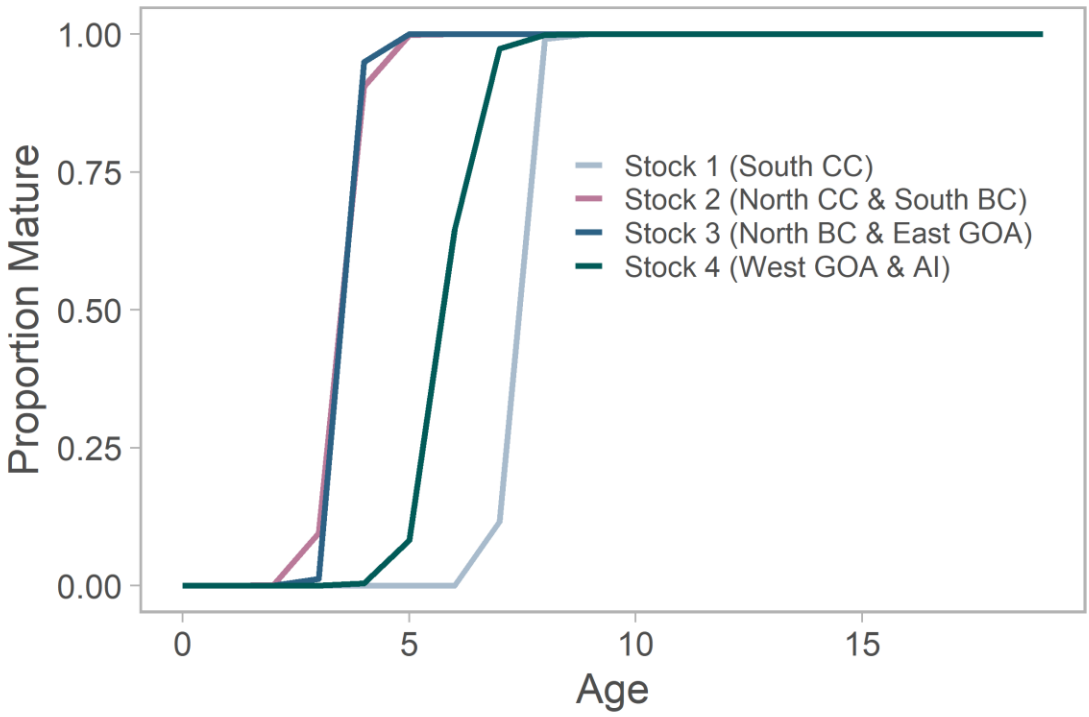


Figure 3.5: Maturity-at-age for female sablefish. Values are only shown through age 20 for clarity. Ages-at-50% and -75%-maturity are estimated by fitting a logistic curve to macroscopic data from all three regions in an analysis external to this analysis.

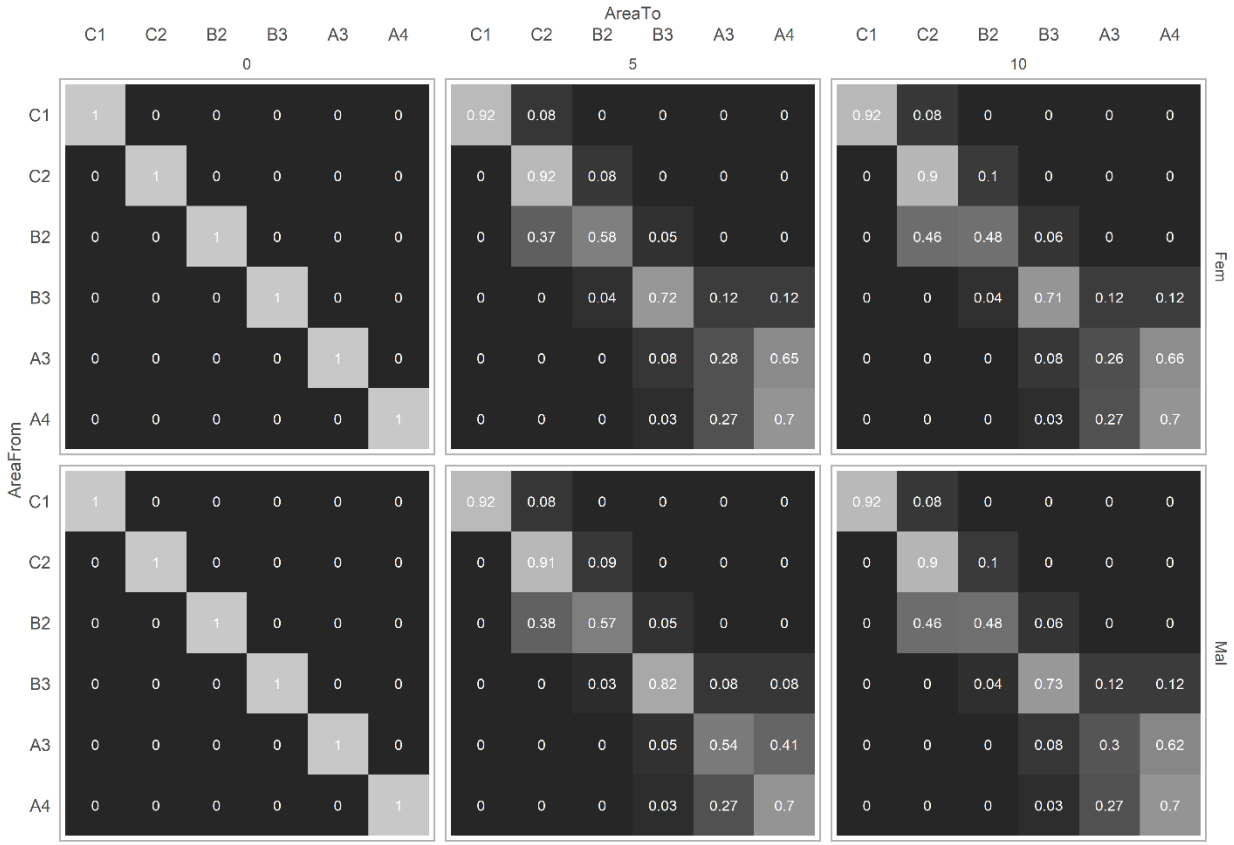


Figure 3.6: Externally estimated input movement rates between sub-areas at select ages (columns) by sex (rows). Movement rates are constant after age 10. Animals younger than age one do not move among sub-areas.

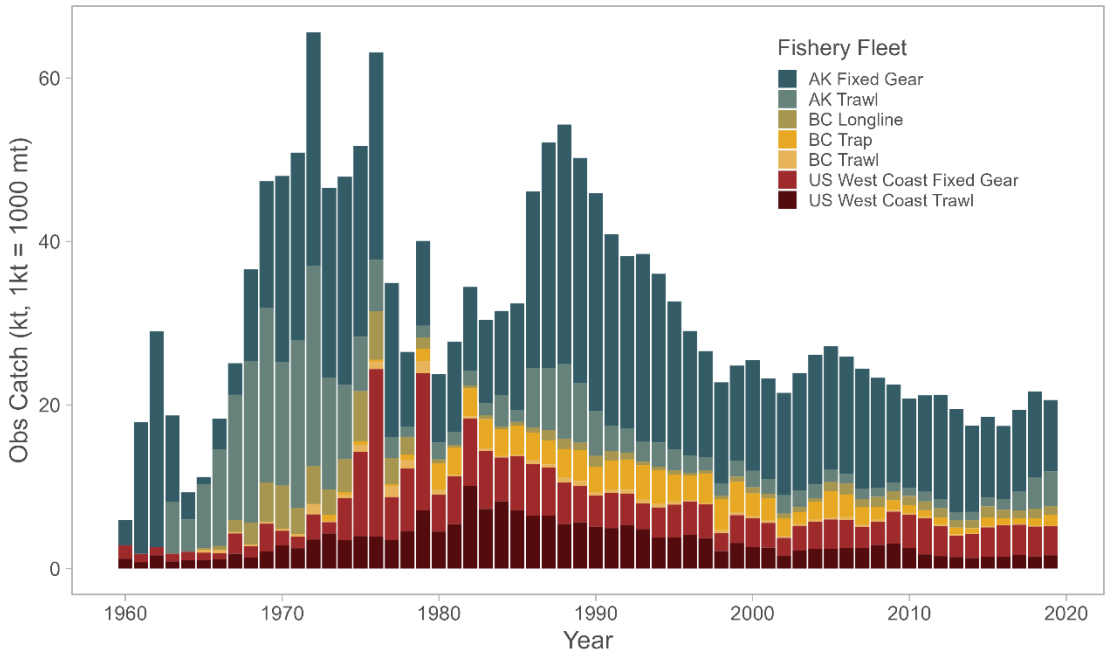


Figure 3.7: Annual fishery catches in kilotons (1kt = 1000 t) used in the operating model.

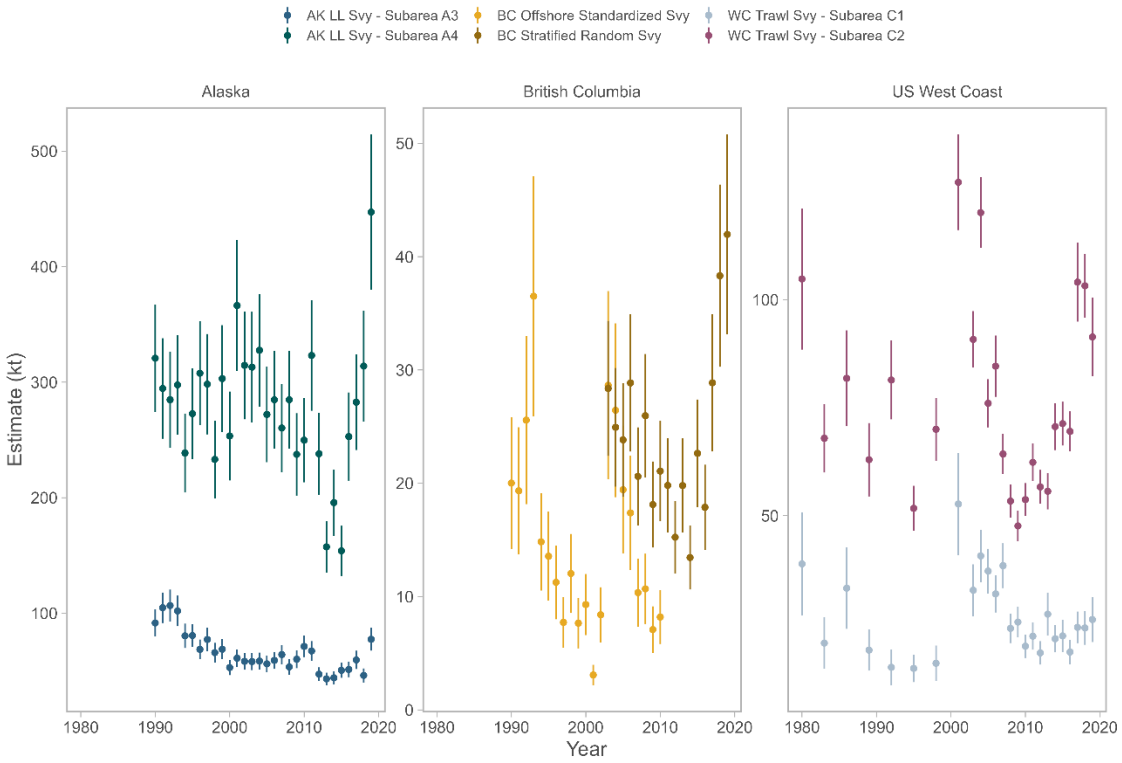


Figure 3.8: Indices of relative abundance developed for and/or included in the operating model. Vertical bars indicate ±1 standard error.

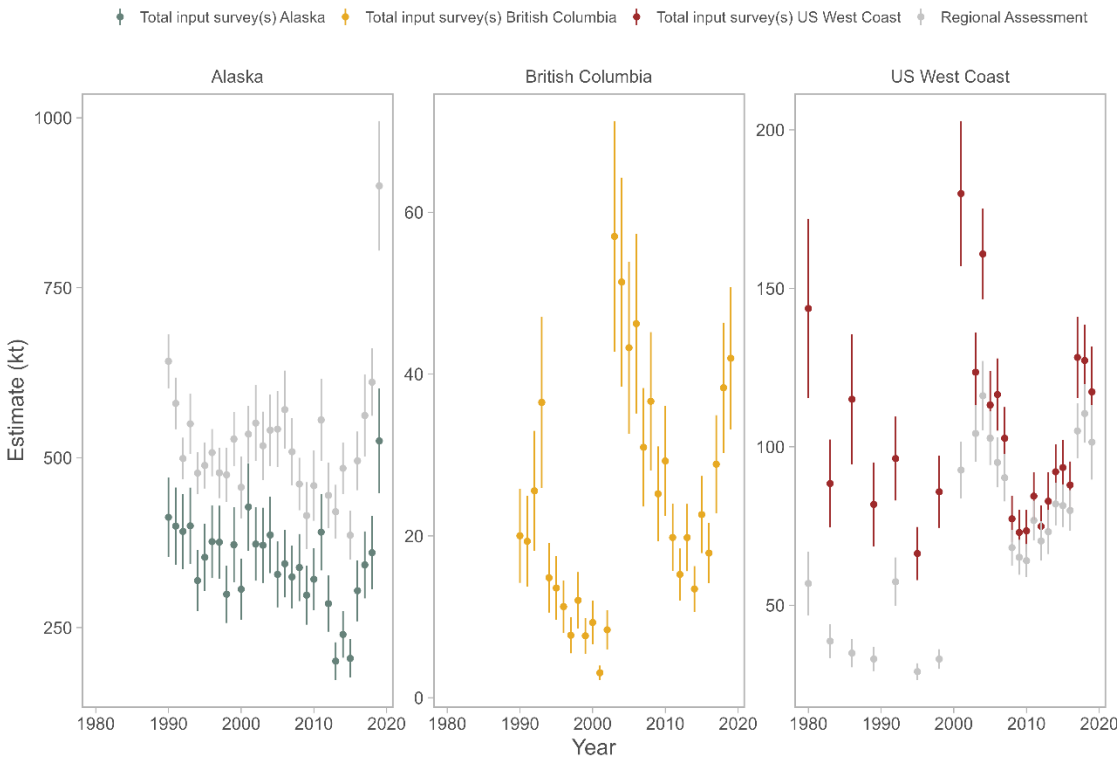


Figure 3.9: Indices of relative abundance included in the operating model (colored points) and those used in recent regional assessments (grey points and bars). Vertical bars indicate ±1 standard error. To aid in comparison, the stratified estimates for Alaska (AK_VAST_A4 and AK_VAST_A3) and the US West Coast (CC_VAST_C2 and CC_VAST_C1) have been summed within their respective regions. The indices for British Columbia are identical to those used in the regional operating model.

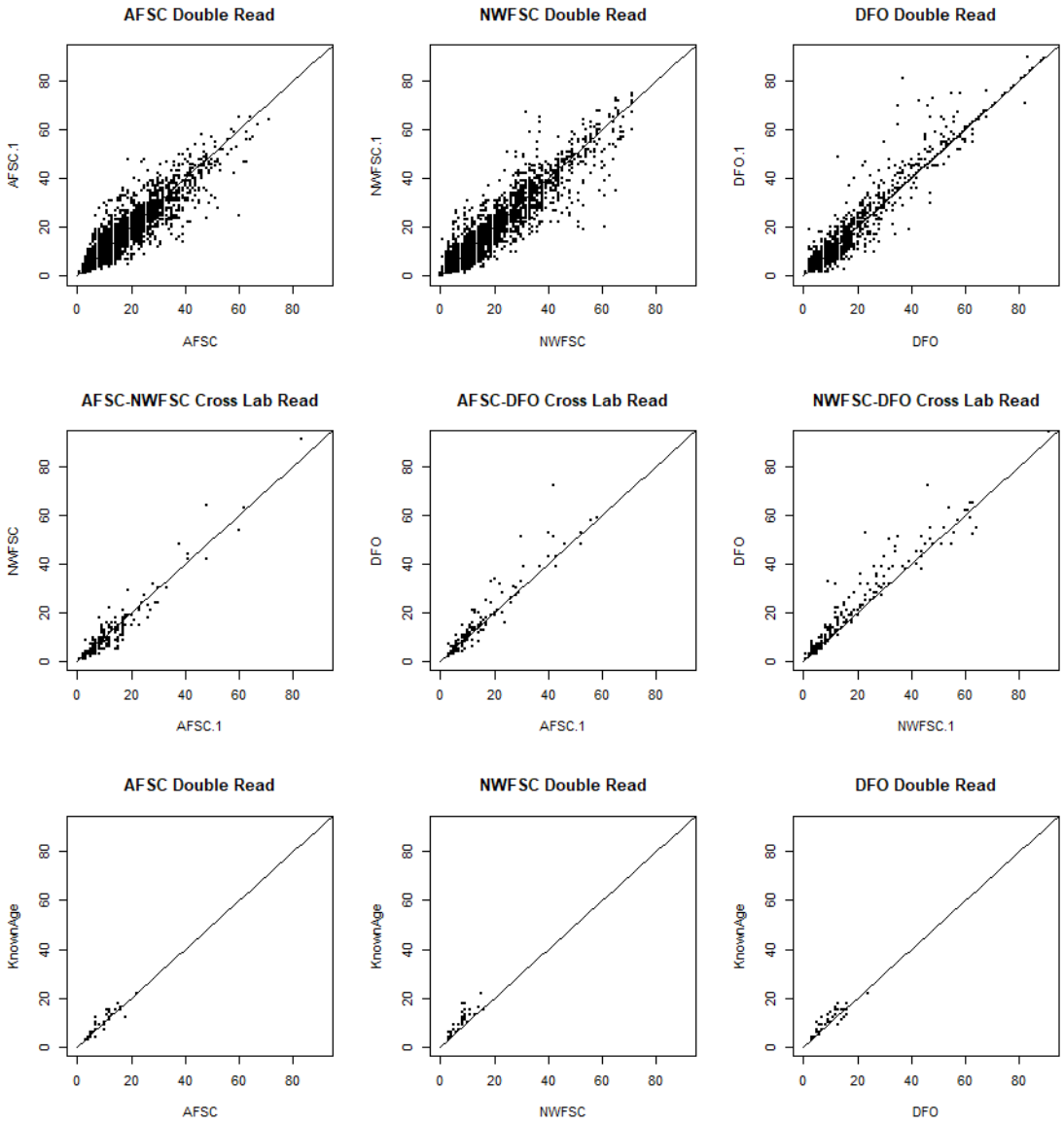


Figure 3.10: All within- and between-lab age data for northeast Pacific sablefish.

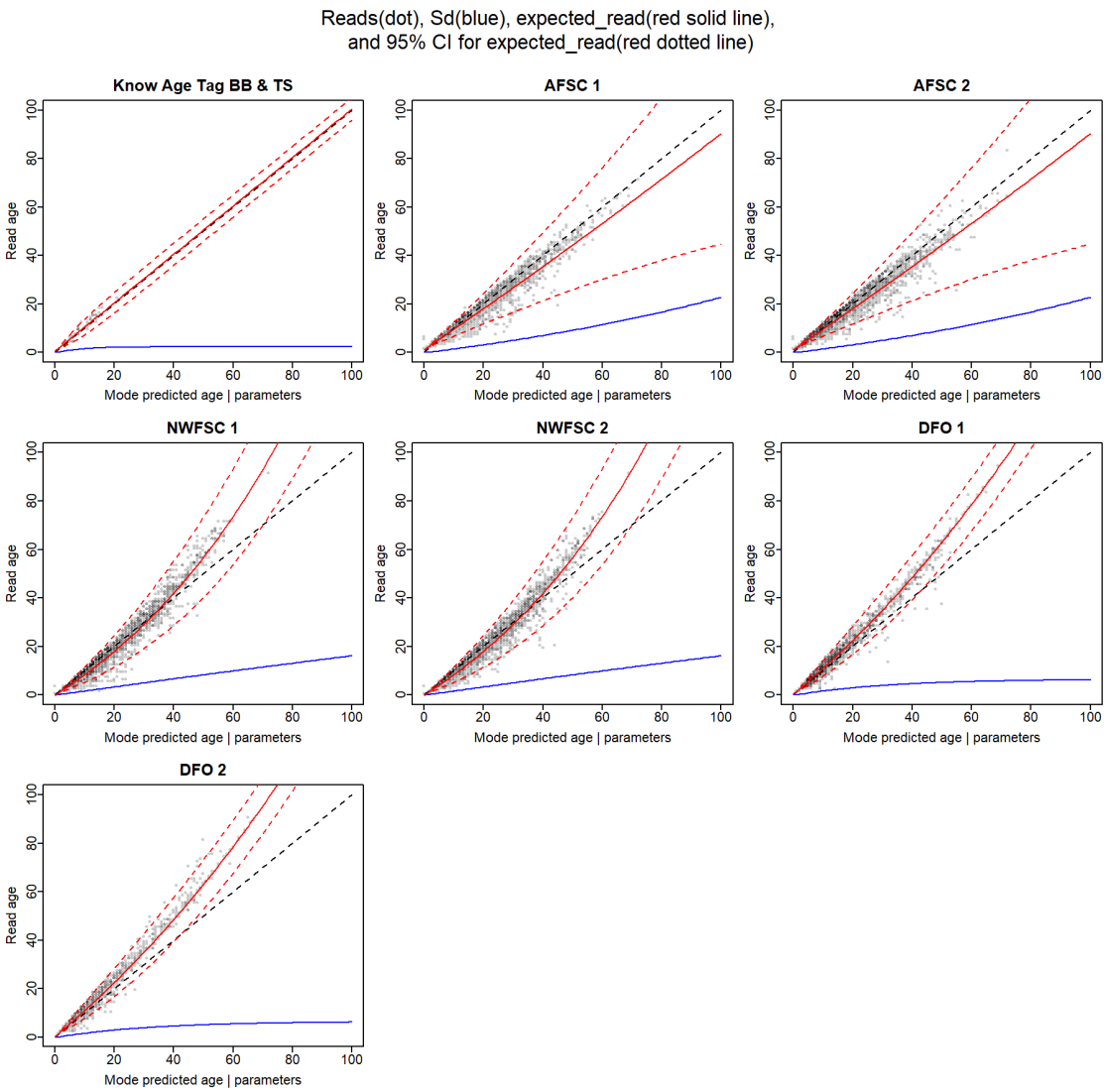


Figure 3.11: True age vs. reads, by reader. Data (points) and model-predicted versus observed ages (red), and standard deviations (blue lines) for each agency and the known age samples. Note that there are two plots for each lab showing the fit to that lab's double read data. The estimates of bias and precision are based on the combined data from each lab (i.e., each lab's reader 2 parameters are mirrored to those for reader 1 to produce one set of estimates).

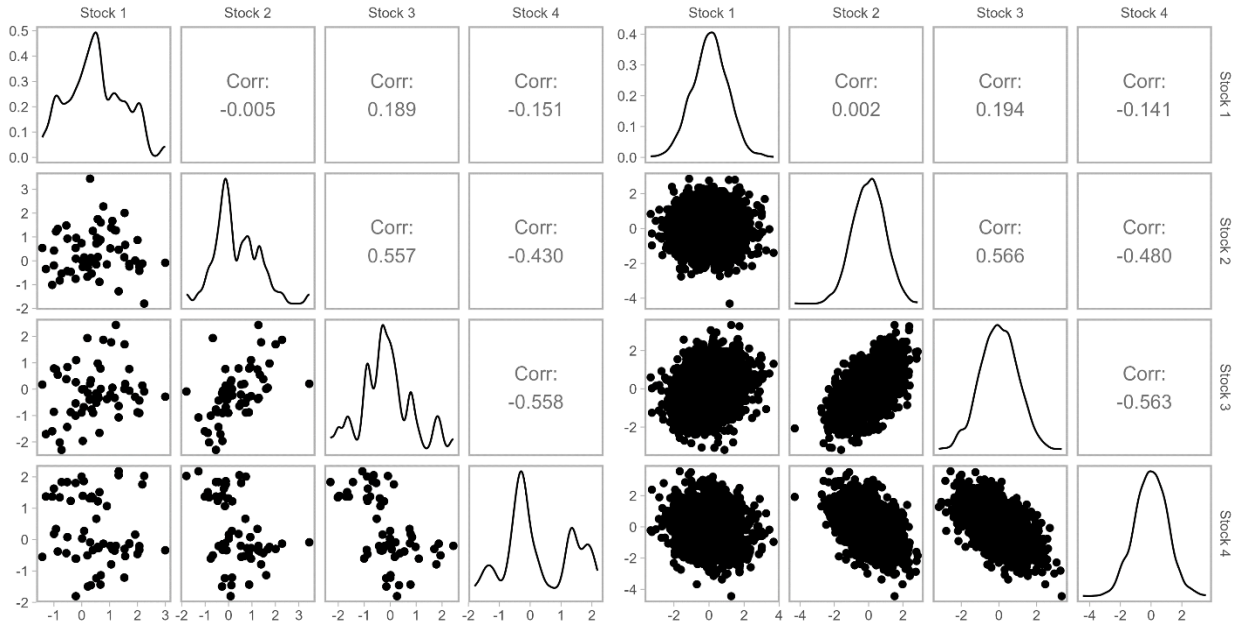


*Figure 3.12: (*Left) Pearson correlations between historical (1960-2019) recruitment deviations from the operating model. (Right) Pearson correlations of simulated recruitment deviations for 100 replicates of 21 years each. Grey values in upper triangle indicate the Pearson correlation coefficient for each stock pair.

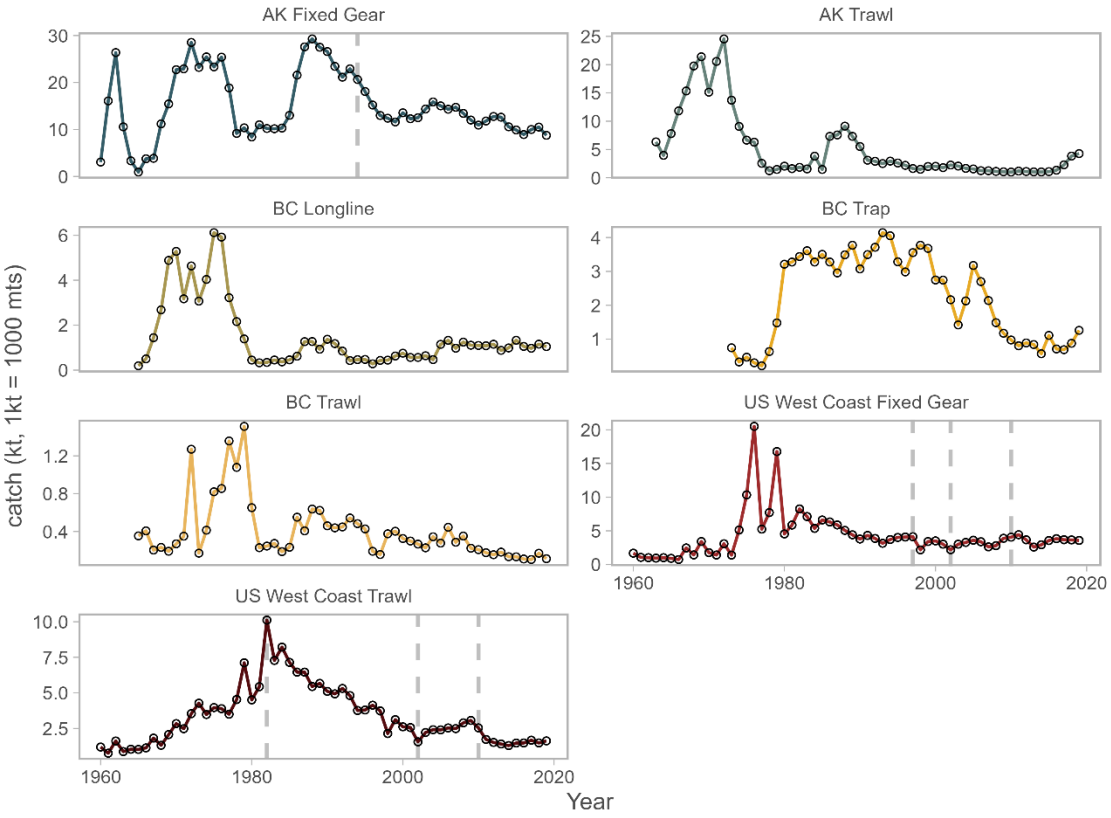


Figure 3.13: Fits to catch data by fishery fleet. Vertical dashed lines indicate time blocks in fishery selectivity.

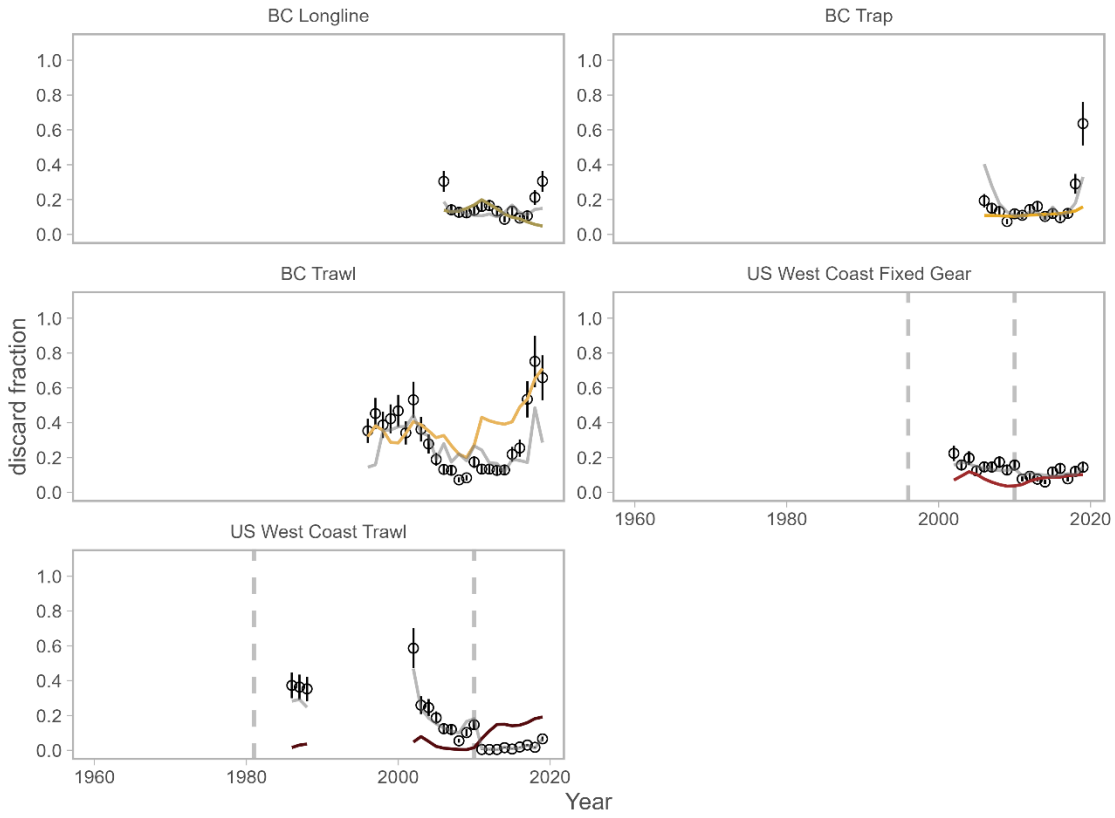


Figure 3.14: Fits to discard ratio data by fishery fleet (colored lines). The light grey lines are the fits to these data in the regional assessments. Vertical dashed lines indicate the time blocks in retention rates.

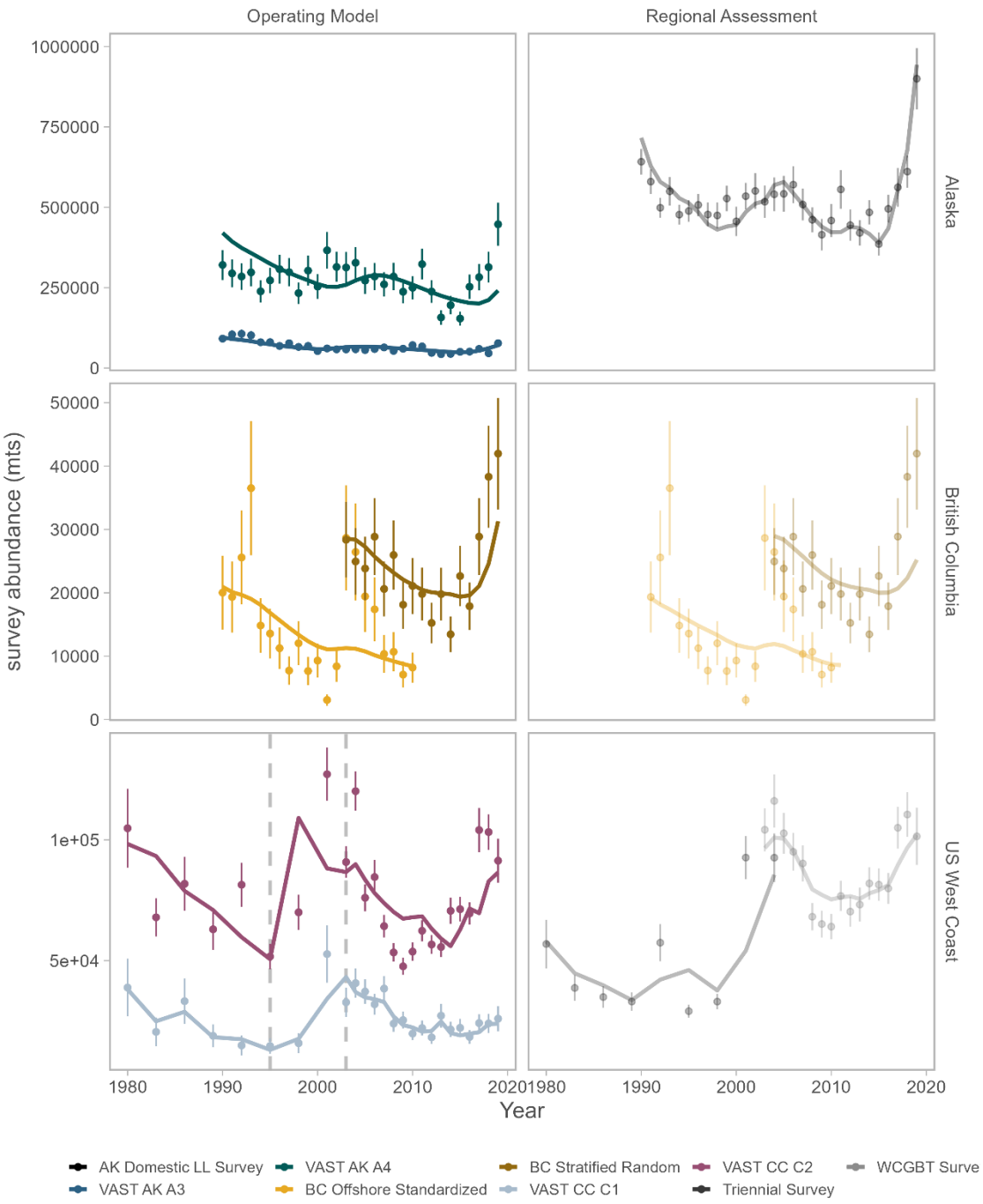


Figure 3.15: Fits to survey data by the OM (left column) and from recent regional assessments (right column). Rows correspond to management regions. The indices of abundance are estimated using VAST at the stratification corresponding to the operating model for Alaska and the US West Coast so the total relative abundance within each management region is roughly comparable to the values used in the regional assessment. Vertical dashed lines indicate time blocks in survey selectivity.

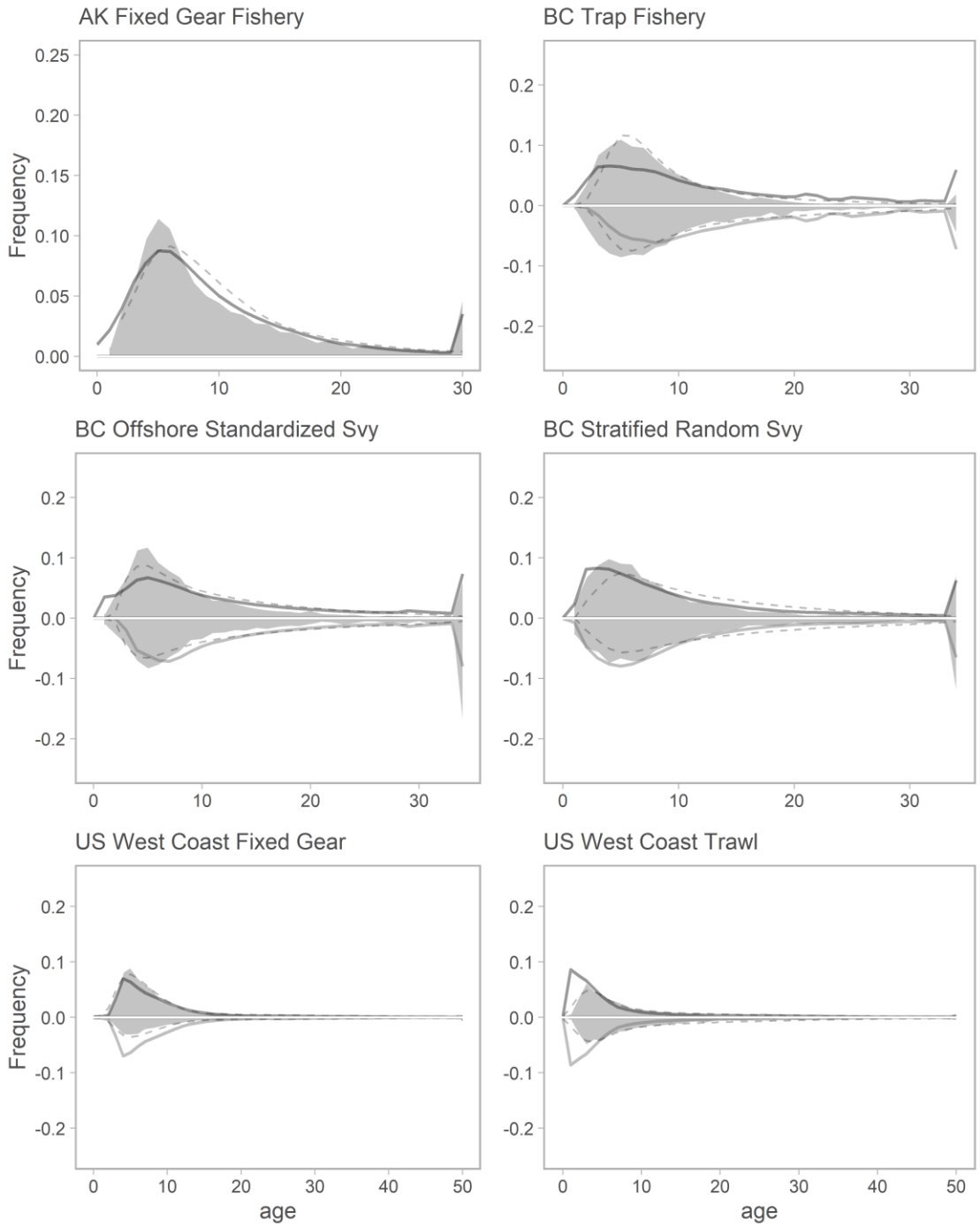


Figure 3.16: Fits to aggregated (all available years) age-composition data by fleet. The grey polygon is observed data; dark grey lines are predicted age-frequencies for females and light grey lines are predicted age-frequencies for males. The Alaska fixed-gear fishery age-composition data is aggregated by sex. Dashed grey lines indicate fits to age composition data in the recent regional assessments.

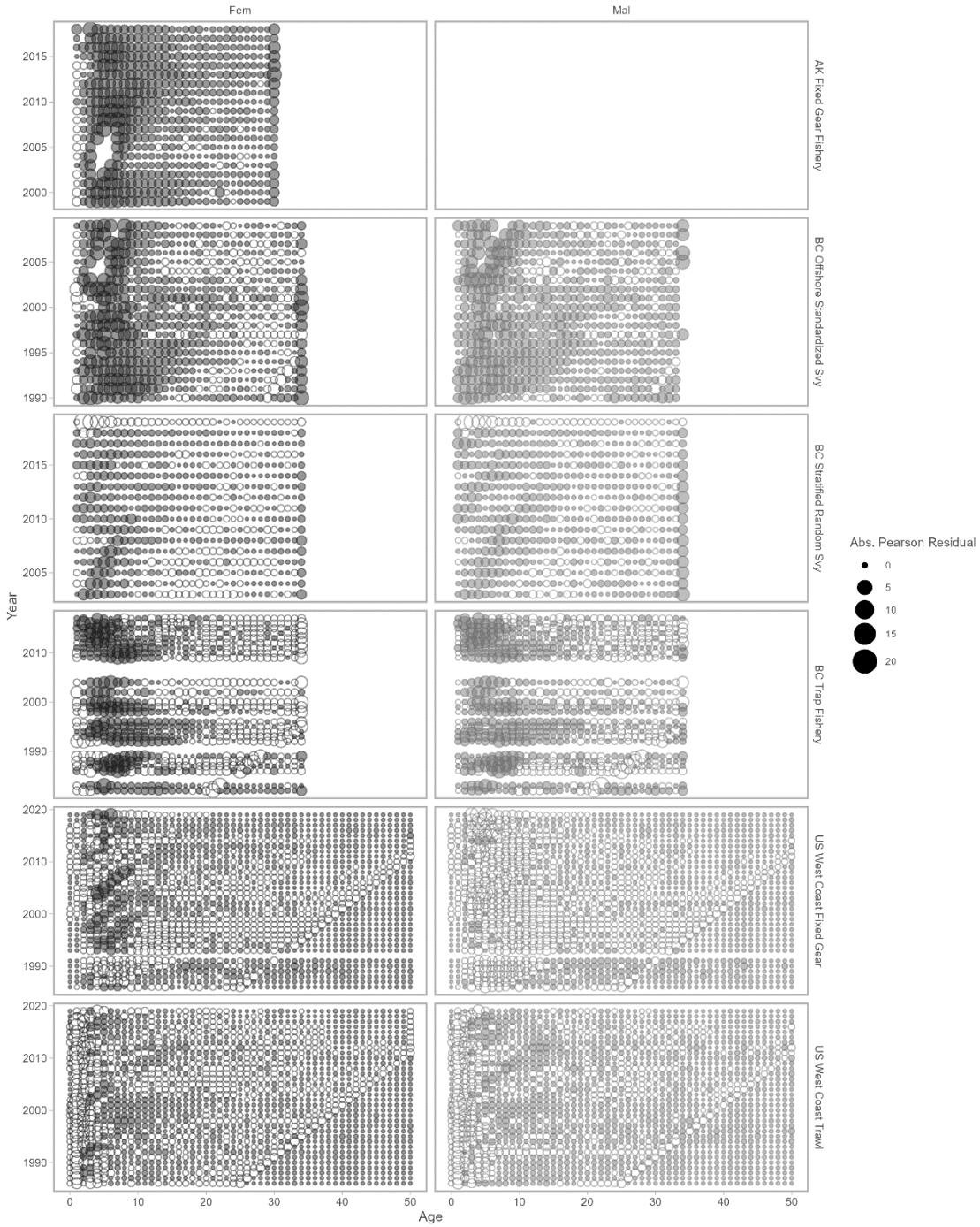


Figure 3.17: Pearson residuals by sex, age, and year for the fishery and survey fleets with age-composition data. Filled circles have positive residuals according to the scale on the right hand side; open circles are negative residuals. The Alaska fixed-gear fishery data are aggregated by sex.

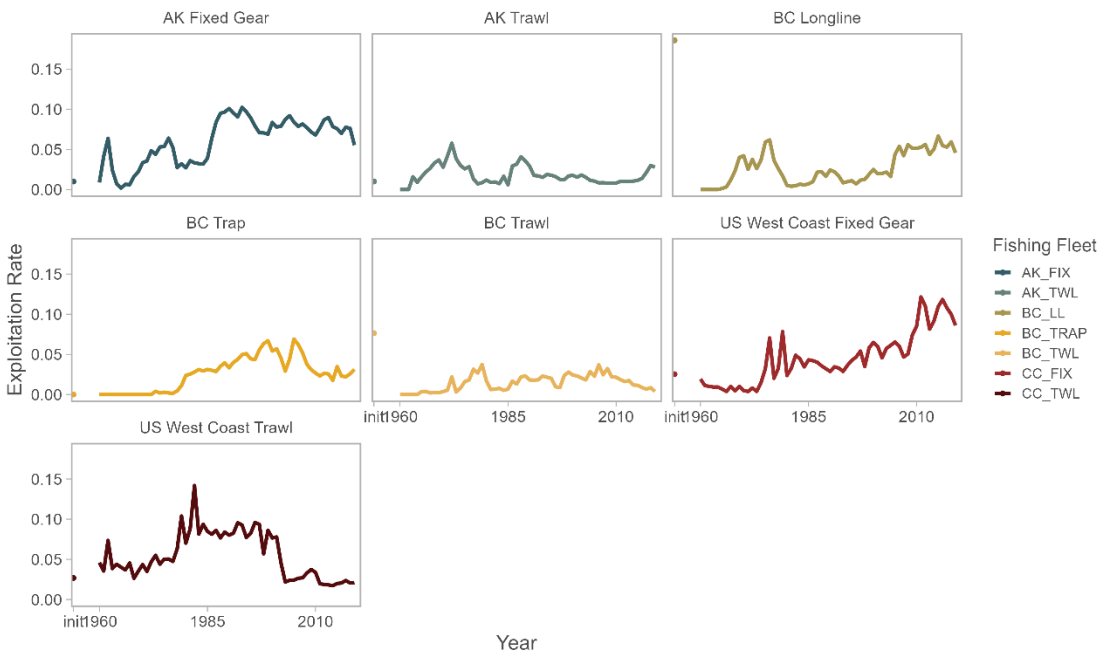


Figure 3.18: Total exploitation rates (both seasons combined) by year and fishery fleet.

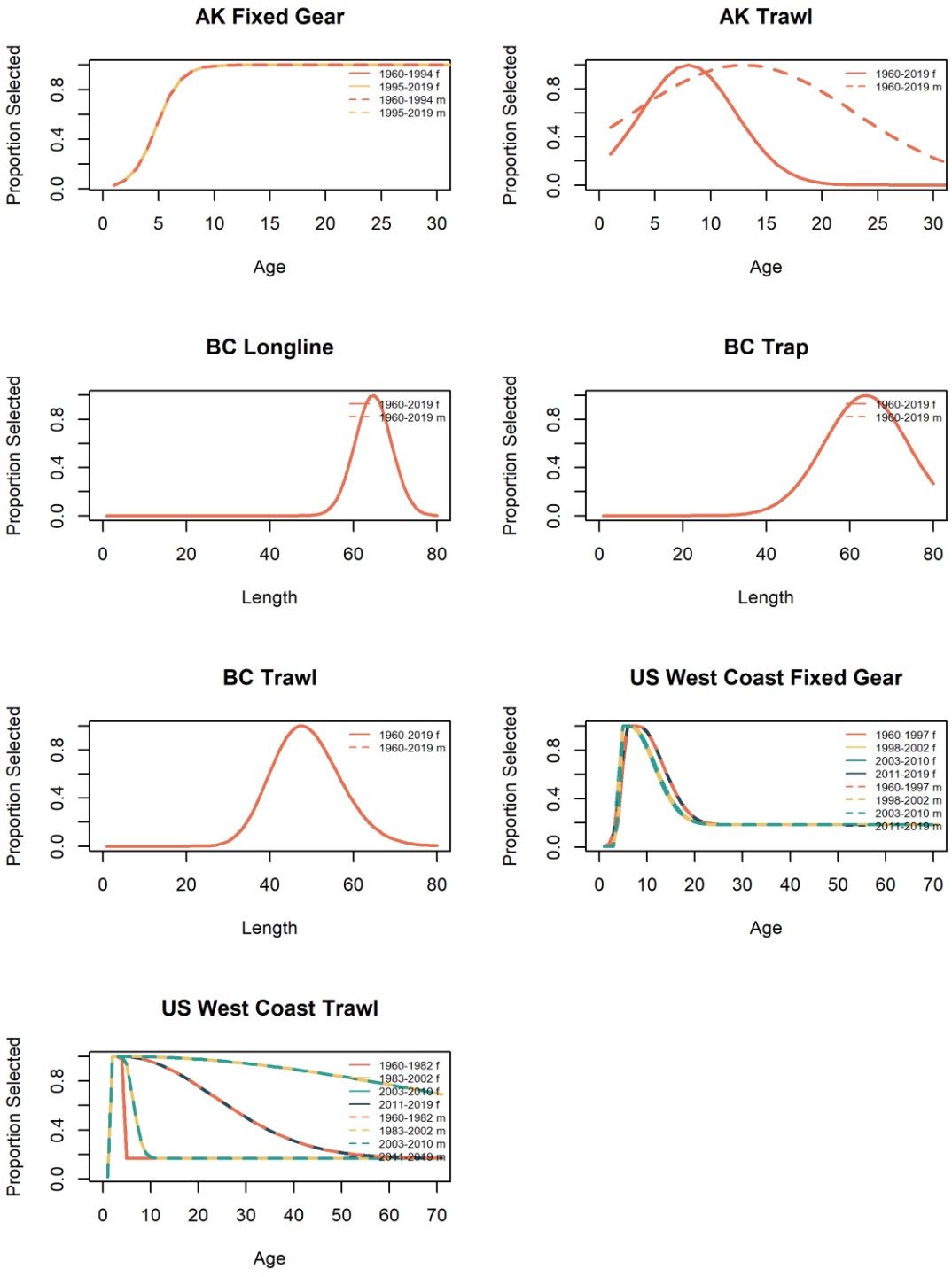


Figure 3.19: Selectivity curves by time block for the fisheries, by age or length and sex.

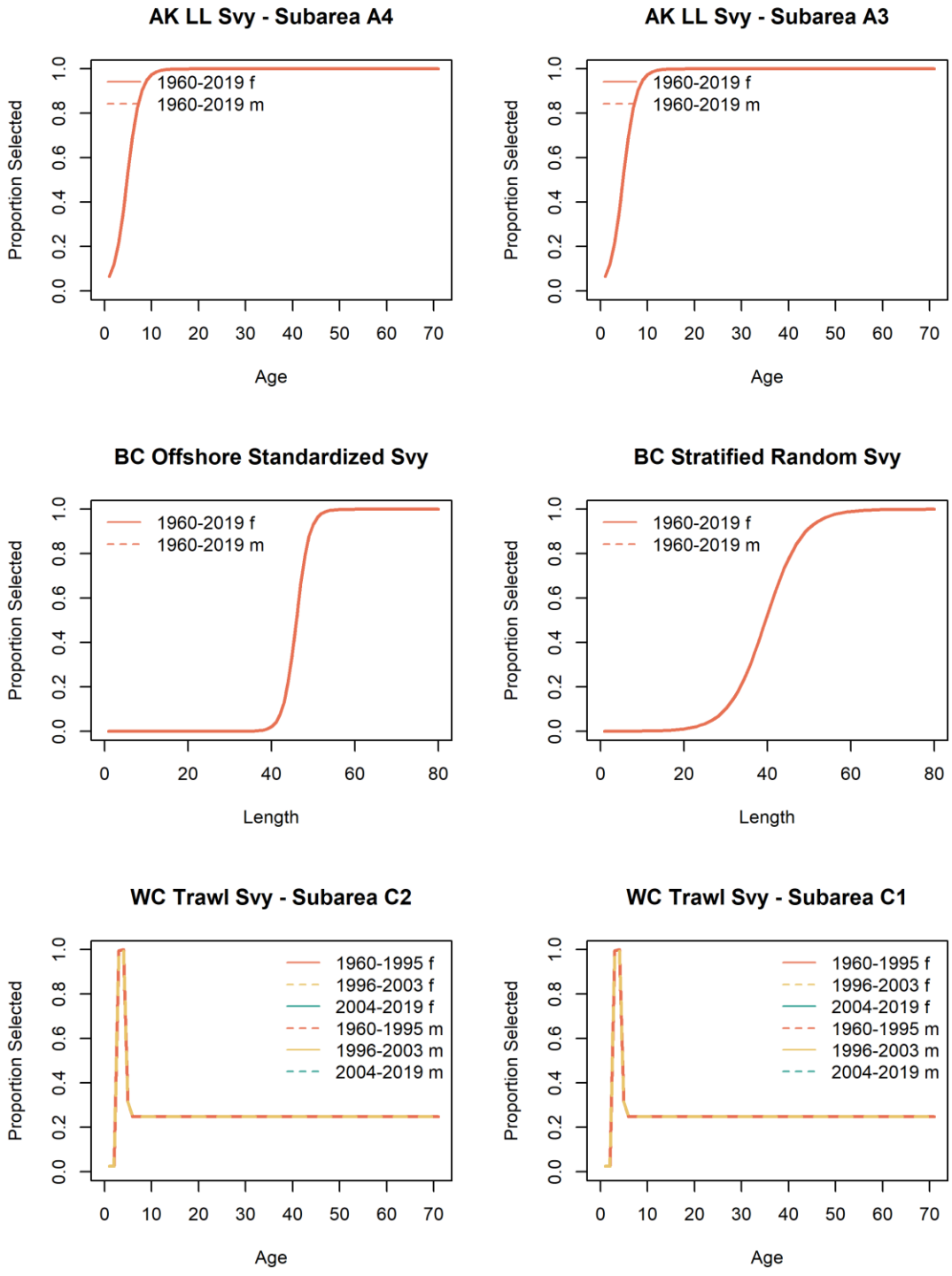


Figure 3.20: Selectivity curves by time block for the surveys, by age or length and sex.

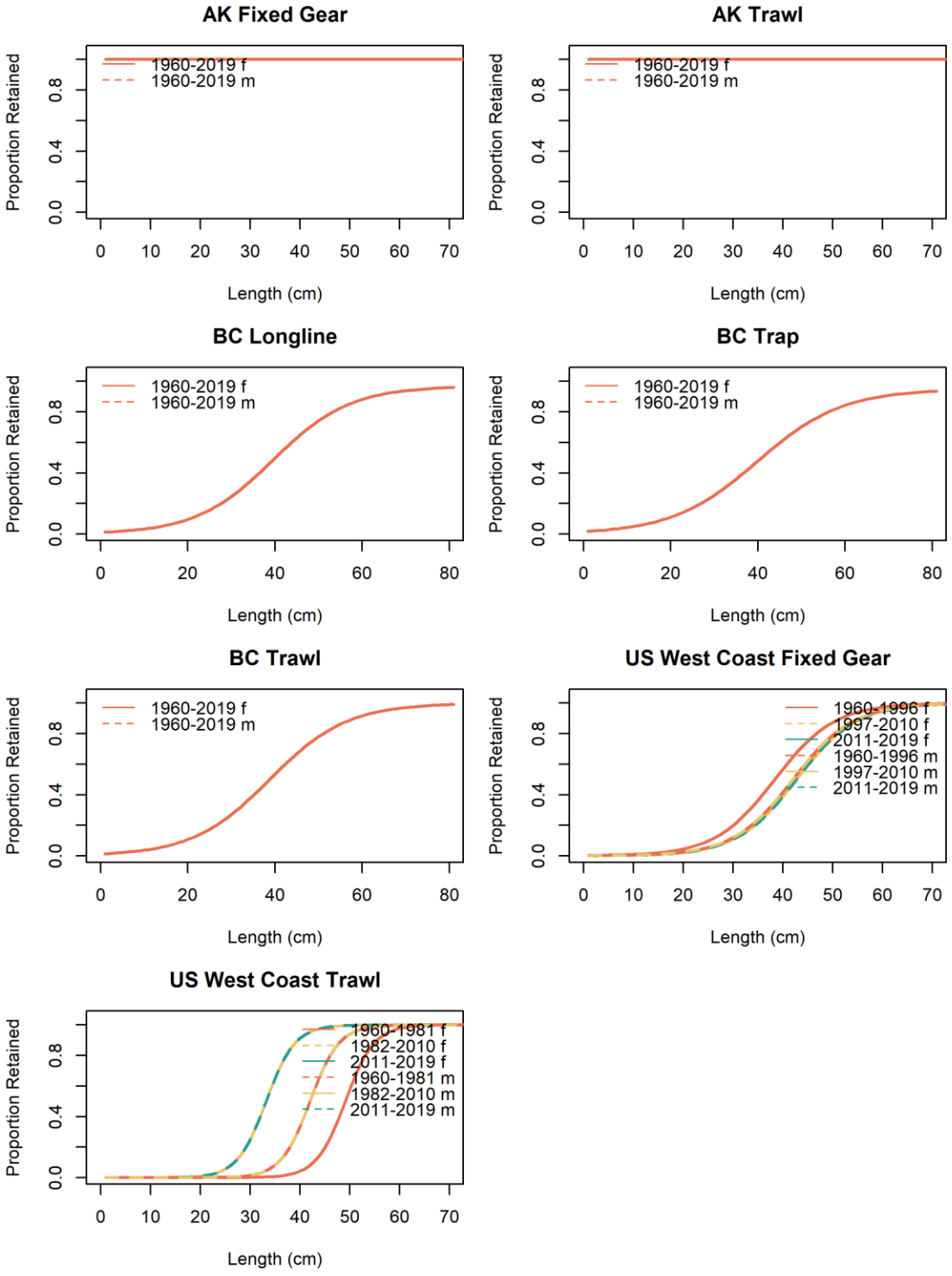


Figure 3.21: Retention curves by time block for the fisheries, by age or length and sex.

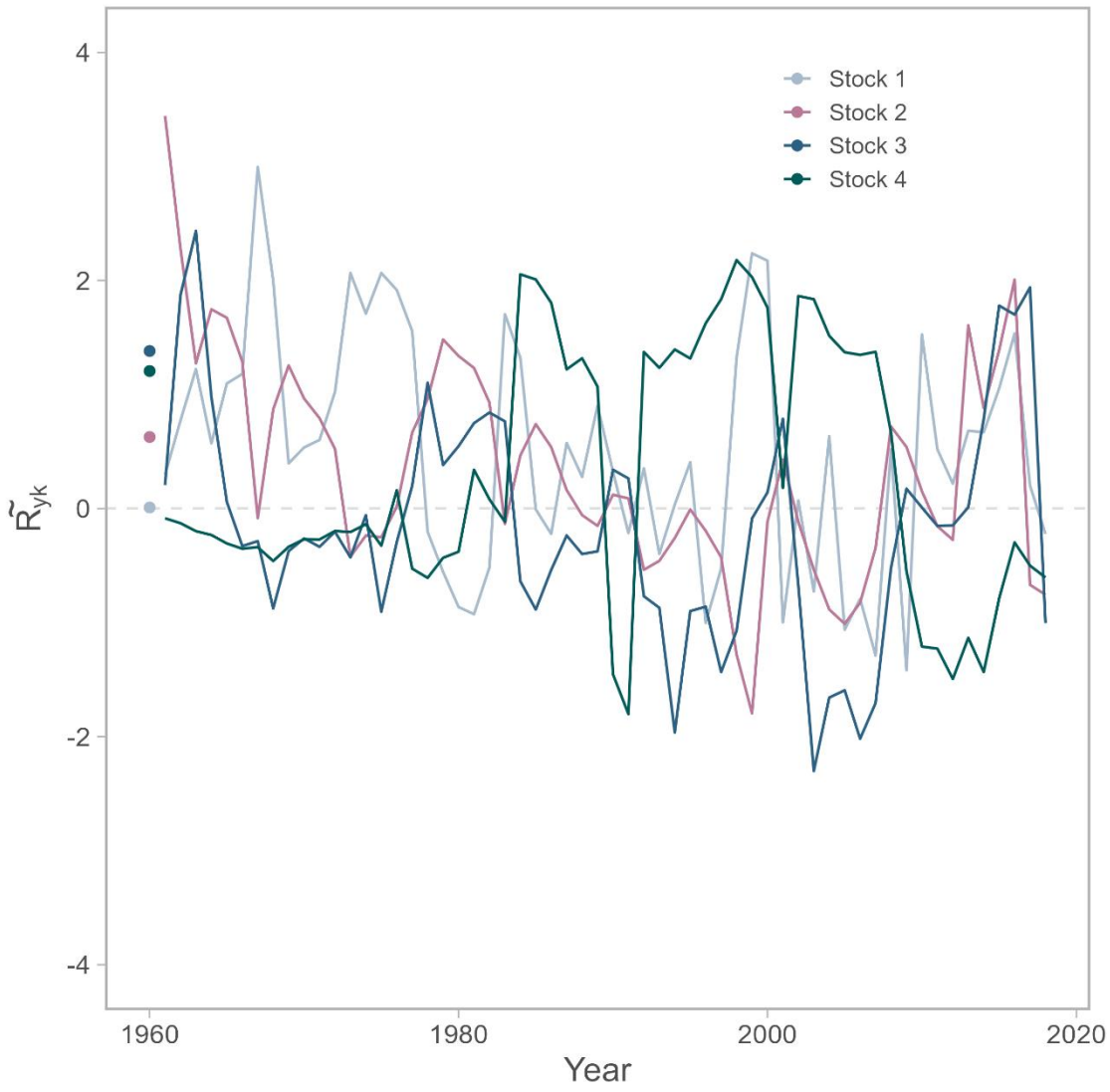


Figure 3.22: Log recruitment deviations by stock. The solid points are the initial recruitment deviation.

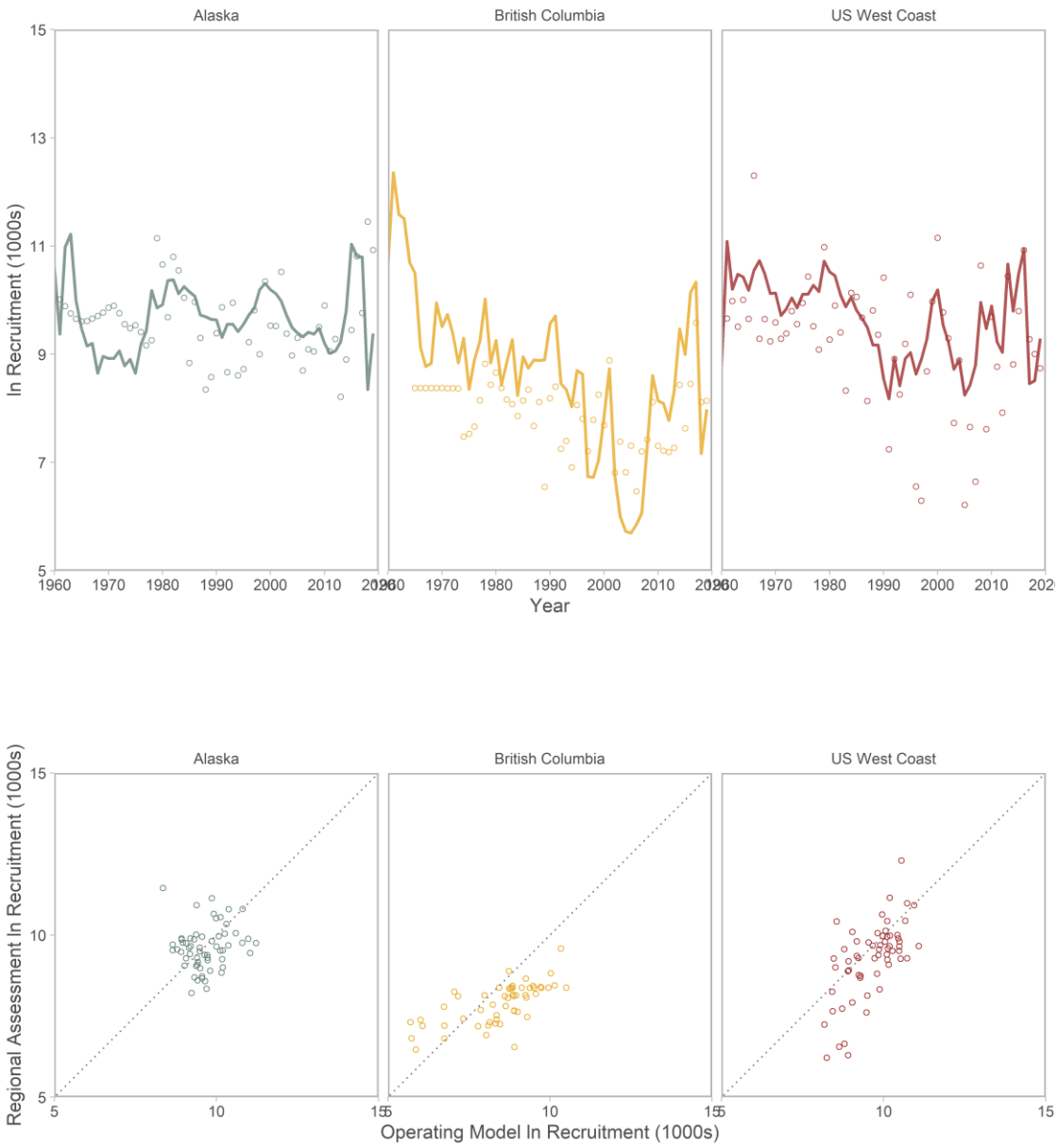


Figure 3.23: (Top) Log-recruitments from regional assessments vs. log-recruitments from the operating model, by management region (panels). (Bottom) Time series of log-recruitments estimated in operating model (lines) vs. recruitment estimated in regional assessments (open points), by management region (panels).

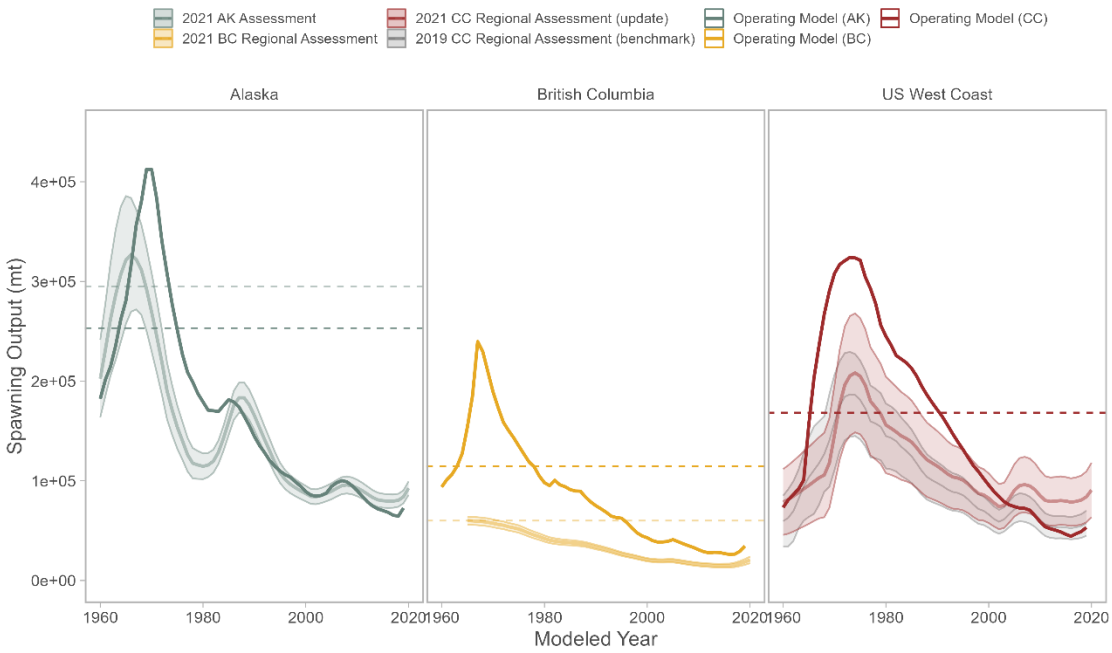


Figure 3.24: Estimated time-trajectories of spawning stock biomass (SSB, opaque lines) by management region (panels). For comparison, transparent lines and shaded areas are estimated SSB and 95% confidence intervals from recent regional assessments. Horizontal dashed lines are unfished SSB from the OM (darker lines) and regional assessments (fainter lines).

Chapter 4. SIMPLE ESTIMATION APPROACHES AVOID POOR MANAGEMENT OUTCOMES FOR TRANSBOUNDARY NORTHEAST PACIFIC SABLEFISH

4.1 Abstract

Sablefish (*Anoplopoma fimbria*) of the northeast Pacific support a highly mobile, valuable fishery resource worth over $112 million USD that employs 8% of the harvesting workforce in Alaska alone. At the federal level, sablefish in the northeast Pacific are currently managed as three separate populations (Alaska, British Columbia and the US West Coast) by the National Oceanic and Atmospheric Administration (NOAA-US) and the Department of Fisheries and Oceans (DFO – Canada). Recent work has shown sablefish to be genetically mixed but across the range, tagging studies have confirmed high movement rates, and biomass trends are synchronous, including declines during the last decade. A management strategy evaluation (MSE) was developed with the collaboration of stakeholders and scientists from the three regions to investigate whether spatially-structured management paradigms, where the population is evaluated as a single but spatially-structured population, might result in better conservation and economic outcomes for sablefish. This effort included synthesizing demographic information (growth, maturity, and movement) and holding stakeholder workshops to determine management strategies and performance metrics. The MSE is comprised of an operating model to represent spatial dynamics, delay-difference estimation method(s) at various spatial stratifications and five harvest control rules to evaluate management strategies of various spatial resolutions. Management strategies that include movement between regions allow for better economic outcomes for the US West Coast and Alaska, with tradeoffs in catch variability. Management performance for British Columbia is not greatly affected by accounting for mixing with adjacent regions. A strategy where the three management regions are treated as separate, unconnected stocks – the closest to the current approach – resulted in neither the optimal nor worst outcomes for any performance metric. Importantly, this work revealed that a mismatch in the spatial scale of the harvest control rules and the conservation units poses a crucial risk of ‘cryptic’ or localized depletion in isolated stocks in the southern U.S. West Coast. This work serves as an example of a stakeholder-informed MSE and illustrates the value of incremental management changes towards meeting biological and economic fishery objectives. Results underscore the influence of movement rates and stock structure on management

performance. Future work should incorporate additional spatial structure hypotheses, estimate movement rates, and investigate the drivers of recruitment patterns coastwide.

4.2 Introduction

Stock assessment scientists and managers of commercial fisheries have long been concerned with spatial structure in assessment modeling and the setting of catch limits because fish populations are not evenly distributed across the seascape. Numerous studies (e.g., Thorson et al. (2015b); Ying et al. (2011)) have illustrated that mis-specifying spatial structure can preclude the ability to accurately assess the size and status of a fish stock, and complicates development of effective management strategies to meet conservation and economic objectives.

Transitioning to a new, spatially-structured paradigm of assessing or managing a fished population can be costly and uncertain. At a minimum, such efforts require an agreed-upon set of spatial areas, a reconfiguration of the input datasets to meet the new spatial stratification, and an approach for reconciling the outputs of the new assessment method with the specification of catch limits. It is therefore helpful to employ a management strategy evaluation (MSE) to assess the tradeoffs associated with changing assessment and management frameworks because such changes add uncertainty, and can lead to dramatic changes to the perception of the status of a stock (Kerr et al., 2014; Szuwalski and Punt, 2015) and of maximum yields (Ralston and O'Farrell, 2008).

Management Strategy Evaluation (MSE) is a modeling approach that simulates the effects of alternative management actions on a fish stock. MSE has been used successfully in several fisheries around the world and has been shown to be an effective way to address the issue of spatial structure in fisheries management. Pacific hake (*Merluccius productus*) are managed via international treaty across the US-Canada border, and recent MSE work has shown that ignoring climate-induced changes in the spatial structure of the stock can lead to lower catches (Jacobsen et al., 2022). Similarly, area-based management was shown to insulate commercial catches from climate-related range contractions of Atlantic surfclam (*Spisula solidissima*), (Kuykendall et al., 2019). The results from an MSE can help managers understand how changes in fishing practices and other factors will impact a stock, and identify the most effective management strategies, before making a formal change.

Sablefish (*Anoplopoma fimbria*) in the northeast Pacific, in common with other high-value and long-lived stocks, have attracted scientific attention due to spatially synchronous biomass trends, including a large-scale decline from 2010-2018 that has only recently started to improve. Recent work indicates that sablefish are well mixed, with little genetic structure from California to the Bering Sea (Jasonowicz et al., 2016), and tagging studies have shown that sablefish readily traverse this distance within one or two years (Goetz et al., 2018; Hanselman et al., 2014). In 2018, scientists from the United States and Canada convened to begin development of a transboundary MSE for sablefish. The principal goal of this effort was to investigate how ignoring the spatial dynamics of northeast Pacific sablefish might affect management outcomes. This study is the culmination of a multi-year MSE endeavor, which included stakeholder input on the operating model and the selection and specification of performance metrics (Fenske et al., 2019; Kapur et al., 2021a). The MSE is comprised of a spatially-structured operating model, a set of management strategies constructed from a state-space delay-difference estimation method, and some harvest control rules applied to the management regions.

The principal questions of this work were: 1) whether ecological and/or economic outcomes for the northeast Pacific sablefish fishery vary across spatial management strategies, and 2) what, if any, characteristics of the northeast Pacific sablefish population present unique concerns for conservation, that may be applicable to other managed populations. The management strategies evaluated represent a gradient of spatial complexity, ranging from the hypothesized demographic structure of the sablefish population to full panmixia. The results illustrate the tradeoffs between assuming, simplifying, or ignoring spatial structure in a management context. Finally, the MSE can be used as a tool to explore additional hypotheses about the ecological and bioeconomic implications of transboundary sablefish management.

4.3 Methods

The management strategy evaluation (Figure 4.1) uses a model of the population dynamics (the "Operating Model", OM) to project a population forward based on a set of management actions (total allowable catches (TACs)). The OM is assumed to represent the truth regarding the population dynamics and generates future data for use in an estimation method (EM). The management actions are based on the derived quantities from the EM, which is fit to the historical and projected data from the OM for each year of a projection period. TACs are determined by

applying the harvest control rules to estimated quantities (current biomass and reference points) from the EM and are assumed to be removed exactly.

4.3.1 Operating model

4.3.1.1 General structure

The operating model (OM) is a two-sex, age-structured, spatially-structured population dynamics model, with an annual timestep from 1960-2019 (the "historical model period"), and intra-year dynamics accounted for using three "seasons". In the OM, six modeled spatial areas are the union of demographic regions, or "stocks", and three political management regions (Figure 4.2a). The spatial structure of the OM allows for nesting of modeled spatial areas within stocks. Spatial areas are the smallest unit at which population dynamics are tracked, and each spatial area contains animals from only one stock and is part of only one management region. Stocks are sets of spatial areas with the same demographic parameters, including those for growth and maturity. The four biological stocks are defined by distinct demographic regimes, with growth described by Chapter 2. Density-dependence (recruitment following a Beverton-Holt stock-recruitment relationship) occurs at the stock level. Fish that move to a new stock assume the demographic characteristics of that stock and contribute to the stock biomass for that stock. The proportion of recruits assigned to each stock's constituent spatial areas is defined by an annual parameter. Movement rates at age are specified among spatial areas based on the results of a tag-recapture study.

The plus-group age A for the OM is 71 years. The plus-group length bin is 81 cm+; numbers- and biomass-at-age greater than age A or numbers- and biomass-at-length greater than 81 cm are collapsed into these terminal groups, after which size-at-age and selectivity are assumed to be constant. The numbers-at-age after movement and all mortality inform the stock biomass and recruitment levels at the start of the subsequent year. For ages 1 through A, the number of individuals in spatial area i of sex γ and age a at the start of year $y + 1$ is the sum of individuals of age $a - 1$ in that spatial area at the start of year y that survive and do not emigrate, or those that immigrate into spatial area i after surviving year y within another spatial area. The plus group in spatial area i is comprised of individuals that are at least of age A and remain in spatial area i, plus individuals that are of at least age A and move to spatial area i from spatial area j (Figure 4.1).

The operating model was conditioned on historical survey, catch, discard, and age-composition data from the entire study domain using Template Model Builder (Kristensen et al., 2016). A detailed description of the operating model, including equations governing the dynamics and the sensitivity analyses is provided in *Chapter 3.*

4.3.1.2 Data Generation

The management strategies investigated in this study (described below) are evaluated based on 100 replicate projections of a single operating model, which each include a 60-year historical period (1960-2019) and 21 'projection' years. The maximum time considered in the MSE spans 80 years (1960-2040). This timescale is appropriate given that sablefish reach maturity and are selected into the fishery for all regions by age 10 (Figures 3.5 and 3.20). The data used during the historical period are unchanged across replicates. Each replicate consists of a unique time-series of deviations from expected recruitment for each of four stocks across the projection period. These future recruitment deviations are simulated to retain the variation within and spatial autocorrelation among stocks. These correlations are highest, in an absolute sense, between adjacent stocks in the US and Canada, with a 57% correlation between the northern California Current/southern British Columbia, and northern British Columbia/the Gulf of Alaska and a correlation of -57% between northern British Columbia/the eastern Gulf of Alaska and the western Gulf of Alaska/Bering Sea/Aleutian Islands (Figure 3.12). The positive correlation between the northern California Current through the western Gulf of Alaska is likely caused by the high level of mixing between the four areas composing these two stocks, and the fact that there is a single survey for British Columbia, which straddles both stocks. The negative correlation across the 145°W line (splitting the eastern Gulf of Alaska from the western Gulf of Alaska, Bering Sea and Aleutian Islands) is consistent with the ecological perception of those areas as separate, large marine ecosystems (Cleaver and Evans, 2019).

Projection involves generating future survey estimates of abundance, but not future discard nor length- nor age-composition data, because these data types are not used by the EM. The future survey observations (Figure D.1) occur at the same frequency as the historical surveys, such that the two US West Coast surveys, the two Alaskan surveys, and the British Columbia Stratified Random survey are each conducted annually. Of the two historical British Columbia surveys only the Stratified Random survey is simulated into the future, because the British Columbia Offshore

Standardized survey ended in 2010 (although that historical dataset is provided to the EMs). Survey estimates of biomass are generated by assuming they are log-normally distributed about their expected values (proportional to the biomass vulnerable to each fishery in the middle of the year), with survey observation variance corresponding to the survey-specific input variance used in the terminal year of the OM (2019).

The future catches provided by the management strategy are assumed to be taken exactly; the harvest rate is calculated to fully realize these catches from the vulnerable biomass in the OM.

4.3.2 Management strategies

Each management strategy consists of an estimation method (a stock assessment method with an assumption about the spatial structure of the population) and a harvest control rule applied to the quantities estimated using the EM (Table 4.1). The management strategies are specified to provide Total Allowable Catches (TACs) for the appropriate regions, as defined by the harvest control rule(s), and to fit the historical and projected survey indices of abundance.

4.3.2.1 Stock assessment (estimation method/EM)

The EM is a state-space delay-difference model with spatial structure characterized by a set of spatial areas. It does not consider age- nor size-structure and is not sex-specific. The spatial areas may or may not coincide with the boundaries of a demographic stock. Fish may move between spatial areas and assume the demographic characteristics (growth rates and stock-recruitment parameters) of their present stock. The EM requires a time series of survey biomass and catches for each fishery. The biomass in each spatial area is defined by the proportion of fish that stay, immigrate, and survive capture and natural mortality in that spatial area; this biomass grows and contributes to the expected Beverton-Holt recruitment for the applicable stock for the following year. For all scenarios, TAC setting occurs at the management level and TACs are allocated to fisheries within each management region. Detailed descriptions, including equations of the EM dynamics, parameterization and objective functions are provided in *Appendix D*.

4.3.2.1.1 Applying the estimation method

4.3.2.1.1.1 Overview

The first five management strategies each involve one of five estimation methods (EMs). These methods are defined by their spatial stratifications and whether fish move among spatial areas. There are two EMs with six spatial areas, two with three spatial areas, and one panmictic, single-area EM (Figure 4.2). Each estimation method is applied to the entire time series of estimates of survey biomass and catches (1960 onwards).

Parameters that can be estimated using the EM include: natural mortality, Beverton-Holt steepness, unfished recruitment, the initial recruitment deviate, the initial harvest rate, a time series of recruitment deviates and the standard deviation thereof, movement rates between spatial areas, and the proportion of recruits in stock k that are assigned to spatial area i in each year (assuming stock k is comprised of more than one spatial area). Estimation is performed using maximum likelihood, implemented in Template Model Builder. The estimation method assumes that annual recruitment represents female sablefish at age 10 because the biomass estimates from the delay-difference model are used as proxies for fully-selected vulnerable biomass and stock biomass. In the OM, 10 years is the age at which 100% of fish in each stock are assumed to be mature, fish are fully selected and after which (where applicable) movement rates are unchanged. Therefore, the values for Ford-Brody demographic parameters concerning the weight of recruits and annual growth rate are estimated externally to the EM at the appropriate spatial stratification using simulated length- and weight-at-age of age-10 female sablefish from the OM. This is achieved by first randomly sampling 5,000 observations of female length-at-age from the OM from each of the six spatial areas shown in Figure 4.2a. These data are then aggregated at the appropriate stock stratification for the EM in question (either the four stocks in Figure 4.2a, three stocks as in Figure 4.2b-c, or a single panmictic stock (Figure 4.2d). The parameters of the von Bertalanffy growth curve (asymptotic length, growth rate, and age at length zero) are then estimated for each stock using non-linear minimization, and are then used to calculate the parameters of the Ford–Walford growth model, which approximates weight-at-age (as in Thorson et al. (2015a); see *Appendix D*.

For all strategies, the initial harvest rate is set to 0.01, as is the case for the U.S. West Coast assessment (and to avoid confounding with initial estimates of recruitment). Natural mortality and

steepness are set to the stock-specific values used in the OM. For the panmictic operating model, these parameters are set to the means of the values used in the OM. An analysis of 30 years of tag-recapture data (Rogers et al., in prep) provided time-invariant movement rates among the same six spatial areas as in *SixAreaFourStockMove* and *SixAreaThreeStockMove* (see below for definitions) as well as among the three management regions for *ThreeAreaThreeStockMove*. Movement rates are not estimated within the EMs. Attempts to estimate the spatial apportionment of recruits as random effects were intractable. Initial runs of the estimation method revealed very little (<5%) change in recruitment apportionment across years within stocks (Table 3.17) so this parameter is assumed to be time-invariant.

Before performing the management strategy evaluation, EMs for each of the five spatial scenarios were evaluated based on their ability to fit the historical survey data, realize the historical catches, and return a time series of biomass in recent years of the same order of magnitude as the recent regional assessments. For each projection year of the MSE, parameters controlling the unfished population size (unfished recruitment and the initial recruitment deviation for each stock), the entire time-series of recruitment deviations and the partitioning of recruitment among spatial areas within a stock (if applicable) were estimated; the variation in recruitment was estimated once using the historical data for EM1 and was set to this value for all other EMs. The treatment of all parameters for each EM and additional figures for these inputs are described in more detail in *Appendix D.*

The study design is balanced so that the same unique replicates were used for all management strategies. Replicates were discarded for all strategies if any estimator failed to minimize for any year during the projection period. For all successful model runs, the invertibility of the Hessian matrix was recorded. This approach follows the recommendation by Punt et al. (2016a) that the management strategy be based only on how it would be applied in practice. The only information shared between the OM and EM are the simulated datasets passed during the projection period, the values of parameters such as natural mortality and steepness that were assumed to be known and were not estimated in any EM (described above), and a subset of parameters (such as those related to growth) are based on data simulated from the OM.

4.3.2.1.1.2 Spatial scenarios

Spatial scenarios (Figure 4.2; Table 4.1) are used to evaluate the five estimation options. These range from complex, biologically-representative (spatial and stock structure determined by demographic analysis) configurations to approximations of the status quo (spatial and stock structure determined by political boundaries, with or without movement among spatial areas). The final scenario is a panmictic model (no spatial structure). Stock- or spatial area-specific parameters are estimated at the stratification applicable to the scenario. The five spatial estimation methods are as follows, with the name used to refer to each in italics:

- Management Strategy 1: *SixAreaFourStockMove*. This mimics the spatial structure of the OM, in that there are six spatial areas, with movement among them, and four stocks. Two of the stocks are each comprised of two spatial areas and straddle an international boundary (Figure 4.2a).

- Management Strategy 2: *SixAreaThreeStockMove.* Management as stocks, with six spatial areas and movement. This EM has the same six spatial areas as the OM, with movement among them. Instead of four stocks, demographic parameters are instead shared only within the three management regions, each of which is comprised of two spatial areas (Figure 4.2b).

- Management Strategy 3: *ThreeAreaThreeStockMove*. This is a three-area EM with movement among the three spatial areas. The spatial areas reflect the political borders of the United States and Canada (British Columbia, BC). As in *SixAreaThreeStockMove* demography is shared within each of the three spatial areas, which are now each comprised of a single, coincident spatial area. This represents the current spatial stratification of sablefish assessments in the region (though movement is not part of the current management paradigm). The observed survey abundances specific to the six OM areas ($B^{f}_{obs,y}$) and their uncertainties area summed into the appropriate management region (Figure 4.2c).

- Management Strategy 4: *ThreeAreaThreeStockNoMove*. This is identical to *ThreeAreaThreeStockMove*, but there is no movement among the three spatial areas (Figure 4.2c).

- Management Strategy 5: *Panmixia*. This EM treats the entire region as a single stock, with a single set of demographic parameters. The data from the fisheries and survey observations are summed into a single fishery and survey, assumed to act throughout the assessed area (Figure 4.2d).

4.3.2.1.2 Stock status and harvest control rules (HCRs)

4.3.2.1.2.1 Reference points

Several of the HCRs considered in the MSE require estimates of the reference points MSY and B_{MSY}, which are computed using the estimation method for each stock. However, the calculation of equilibrium reference points for spatially-structured models is an open area of research (Kapur et al., 2021c). Because the management strategies explored here already differ substantially from the current management paradigm (both in terms of model type and spatial structure), the reference points are calculated using equilibria that do not consider movement among areas. This provides reasonable confidence that differences in performance among management strategies can be attributed to the management strategy itself, and not the method used to calculate reference points.

The principal reference point from the delay-difference model for the U.S. West Coast and British Columbia is U_{MSY} (the exploitation rate corresponding to MSY), which requires the equilibrium quantities defined in the following equations. Calculation of U_{MSY} employs a search algorithm to numerically solve the equation:

$$\left.\frac{\delta C(U)}{\delta U}\right|_{U=U_{MSY}} = 0 \qquad (4.1)$$

where $C(U)$ is the equilibrium yield when the harvest rate is U, i.e., $U\phi_{eq}(U)R_{eq}(U)$. The yield corresponding to U_{MSY} is the maximum sustainable yield MSY; and B_{MSY}, the biomass supporting this yield is given by $\phi_{eq}(U)R_{eq}(U)$. For Alaska, $U_{40\%}$ is used as a proxy for U_{MSY}, where $U_{40\%}$ is the harvest rate associated with a spawning potential ratio of 0.4. $U_{40\%}$ is then the maximum harvest rate applied when biomass is at or above $B_{40\%}$ (the long-term average biomass that would be expected under average recruitment and $U = U_{40\%}$). The yield corresponding to U_{MSY} is $Y_{40\%}$. A search algorithm is used to solve for this quantity given the life history parameters for Alaska:

$$\left.\frac{\phi_{eq}(U)}{\phi_0}\right|_{U=U_{40\%}} = 0.4 \qquad (4.2)$$

Reference points need to be computed for TAC setting purposes at the scale of the management regions, but the boundaries of the modeled stocks do not always match those of the management regions. For these cases, the biomass reference point B' for management region m is computed by

summing the estimates of B' by stock weighted by the estimated proportion of the biomass of stock k in management region m, i.e.:

$$B'^{m} = \sum_{y,i\in m} B'^{i} \frac{B^{i}_{Obs,y}}{\sum_{i\in k} B^{i}_{Obs,y}} \qquad (4.3)$$

This weighting is based on the observed (for the past) and simulated (for the future) survey abundance in each spatial area, averaged across all available years of data from the OM. For all U.S. spatial areas, a single survey operates in each spatial area i. For British Columbia, the Stratified Random Survey is assumed to sample both spatial areas in British Columbia, so the observed biomass is instead weighted by the expected survey biomass for year y in each British Columbia spatial area. The resultant target harvest rate U^{m}_{MSY} is then re-calculated as MSY^{m}/B'^{m}_{MSY}. For Alaska, the target harvest rate $U^{m}_{40\%}$ is calculated in the same fashion using $Y^{m}_{40\%}/B'^{m}_{40\%}$.

4.3.2.1.2.2 Harvest Control Rules

Catch limits are set on a per-management-area basis using HCRs, which may or may not be unique to each management region. The actual observed catches are used for year 2020. There are three sets of HCRs (Figure 4.3; Table 4.1). The first set includes those currently used in each management region. Generally, these are recti-linear algorithms that ramp harvest rates (and subsequent TACs) up or down between two pre-specified limits and scale to the current estimated biomass from the EM (B_y). A general form for these algorithms is:

$$\mathrm{TAC_y} = \begin{cases} 0 & if B_y \le lim_1 B' \\ B_y U_{targ} \dfrac{B_y - lim_1 B'}{lim_2 B' - lim_1 B'} & if lim_1 B' < B_y \le lim_2 B' \\ B_y \mathrm{U}_{targ} & if lim_2 B' < B_y \end{cases} \qquad (4.4)$$

For the US West Coast, U_{targ} is set to U_{MSY} (Equation 4.1), B' to B_0, and lim_1 and lim_2 are 0.1 and 0.4, respectively. For Alaska, U_{targ} is set to $U_{40\%}$ (Equation 4.2) B' to $B_{40\%}$ (the long-term average biomass that would be expected under average recruitment and $U = U_{40\%}$), and lim_1 and lim_2 are 0.05 and 1, respectively. For British Columbia, U_{targ} is 0.055, B' is B_{MSY}, and lim_1 and lim_2 are 0.4 and 0.6, respectively. Currently, the management procedure for British Columbia

modifies the TAC returned by the HCR shown above in accordance with the "two stage method" (S.D.N. Johnson, Landmark Fisheries, *pers. comm.*): the proposed TAC is only accepted if it exceeds the previous year's TAC by greater than or equal to 200 t (otherwise, the previous year's TAC is retained). This modification stabilizes TACs in that region. For the analysis presented here, the second stage is not implemented to ease comparison in performance across regions.

The second HCR is an empirical HCR (*eHCR*) that considers the slope of a regression of the logarithms of survey observations in each management region during the most recent five years. The TAC is then the proportional increase or decrease based on the slope (or mean of slopes, if more than one active survey) multiplied by the mean catch from 2009-2019.

$$\mathrm{TAC}_y^m = (1 + s_y^m)\bar{C}_{2009-2019}^m \qquad (4.5)$$

where s_y^m is the slope (or mean of slopes) of a linear regression performed on the most recent five years of log survey observations for all surveys operating in management region m, and $\bar{C}_{2009-2019}^m$ is the mean total catch by all fisheries operating in management region m during the terminal ten years of the OM (2009 to 2019). The five-year averaging window does not continue into the projection period because this can lead to undesirable oscillatory behavior, as the survey data responds to the changing catches (A.E. Punt, University of Washington, *pers. comm*).

The final HCR (*meanCatch*) also does not consider the outputs of the estimation method, and instead sets the catch for each fishery to the corresponding mean catch during the last 20 years of the historical period to introduce some contrast between this and the empirical HCR.

The first set of HCRs is combined with each of the five estimation methods to create the first five management strategies (Table 4.1); the other two HCRs do not require outputs from the estimation method and are treated as standalone management strategies, for a total of seven management strategies.

4.3.2.1.2.3 Fishery apportionment of TACs

For all HCRs, TACs are apportioned from management regions to fishery based on the proportion of the total 2019 catch taken by each fishery by weight. This retains the current apportionment paradigm used in each region for the duration of the projection period.

4.3.2.2 Performance Metrics (PMs)

Two types of metrics are used to quantify the performance of each management strategy: those related to biological conservation, and those related to economic goals (Table 4.2). These goals were taken directly and/or adapted from those determined during a series of stakeholder workshops held during April 2021 (Kapur et al., 2021a). These criteria were not used to discard management strategies, but to compare among them. The biological performance metrics include: the number of years in which the stock biomass is above the overfished limit (see below), the number of years in which the stock biomass is above a general precautionary limit, here 40% of unfished biomass, and the mean fish length during the first ten years of the projection period. These metrics are evaluated for each of the four biological stocks in the OM (Figure 4.2a). Economic metrics include: the average catch during the first and last five years of the projection period, the total catch during the entire projection period, the average annual variability in the catches, and how often that annual variability exceeds 15% (a threshold selected by stakeholders). The economic metrics are evaluated for each of the three management regions (Figure 4.2a).

In practice, the definition of "overfished" varies by management region: in the U.S. West Coast, this occurs when the level of depletion (the ratio of stock biomass in year y to unfished stock biomass) is less than 0.25; for British Columbia, it occurs when the stock biomass falls below B_{MSY}; for Alaska, it occurs when the ratio of stock biomass to the stock biomass at maximum sustainable yield falls below 0.5. For this study, the U.S. West Coast definition is applied to all three regions. This avoids the uncertainties associated with calculating equilibrium-based reference points (i.e., B_{MSY}) from a spatial operating model, and eases interpretation of performance for stocks that straddle management regions (as the regional definitions of "overfished" would vary among adjacent spatial areas within British Columbia).

The performance metrics are described in Table 4.2 and are calculated on a per-replicate, per-strategy basis, which becomes the raw "score". Equal weight is applied to all performance metrics. For summarization, scores are adjusted such that a high value always reflects a positive outcome: for example, the complement of average annual variability in catch is used so that scores closer to one correspond to lower variability. Then, scores for each performance metric are scaled to the maximum for each unique performance metric-area combination, such that a score of 1 represents the highest score obtained for each stock or management region for a given metric across all

management strategies. The sum of the medians of the scaled scores for each performance metric is used to compare across strategies within each spatial unit (either stocks, for biological metrics, or management regions, for economic metrics). The median raw scores (i.e., actual catches obtained in t) are used to evaluate the relative performance and describe tradeoffs among strategies.

The performance of each strategy is also evaluated by qualitative fits to the historical and projected survey and catch data. The mean relative error (MRE) of key derived quantities (unfished stock size and depletion) is summarized to aid investigation of the impacts of spatial misspecification on estimation performance.

4.4 Results

The size of the OM population in terms of unfished spawning stock biomass S_0 is comparable to values estimated in recent regional assessments (Figure 3.24). S_0 values in the conditioned OM are 253,237 t, 114,468 t, and 168,072 t for Alaska, British Columbia and the US West Coast, respectively. The recent regional assessment estimates of S_0 are 295,351, 56,560 mt, and 168,875 mt, respectively. The SSB trend falls within the confidence interval for most of the period after the onset of survey and compositional data for Alaska and the US West Coast, whereas the trend for British Columbia matches that of the regional assessment although slightly higher in magnitude (Figure 3.24). OM depletion is above the "overfished" cut off at the end of the historical period, but is close to the threshold in Alaska (0.29). Because the OM does not fit the unprecedented high survey biomasses observed in the terminal years for Alaska nor British Columbia, the simulated survey observations for the projection period are more in line with observations for earlier years (Figure D.2). A detailed description of the conditioning results and projected data are provided in *Chapter 3*.

4.4.1 Estimation Method Performance

All EMs fit the historical and projected survey data well and fit the high survey observations in the late 2010s better than the conditioned OM (Figure D.2 vs Figure 3.15), with the peak observation at the end of the historical time series for Alaska best fit by the three-area models. Catches are fit perfectly by all EMs in all years. *SixAreaFourStockMove*, *ThreeAreaThreeStockNoMove* and *Panmixia* have the highest proportion of models with invertible Hessian matrices (96%, 99% and 100%, respectively), followed by *SixAreaThreeStockMove*

(76%) and *ThreeAreaThreeStockMove* (9%, Table D.1). The proportions of simulations with invertible Hessians does not vary greatly among years and were not used to discard model runs, as this is not often used as a criteria for estimators in an MSE context (e.g., Heller-Shipley et al. (2021); Punt et al. (2016)). Results shown here correspond only to simulations that successfully minimized (but did not necessarily produce an invertible Hessian matrix) for all projection years.

Strategies that treat Alaska as a single area (*ThreeAreaThreeStockMove* and *ThreeAreaThreeStockNoMove*) estimate the population to be overfished in 14% (*ThreeAreaThreeStockNoMove*) to 26% (*ThreeAreaThreeStockMove*) of simulations during the first five years of the projection period (Figure D.3); the median estimated depletion during this period ranges from 28% to 33%. The stock was estimated to be overfished in fewer than 2% of simulations for Alaska thereafter. The British Columbia population is not estimated to be overfished during any projection year for any simulation under the spatially-structured estimation methods (Figure D.3). *SixAreaThreeStockMove* and *ThreeAreaThreeStockMove* estimate the population off the US West Coast to be overfished at least once in 63% of simulations (Figure D.4), and *ThreeAreaThreeStockNoMove* estimates the population to be overfished at least once in 92% of simulations; the median estimated depletion for the entire projection period for the US West Coast ranges from 21% (*ThreeAreaThreeStockNoMove*) to 33% (*SixAreaFourStockMove*, Figure D.3).

Estimated unfished biomass does not vary greatly among years or management strategies (Figure 4.4; Figure D.4). Median estimated B_0 for Alaska ranges from 299 to 305 kt (Figure 4.4; Figure D.4); the OM value is 255 kt. For British Columbia, B_0 estimates do not vary by more than 100 t across estimators (Figure 4.4; Figure D.4); the OM value is 144 kt. Median estimated B_0 for the US West Coast ranges from 171 to 192 kt (Figure 4.4; Figure D.4); the OM value is 167 kt. In comparison to the operating model, the EM estimates of unfished biomass are slightly higher for Alaska and the US West Coast, and lower for British Columbia (where the median estimate of B_0 across all years and strategies is 61 kt, roughly half the value in the OM and more similar to the estimate from the recent, regional assessment, Figure D.4 and Figure 3.24). The median panmictic model estimate of B_0 is 575 kt (Figure D.4); the sum of the OM values is 535 kt.

Estimation uncertainty, represented by the variance of the estimated biomass across all replicates and years during the projection period, is lowest for *ThreeAreaThreeStockNoMove* (Figure 4.4).

Variance in estimated biomass is largest for Alaska for *ThreeAreaThreeStockMove*, and for British Columbia and the US West Coast for *SixAreaFourStockMove* (Figure 4.4). The panmictic model exhibits very little uncertainty on a year-to-year basis (Figure 4.4).

The discrepancy between estimates of stock biomass and depletion and their true values is measured by the mean relative error during the projection period (Figure 4.4). The MRE for both measures is the most positive for all regions and estimation methods in the first projection year, after which MRE in stock biomass and depletion declines and stabilizes (Figure 4.4). The MRE trajectories for biomass and depletion in Alaska are virtually identical for *SixAreaFourStockMove* and *SixAreaThreeStockMove*; the MREs for depletion in Alaska are negative for all EMs. The MRE trajectories for biomass in British Columbia are all negative, while the MRE trajectories for depletion in British Columbia are positive after 2021. The least-biased depletion trajectory for British Columbia occurs for *SixAreaThreeStockMove*, where the median MRE across all projection years is 5%, although the biomass trajectory is the most biased with a median relative error of -44%. *ThreeAreaThreeStockMove* results in biased biomass trajectories for British Columbia (-44% median relative error across all years), but minimally biased depletion (4%, Figure 4.4). *SixAreaFourStockMove* is the least biased for the US West Coast in terms of stock biomass (average MRE across all projection years = 31%, Figure 4.4) and depletion (average MRE across all projection years = 28%); all other spatially-structured EMs resulted in negative median relative errors for the entire projection period for the US West Coast (Figure 4.4). *ThreeAreaThreeStockNoMove* was the most biased by both measures for the US West Coast (median MRE across all projection years = -52% for stock biomass, -28% for depletion). The panmictic estimation model led to the largest discrepancy of all estimators, with a median MRE for stock biomass of 71% and depletion of -73%.

All spatially-structured management strategies estimate at least one large recruitment event in all areas during the late 2010s followed by a period of low recruitment leading into the start of the projection period, and relatively stable recruitment deviations thereafter (Figure 4.5). The large late-2010 recruitment pulse occurs in 2017 for the US West Coast and southern British Columbia, in 2016 and 2019 for northern British Columbia/eastern Gulf of Alaska, and 2016 and 2019 for the western Gulf of Alaska/Eastern Bering Sea and Aleutian Islands by *SixAreaFourStockMove* (Figure 4.5). *SixAreaThreeStockMove*, *ThreeAreaThreeStockMove*, and

ThreeAreaThreeStockNoMove all place this event in 2016 for the US West Coast and British Columbia, and 2015 for Alaska (with a secondary pulse in 2018 or 2019, Figure 4.5). In the OM, the largest recruitment event occurs during 2016 for the southerly stocks, and during 2019 for northern British Columbia/eastern Gulf of Alaska. Estimated recruitment is more stable during the projection period than during the historical period for all EMs. The recruitment deviations for *SixAreaFourStockMove* are negative for the entire projection period for the southern California Current and near zero for the northern California Current/southern British Columbia. Estimates of the recruitment deviations from *Panmixia* are nearly flat and slightly negative for the entire projection period, with no large recruitment event in the late 2010s and very little variability among simulations (Figure 4.5).

4.4.2 Management Strategy Performance

Management strategy performance is contextualized by broad category (economic or conservation) and by the absolute and relative performance of a given strategy, within either the management region (for economic metrics) or biological stock (for conservation metrics). There is no single management strategy that out-performs others for all metrics across the northeast Pacific. Instead, there is greater variation among strategies in biological performance (specifically relative depletion $[B/B_0]$) for more southerly stocks (Figure 4.2), where total biological scores range by up to 68% (Table 4.3), than for northerly stocks, where total scores range by only 14% (Table 4.3). All strategies result in median depletion greater than the target level (40%) in at least 50% of simulations for the stocks spanning northern British Columbia and Alaska (Table 4.3). Mean length during the first 10 projection years varies by two centimeters at most within stocks among management strategies (Figure D.5). The southern California Current is overfished for 90% of *SixAreaFourStockMove* simulations (all replicates, all years) (Table 4.3; Figure 4.6). The populations in the stocks encompassing the US West Coast and British Columbia are never above $B_{40\%}$ under *SixAreaFourStockMove*, *eHCR* nor *meanCatch* (Table 4.3; Figure 4.6).

Variation in economic performance is also more pronounced across management strategies in the US West Coast than in Alaska or British Columbia. Spatially-structured management strategies all result in positive catch trends for Alaska (Table 4.4; Figure D.5). All strategies result in catches above the historical minima in at least 50% of projection years (Table 4.4), although the panmictic strategy has periods of very low catch early during the projection period (Figure D.5).

SixAreaFourStockMove and *SixAreaThreeStockMove* have nearly equivalent overall scores for catch metrics in Alaska (5.49 vs 5.41, Table 4.4). The best-performing strategy for British Columbia in terms of catch (*eHCR*) has low AAV during the projection period (4%, Table 4.4) and AAV falls below the 15% threshold in all projection years (Table 4.4). The empirical HCR dominates the mean catch strategy for British Columbia because the catch trend is positive under *eHCR*. The empirical HCR outperforms all other strategies for the US West Coast on every economic metric total catch during the projection period, which was highest for the US West Coast under *SixAreaFourStockMove* at 130 kt, nearly a third larger than catches under *eHCR* (110 kt, Table 4.4). Total catches for British Columbia under *ThreeAreaThreeStockNoMove* were 38 kt, roughly at the middle of the range among strategies (29 kt under *SixAreaThreeStockMove* to 79 kt under *Panmixia*). In all regions, strategies that score highest for economic metrics are distinguished by high catches during the first five projection years and overall, low annual average variability in catch, and many years with average annual variation in catch below the 15% threshold identified by stakeholders (Table 4.4; Figure D.5). This suggests that strategies that perform well in this category do so by meeting more than one dimension of economic performance (stability through time and volume in the near-term).

Finally, the model that approximates the current management paradigm (*ThreeAreaThreeStockNoMove*) is neither the preferred nor the worst-performing strategy for any region, with larger absolute tradeoffs in catch and biological performance for southern regions. This strategy is among the top-three performers in terms of biological metrics for all stocks, resulting in biomass levels above $B_{40\%}$ in at least three-fourths of simulations for all stocks (Table 4.3). Catches under *ThreeAreaThreeStockNoMove* were in the middle of the range for Alaska (290 kt overall, about 85% of the maximum for that region) with similar AAV to *SixAreaFourStockMove* (11%, Table 4.4 and Figure D.5). This strategy results in the lowest or second-lowest catches for the US West Coast in the first and last five years of the projection period (6-11 kt, Table 4.4) and overall (37 kt, Table 4.4; Figure D.5).

4.5 Discussion

4.5.1 Summary

In relative and absolute terms, the differences among management strategies are far less pronounced for Alaska and the Alaskan stocks than for the US West Coast and stocks therein

(Tables 4.3 and 4.4). This is likely a result of the interaction between movement paradigms, the specification of stock structure and how the management-region-specific HCRs respond to changes in stock structure and size (discussed below). Findings suggest that Alaska may obtain better catch outcomes under management strategies that consider connectivity between areas than those that do not, or those that do not use an estimation method, though the difference in near-term catch across strategies is less than 10%. The US West Coast has the largest range in catches and resultant depletion across strategies (Figures 4.6 and D.6). Crucially, southerly stocks encompassing this management region are the only stocks for which biomass was never above 40% of unfished biomass (with median depletion as low as 21% for the southern California Current under the four-stock strategy, Figure 4.6; Table 4.3), an important threshold for fishery management in that region. The US West Coast must carefully consider the interaction between movement rates, stock structure, and the allocation of catches south of 36°N to avoid localized depletion. Future research should explore the drivers of recruitment dynamics range-wide and investigate the integration of tag-recapture data to better estimate movement rates both within and across management areas.

No single management strategy consistently out-performed the others for all stocks, management regions, and performance metrics. Each strategy presents trade-offs with respect to the risk of localized depletion, and the stability and volume of catches through time. The empirical HCR and mean-catch strategy perform well on economic measures for British Columbia and the US West Coast largely because of the stability inherent in these methods. Total catches are equal under these strategies for the US West Coast (110 kt), and AAV in catch is below the 15% threshold in 100% of simulations. Empirical HCRs can act as a "stand-in" for cases where the status-quo management paradigm is a failure; while that is not the case for sablefish, it is useful to know the potential for this approach given that the straightforward link between abundance and management advice is often attractive to stakeholders (Walter et al., 2023). The strategy with four demographic stocks and six linked spatial areas leads to much higher catches than the other strategies for the US West Coast early and late during the projection period (~40 kt vs ~30 kt, Table 4.4) yet meets the AAV cutoff in fewer than half of the simulations.

The panmictic strategy performs best for biological measures for all stocks due to periods of low catches (Figure D.5). The panmictic strategy assesses stock biomass as a single unit, and the

resultant TACs are calculated assuming that the entire population is available for harvest by each fishery. This means that even if the stock is estimated to be very depleted (as it is during the early part of the projection period, Figures D.3 and D.6), driving the population even lower for several years before catches are reduced and the population rebounds dramatically (Figure 4.6). Catches under the panmictic strategy are especially high in British Columbia from 2025 onwards, but this does not result in a strong decline in biomass (Figure 4.6), since the catches taken under this strategy are the lowest (Alaska) or among the lowest (US West Coast) in the regions that otherwise dominate the catch totals for these stocks (Figure D.5). The unstable nature of *Panmixia* is reflected in the near-zero catches during the first five projection years (Table 4.4) and low total catch (e.g., ~30% of total catches for other strategies for Alaska and the US West Coast, Table 4.4 and Figure D.6).

4.5.2 What drives differences in management performance?

The results indicate tradeoffs between economic and conservation performance, with more pronounced discrepancies among management strategies in the southern regions. Additionally, all EM-based management strategies involve a mismatch between the scale of the conservation unit (stocks) and the management paradigm (three political areas); such incoherence has been shown to lead to undesirable outcomes, particularly when demographic parameters vary among modeled areas (Berger et al., 2020). The sensitivity of the outcomes for the southerly regions to the spatial misspecification in management strategies corroborates warnings from simulation studies that such mismatches can mask localized depletion of small stock units (Bosley et al., 2019; Okamoto et al., 2020)

Economic and biological outcomes do not vary greatly among strategies in Alaska (aside from the *Panmixia* strategy), with total catches ranging from 26 kt to 34 kt (Table 4.4; Figure D.5), and median depletion in the two northern stocks from 45-58% (Figure 4.3). This effect is compounded by the fact that estimated biomass in Alaska is similar across management strategies (Figure 4.4), although estimation uncertainty is higher in the spatially-structured models, as expected (Punt, 2019a). The consistency in estimated stock biomass is attributable to high mixing rates within the Alaskan spatial areas, while emigration to British Columbia amounts to 4% or less of Alaskan biomass (Figure 4.2e). This means that the movement paradigms examined here do not lead to large differences in the estimated stock biomass for Alaska among strategies. Future research

should investigate movement rates, potentially using tag-integrated assessment models, to confirm whether emigration across the Alaska-Canada border is indeed low and assess how estimation of these rates might impact management performance; such efforts are underway in Alaska (D. Goethel, Alaska Fisheries Science Center, *pers. comm.*).

Conservation and economic objectives in British Columbia were met by several management strategies, with several strategies leading to positive outcomes for both performance categories for this region (Tables 4.3 and 4.4). Two mechanisms likely contribute to the apparent insensitivity of British Columbia populations to the spatial structure of the management strategy used: firstly, the target harvest rate used in British Columbia is set to 5.5%, lower than what the MSY-based target harvest rate would be for this region given the demographic values used in this study (18%, Table D.4). This results in a narrow range of catches under the spatially-structured strategies (Figure D.6), all below the EM-free strategies, leading to less-depleted populations (Figure 4.6). Secondly, OM movement rates assume British Columbia to be highly dependent on biomass subsidies from other management areas, whereby ~26% of Canadian biomass is a result of immigration on an annual basis (Figure 4.2e, f). Simulation work has indicated that recruitment estimates for smaller populations (such as that off the coast of British Columbia) are sensitive to mixing with larger populations (Cadrin et al., 2019).

Estimated recruitment deviations for British Columbia were negative for most of the projection period under strategies where British Columbia is treated as an independent stock; when assumed to belong to two shared stocks, estimated recruitment deviations applicable to British Columbia are positive for the entirety of the projection period (Figure 4.5). This is consistent with the observation that under the movement and recruitment paradigm specified in the operating model, the British Columbia population exports recruits and/or biomass to adjacent areas, as evidenced by the posterior estimates for τ (Figure D.7). Revisions to the movement-density dependence paradigm, particularly were they to suggest that British Columbia is less connected to other stocks, might lead to less-desirable depletion outcomes (as in the southern part of the US, discussed below). Managers in British Columbia could consider the feasibility of moving towards a spatially-structured estimation method, considering that it appears that outcomes are fairly consistent regardless of the movement paradigm used (given the range of movement rates explored here).

The US West Coast and stocks therein exhibit much more variation in economic and biological performance among management strategies. This is likely the result of the movement-recruitment paradigm used within the estimation methods, and how the harvest control rule applied in this region is sensitive to spatial misspecification in the management model. For the US West Coast and British Columbia, assuming four demographic stocks leads to higher total catches (130 kt vs. 110 kt or less for the other strategies, Table 4.4; Figure D.5) at the expense of much lower depletion in the southern California Current (24% median across all projection years for all simulations, Figure 4.6). Target harvest rates would have been lower overall had the US West Coast instead implemented the Alaskan SPR-based HCR (11% vs 14% under *SixAreaFourStockMove*) and would have changed less dramatically among strategies (Table D.4), potentially avoiding the 'overfished' outcomes for the US West Coast and British Columbia under this strategy. However, recent examinations of Atlantic groundfish indicate that unintended overfishing driven by mis-perceptions of stock status might be more influential to management strategy performance than the control rule used (Mazur et al., 2023).

The undesirable biological outcomes for stocks off the US West Coast illustrate several important considerations for the impact of spatial dynamics on management outcomes. Assuming four demographic stocks in the estimation method results in large positive bias for stock biomass in 2020 (Figure 4.4), corresponding to the high estimate of the 2016/2017 recruitment event also captured in the regional assessment (Figure 4.5, Kapur et al., 2021b). The estimated biomass quickly declines, as does the MRE in depletion, to near-zero, where it remains for the first half of the projection period (Figure 4.4). Around 2030 the MRE in biomass becomes more positive for this strategy, even though estimated biomass has decreased slightly (Figure 4.4), suggesting that the EM has failed to accurately perceive a decline in biomass in this region, likely due to the absence of age-composition data since the "missing fish" are not selected by the survey (at 8 years) given the projection period used here.

The OM and first management strategy has six connected sub-areas nested within four independent "stocks" (Figure 4.2). Two of these stocks are each comprised of a single sub-area, one of which (the area south of 36°N on the US West Coast, "C1") receives no immigration from northerly regions (Figure 4.2e). The survey estimates of biomass for C1 are of similar magnitude to those for British Columbia, about 25% of the total for the US West Coast region, and do not exhibit

similar trends to the northerly population (remaining mostly flat as northern stocks increase and decrease, Figure D.2). These ecological characteristics render the area south of 36°N a strong candidate for localized depletion due to its small size and unidirectional movement northward (Figure 4.2e, McGarvey et al., 2017; Ying et al., 2011).

Empirical analyses have suggested that the area south of Monterey Bay (36.6°N) exhibits different sablefish recruitment dynamics from the northern area (Tolimieri et al., 2018), which is represented by the near-zero correlation in recruitment deviations between these two areas (Figure 4.3). In the OM, 80% of estimated recruitment and 90% of mature biomass in the US West Coast comes from the area north of 36°N, consistent with the posterior estimates of τ (Figure D.7) and relative survey biomasses (Figure D.2) for this region. Indeed, the biomass decline in the OM is pronounced in this southern area (Figure 4.6). A slight decline in estimated biomass is visible for the US West Coast during the last three years of the projection period when the southern US is treated as an independent stock (Figure D.3). Given that the HCR acts at the management level (even though assessments are conducted at the stock level), managers must consider the risks of missing undesirable population trends at small spatial scales inherent in using simplified estimation methods (aka. Okamoto et al., 2020). Future work could explore alternative TAC-setting and/or apportionment procedures that consider local dynamics, given that this analysis provides information on survey biomass in each area that could be considered in catch allocation.

4.5.3 Limitations and future work

Results should be considered given several limitations of the study design. Firstly, the MSE assumes that TACs are fully realized by all fisheries in all management regions, whereas it has been common, particularly off the US West Coast, for TAC attainment to be low (on the order of 15% south of 36°N, often due to bycatch avoidance; Somers et al., 2022). This means estimated depletion may be lower than would be expected given the same TAC specification, so it is possible that the fishery dynamics in that region would preclude undesirable depletion outcomes such as those here. It is likely that management strategies based on age-structured models might be required to avoid bias in derived quantities. Coupled with the negative bias in estimated biomass discussed below, it is likely that most results represent a conservative perception of stock size and status.

With one exception (the first management strategy for the US West Coast, discussed above), the estimation methods used by estimation-based management strategies were negatively biased in terms of estimated biomass and relative depletion for the projection period (Figure 4.4). This result is also apparent in the MSE for British Columbia (S.D.N. Johnson, Landmark Fisheries, *pers. comm.*), where the surplus production model used as an estimation method results in lower biomass estimates than the age-structured operating model. The estimation methods on which estimation-based management strategies are based do not use age-composition data. Removing age-composition data from the OM results in lower estimates of stock biomass for Alaska and the US West Coast for some or all of the period from 2000-2020 (Figure C.13), so the absence of that data source from the model-based management strategies is a possible driver of the discrepancy between estimated and true biomass. Regardless, these negatively biased errors act as a precautionary buffer, whereby TACs are automatically lower than they would be had the estimates been closer to unbiased.

The performance metrics were weighted equally. Most of the performance metrics selected by stakeholders were related to catch; only three out of ten performance metrics explicitly concerned the biological aspects of the populations (the probability that the stock is below 40% of unfished biomass, mean length, and the probability that the stock is overfished). It was impossible for the mean catch strategy (mean catch from 2000-2019 throughout) to return a positive catch trend during the projection period; had catch performance been measured without this metric, mean catch would be preferred in each management region, followed by the empirical HCR. These EM-free strategies easily outperform the others in terms of catch stability, provide mid-range catch and depletion outcomes, and present no risk of overfishing (as is largely the case with most other strategies). The choice to adjust TACs under the empirical HCR using a five-year survey window and the last 20 years of the historical period meant that catches were conservative in Alaska and the US West Coast under the EM-free strategies (since the sablefish stock was at low levels through the early 2000s and 2010s in all regions in the OM, and catches were accordingly reduced by management, Figure D.6). Allowing the empirical HCR to be a function of recent catches would have led to higher AAV for the empirical HCR, and possibly greater contrast with mean catch. Future work could readily incorporate different or additional EM-free approaches, performance metrics and weightings thereof.

Quantities output during the MSE that were not formally considered in the scoring of strategies reveal additional tradeoffs related to the feasibility of implementing the management strategies. The percentage of applications of the EMs with invertible Hessian matrices is low for the strategy with three linked stocks (9% on average across replicates and years), a strategy that otherwise performed well for Alaska and British Columbia. This management strategy attempts to estimate population dynamics at a coarser spatial scale than the OM from which data were generated and use a distinct movement paradigm that suggests an enormous influx of fish from Alaska to British Columbia (Figure 4.2). Survey data were also aggregated into a single survey for each management region. This poses a particular challenge for Alaska and the US West Coast, because the two sub-areas that constitute those management regions in the OM have either conflicting recruitment signals (Alaska) or are assumed to be uncorrelated (US West Coast, Figure 3.12), which may explain the low proportion of simulations with invertible Hessian matrices for this strategy. The strategy most similar to the OM had invertible Hessian matrices in 95% of individual estimation method runs. While the invertibility of the Hessian matrix was not used as an evaluation criterion, it is an important consideration for the feasibility of estimation frameworks as applied management strategies, particularly since it appears that the (status quo) approach that includes movement among the three management areas is incapable of returning invertible Hessian matrices using only survey and catch data.

There are several management-region-specific processes that were not included in the simulation framework to ease comparison among management regions. British Columbia uses a two-stage harvest control rule, whereby TACs calculated from the HCR are only implemented if they exceed the TAC for the previous year by greater than 200 t, and a catch floor of 1,992 t is treated as a performance objective (not as a lower limit). No management strategies except *Panmixia* resulted in catches below the floor (14% of simulations), and 5% of simulations where management areas were unlinked by movement and 10% of simulations under *Panmixia* involved year-over-year TAC increases for British Columbia less than 200 t. Including these stabilizing buffers would result in a reduction of the overall catch taken under each management strategy for British Columbia, improving the performance metrics related to depletion and the AAV in catch for all management strategies.

The definition of "overfished" used to calculate performance metrics is based on a threshold of 25% of unfished biomass; it is undesirable to conflate management performance with the complex, and unresolved, problem of calculating equilibrium reference points (e.g., B_{MSY}) from a spatially-structured operating model (Kapur et al., 2021c). Similarly, the application of management region-specific thresholds would have over-complicated the interpretation of management performance. Future explorations of this system could choose to focus on economic outcomes, such as the percentage of years the fishery would be closed, using region-specific thresholds.

The OM and EMs differ in terms of the underlying population dynamics model (age-structured vs. delay-difference model), as well as the types of data included (the EMs do not consider compositional data nor discard rates). The delay-difference model used in the five spatial EMs assumes that the parameters of the population dynamics can be estimated using catches and regional survey indices alone. The reliance on these data and the inclusion of spatial processes result in more uncertainty in estimated quantities for all regions, which is reduced as the spatial complexity of the EM decreases (Figure 4.4).

The additional data used when conditioning the operating model led to compromises in how well the OM fitted to individual data sources: it fails to fit the terminal year survey observations for Alaska and British Columbia (Figure 3.15), which are much larger than those for the preceding years, although it does estimate an increase in observed biomass during the early projection period (Figure D.2). This pre-determines that the estimation methods (which fit all survey observations) anticipate higher biomasses than the OM during the early years of the projection period. This should not be a factor in relative management strategy performance because this effect (better fits to survey data) was consistent across all EMs. The delay-difference approach is computationally faster than a statistical catch-at-age model, and is also able to fit all survey data from all regions even in the presence of complex spatial structure, whereas the operating model could not reconcile the high terminal survey observations in Alaska and British Columbia with compositional and discard data, and was computationally intractable to implement as an estimator. Future work could explore methods to incorporate age-structure within the estimation framework. An important next step for this work would be to synthesize the drivers of recruitment dynamics for the sablefish population coastwide, to 1) confirm or modify the structural assumptions present in the operating model, and 2) allow for exploration of how oceanic conditions might lead to changes in the

recruitment paradigm for sablefish, and effects on management performance. This analysis could inform the development of additional operating models and allow for investigation of the interaction of climate change and spatial uncertainty for high-value species such as sablefish. Future phases of this work could also integrate economic models to investigate performance metrics such as fishery profits, or allocation questions, both of interest to stakeholders (Kapur et al., 2021a).

4.5.3 Conclusions

Introducing spatial structure into the data collection, assessment and management process is an immense undertaking for any fishery. Transboundary stocks present the additional challenge of formal, political barriers to constructing a mathematical representation of fish populations throughout their range. This study presents a management strategy evaluation for a valuable groundfish in the northeast Pacific, and finds that spatial models of intermediate structural complexity, including those that match the current assessment and management paradigm in terms of having three modeled areas, can satisfy stakeholder objectives and avoid negative outcomes for the sablefish fishery. Management strategies that include movement between regions allow for better economic outcomes for the US West Coast and Alaska, with tradeoffs in catch variability. Accounting for mixing with adjacent regions does not result in large differences in management performance for British Columbia, given the range of movement rates explored here. A strategy where the three management regions are treated as separate, unconnected stocks – the closest to the current approach – was not preferred for any region, but results in neither the optimal nor worst outcomes for any performance metric. Specifically, scaled scores for biological performance metrics are at or above 88% for all stocks, with no risk of overfishing for any stock under this strategy. Performance on catch-related performance metrics was intermediate (in 3rd or 4th place out of the seven strategies). The mismatch in the spatial scale of the harvest control rules and the conservation units hypothesized here poses a crucial risk of ‘cryptic’ or localized depletion in isolated stocks in the southern U.S. West Coast.

The motivation for this work, and others like it, lies in the observation that population demography varies greatly throughout a stock’s range (Berger et al., 2020), and it is the specification of demography (including the structure of stock units and movement among them) that most heavily influences the relative performance of each management strategy. Though this work implements

many of the “best practices” in spatial modeling for design and specification of the operating models and estimation methods, it is clear the mixing rates and the location and structure of density-dependence (recruitment) remain crucial uncertainties for evaluating the population status of long-lived, highly mobile species such as sablefish. These findings underscore the influence of movement rates on management performance (Goethel et al., 2011), and suggests that allowing for simultaneous estimation of movement (via the construction of tag-integrated models, perhaps on the scale of management regions) would be worthwhile for confirming the magnitude of movement effects, particularly for Alaska and British Columbia.

4.6 Tables

Table 4.1: Components of the management strategies. Each estimation method (rows, top half of table) is combined with the "status quo" control rule (HCR) for a total of seven management strategies.

Spatial Scenarios				
Estimation method	Spatial areas	Demographic areas (stocks)	Movement among areas	
SixAreaFourStockMove	6	4	Yes	
SixAreaThreeStockMove	6	3	Yes	
ThreeAreaThreeStockMove	3	3	Yes	
ThreeAreaThreeStockNoMove	3	3	No	
Panmixia	1	1	n/a	
Harvest Control Rules				
	Alaska	British Columbia	U.S. West Coast	note
Status Quo	NPFMC Harvest Control Rule (Cleaver et al., 2019)	U_{MSY}-based rule, without buffers (Cox et al., 2011)	PFMC "40-10" rule *(Terms of Reference for the Groundfish Stock Assessment Review Process for 2023-2024, 2022)*	See Figure 4.3 and Equations 4.1-4.5
Empirical HCR or eHCR	Mean slope of the logarithms of the most recent five years of survey observations within the management region, multiplied by mean fishery catches during 2009-2019			No estimation method required
meanCatch	Mean fishery catch by fisheries in the management region during 2000-2019			No estimation method required

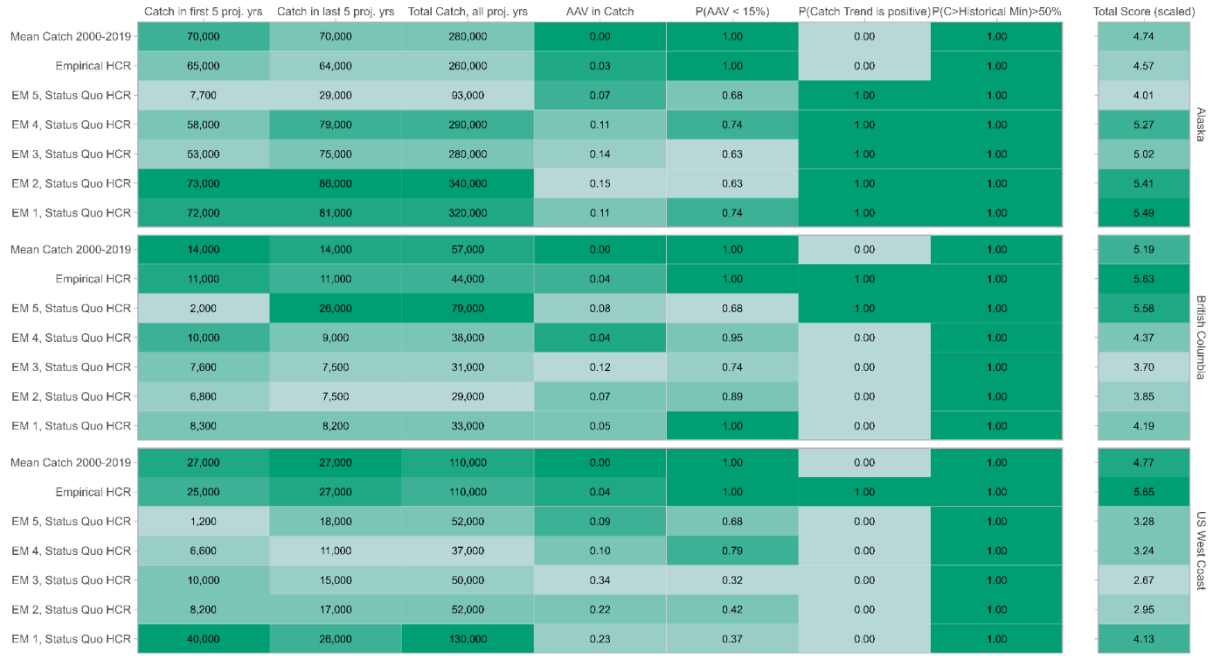
Table 4.2: Objectives and performance metrics identified during the 2021 stakeholder workshop (Kapur et al., 2021a). Equations are provided for how the metric was calculated from the OM during the projection period. When applicable, metrics are first computed for an individual replicate and then averaged across replicates. *These calculations begin in 2021 as actual catches are used for 2020 and do not vary among strategies. Years with zero catch are excluded from this calculation.

Objective	**Performance Metric(s)**	**Equation**
Minimize risk of stock being overfished	Proportion of years during the projection period that the stock biomass is above 25% of unfished biomass	$P(B^m_{2020-2040} > 0.25B^m_0)$
Avoid depleted populations	Proportion of years during the projection period that the stock biomass is above 40% of unfished biomass	$P(B^m_{2020-2040} > 0.40B^m_0)$
Maintain minimum catch level	Whether the proportion of projection years in which the catch is greater than the lowest historical catch is at least 0.5	$P(C^m_{2020-2040} > min(C^m_{1960-2019})) > 0.5$
Minimize annual catch variability	Average annual variation (AAV) in catch over the projection period*	$AAV^m_y = \frac{\sum_{2021}^{2040}\left\|C^m_y - C^m_{y-1}\right\|}{\sum_{2021}^{2040} C^m_{y-1}}$
	Proportion of years* during projection period that the average annual proportional change in catch does not exceed 0.15	$P(AAV^m_{2021-2040} \leq 0.15)$
Maximize catch in the near- and long-term	Average catch during the first* and last 5 years of the projection period	$\frac{\sum_y^{y+4} C^m_y}{5}$
	Total catch during the projection period*	$\sum_{y=2021}^{2040} C^m_y$
	Positive catch trend over the projection period* in each management region	$P(S^m_{2021-2040}) > 0$, where S^m is the slope of a linear model fit to the catches in region m during the projection period
Maximize long term profitability	Mean fish length over a ten-year period, by management region	$\frac{\sum_{2020}^{2029} N^m_y \bar{l}^m_y}{\sum_{2020}^{2029} N^m_y}$

Table 4.3: Median (across replicates) raw scores for the biological performance metrics (columns) for the seven management strategies (rows) and four stocks (panels). Scores have been rounded to two significant digits. The rightmost column indicates the total of the scaled median scores. Color intensity corresponds to the ranking of each management strategy within each management region, where darker green is the best performing, and lighter green is the worst performing.

	P(B>0.40B0)	P(B>0.25B0)	Mean Length, first 10 proj. yrs	Total Score (Scaled)	
Mean Catch 2000-2019	0.00	0.86	33.96	1.71	Stock 1 (South US West Coast)
Empirical HCR	0.00	0.90	34.02	1.76	
EM 5, Status Quo HCR	0.76	1.00	35.31	2.89	
EM 4, Status Quo HCR	0.76	1.00	35.13	2.88	
EM 3, Status Quo HCR	0.67	1.00	34.85	2.75	
EM 2, Status Quo HCR	0.76	1.00	35.00	2.88	
EM 1, Status Quo HCR	0.00	0.10	32.86	0.92	
Mean Catch 2000-2019	0.00	1.00	42.45	1.89	Stock 2 (North US West Coast & South BC)
Empirical HCR	0.00	1.00	42.56	1.89	
EM 5, Status Quo HCR	0.86	1.00	43.89	2.92	
EM 4, Status Quo HCR	0.81	1.00	43.64	2.86	
EM 3, Status Quo HCR	0.76	1.00	43.49	2.80	
EM 2, Status Quo HCR	0.81	1.00	43.59	2.86	
EM 1, Status Quo HCR	0.00	0.69	41.60	1.56	
Mean Catch 2000-2019	0.67	1.00	49.37	2.62	Stock 3 (North BC & East GOA)
Empirical HCR	0.81	1.00	49.49	2.78	
EM 5, Status Quo HCR	0.90	1.00	50.30	2.90	
EM 4, Status Quo HCR	0.81	1.00	49.51	2.78	
EM 3, Status Quo HCR	0.86	1.00	49.65	2.83	
EM 2, Status Quo HCR	0.62	1.00	49.38	2.56	
EM 1, Status Quo HCR	0.55	1.00	49.30	2.48	
Mean Catch 2000-2019	0.76	0.90	56.44	2.82	Stock 4 (West GOA & AI)
Empirical HCR	0.81	0.90	56.57	2.87	
EM 5, Status Quo HCR	0.86	0.90	57.58	2.95	
EM 4, Status Quo HCR	0.81	0.90	56.64	2.87	
EM 3, Status Quo HCR	0.81	0.90	56.80	2.88	
EM 2, Status Quo HCR	0.57	0.90	56.41	2.59	
EM 1, Status Quo HCR	0.55	0.90	56.46	2.57	

Table 4.4: Median raw scores (across replicates) for the economic performance metrics (columns) for the seven management strategies (rows) and three management regions (panels). Scores have been rounded to two significant digits. The rightmost column indicates the total of the scaled median scores. Color intensity corresponds to the ranking of each management strategy within each management region, where darker green is the best performing, and lighter green is the worst performing (for example, higher raw values of average annual variability are lighter in color).



	Catch in first 5 proj. yrs	Catch in last 5 proj. yrs	Total Catch, all proj. yrs	AAV in Catch	P(AAV < 15%)	P(Catch Trend is positive)	P(C>Historical Min)>50%	Total Score (scaled)	
Mean Catch 2000-2019	70,000	70,000	280,000	0.00	1.00	0.00	1.00	4.74	Alaska
Empirical HCR	65,000	64,000	260,000	0.03	1.00	0.00	1.00	4.57	
EM 5, Status Quo HCR	7,700	29,000	93,000	0.07	0.68	1.00	1.00	4.01	
EM 4, Status Quo HCR	58,000	79,000	290,000	0.11	0.74	1.00	1.00	5.27	
EM 3, Status Quo HCR	53,000	75,000	280,000	0.14	0.63	1.00	1.00	5.02	
EM 2, Status Quo HCR	73,000	86,000	340,000	0.15	0.63	1.00	1.00	5.41	
EM 1, Status Quo HCR	72,000	81,000	320,000	0.11	0.74	1.00	1.00	5.49	
Mean Catch 2000-2019	14,000	14,000	57,000	0.00	1.00	0.00	1.00	5.19	British Columbia
Empirical HCR	11,000	11,000	44,000	0.04	1.00	1.00	1.00	5.63	
EM 5, Status Quo HCR	2,000	26,000	79,000	0.08	0.68	1.00	1.00	5.58	
EM 4, Status Quo HCR	10,000	9,000	38,000	0.04	0.95	0.00	1.00	4.37	
EM 3, Status Quo HCR	7,600	7,500	31,000	0.12	0.74	0.00	1.00	3.70	
EM 2, Status Quo HCR	6,800	7,500	29,000	0.07	0.89	0.00	1.00	3.85	
EM 1, Status Quo HCR	8,300	8,200	33,000	0.05	1.00	0.00	1.00	4.19	
Mean Catch 2000-2019	27,000	27,000	110,000	0.00	1.00	0.00	1.00	4.77	US West Coast
Empirical HCR	25,000	27,000	110,000	0.04	1.00	1.00	1.00	5.65	
EM 5, Status Quo HCR	1,200	18,000	52,000	0.09	0.68	0.00	1.00	3.28	
EM 4, Status Quo HCR	6,600	11,000	37,000	0.10	0.79	0.00	1.00	3.24	
EM 3, Status Quo HCR	10,000	15,000	50,000	0.34	0.32	0.00	1.00	2.67	
EM 2, Status Quo HCR	8,200	17,000	52,000	0.22	0.42	0.00	1.00	2.95	
EM 1, Status Quo HCR	40,000	26,000	130,000	0.23	0.37	0.00	1.00	4.13	

4.7 Figures

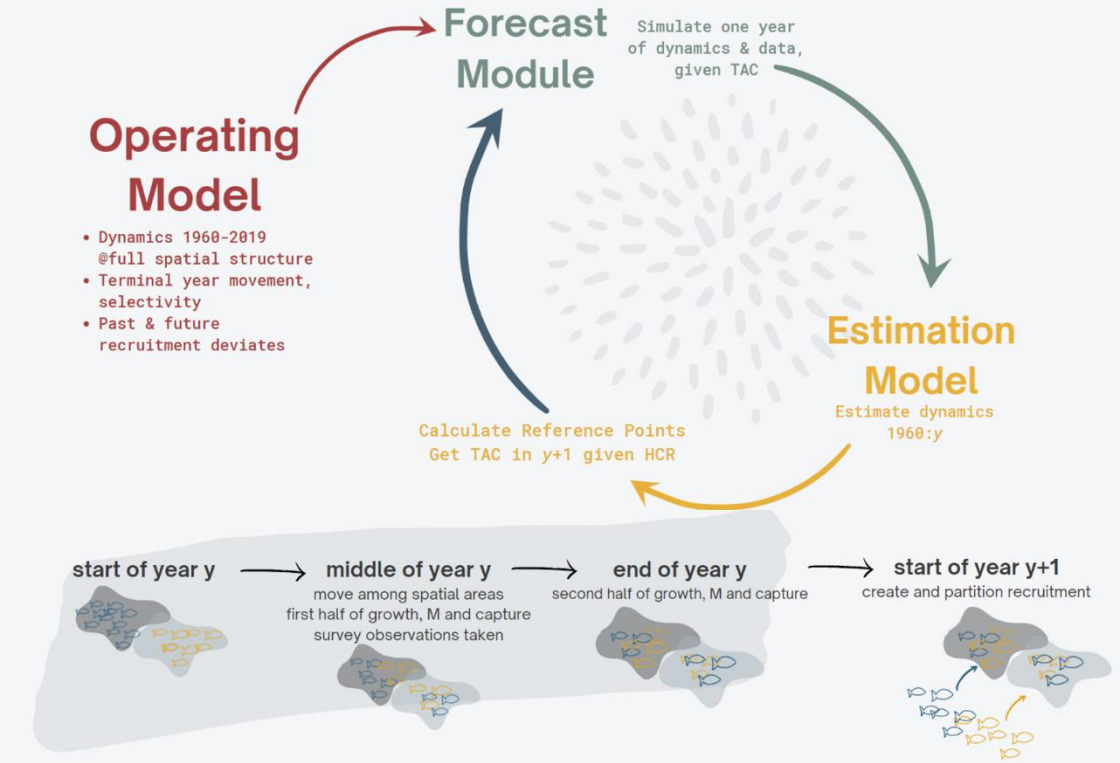


Figure 4.1: (Top) Schematic of the management strategy evaluation framework. The operating model (red text) is conditioned to the observed data from 1960-2019, and is used to determine unfished recruitment, the biomass and age-structure at the start of 2020, and terminal year movement and selectivity. Stock-specific recruitment deviates are drawn for each of 21 projection years for all OM replicates; these determine variability in the simulated dynamics during the projection period (green text). Each annual projection step updates the OM by one year given the fishery-specific catches during year *y* and provides these catches and observed survey biomass data to the management strategy. The management strategy consists of an EM and a harvest control rule, and is used to calculate TACs, shown here in yellow. This cycle is repeated for 30 projection years for each simulation experiment. (Bottom) Schematic of the order of events in the operating model. Fish are subject to movement, half of growth, natural mortality (*M*) and capture by fisheries in the first "season". Survey biomass and associated compositional data are observed at this time. Fish then undergo the second half of growth, movement, and capture. The resultant biomass is used to calculate the expected recruitment for the start of the following year. This recruitment may be split among multiple spatial areas if a given stock is comprised of more than one area.

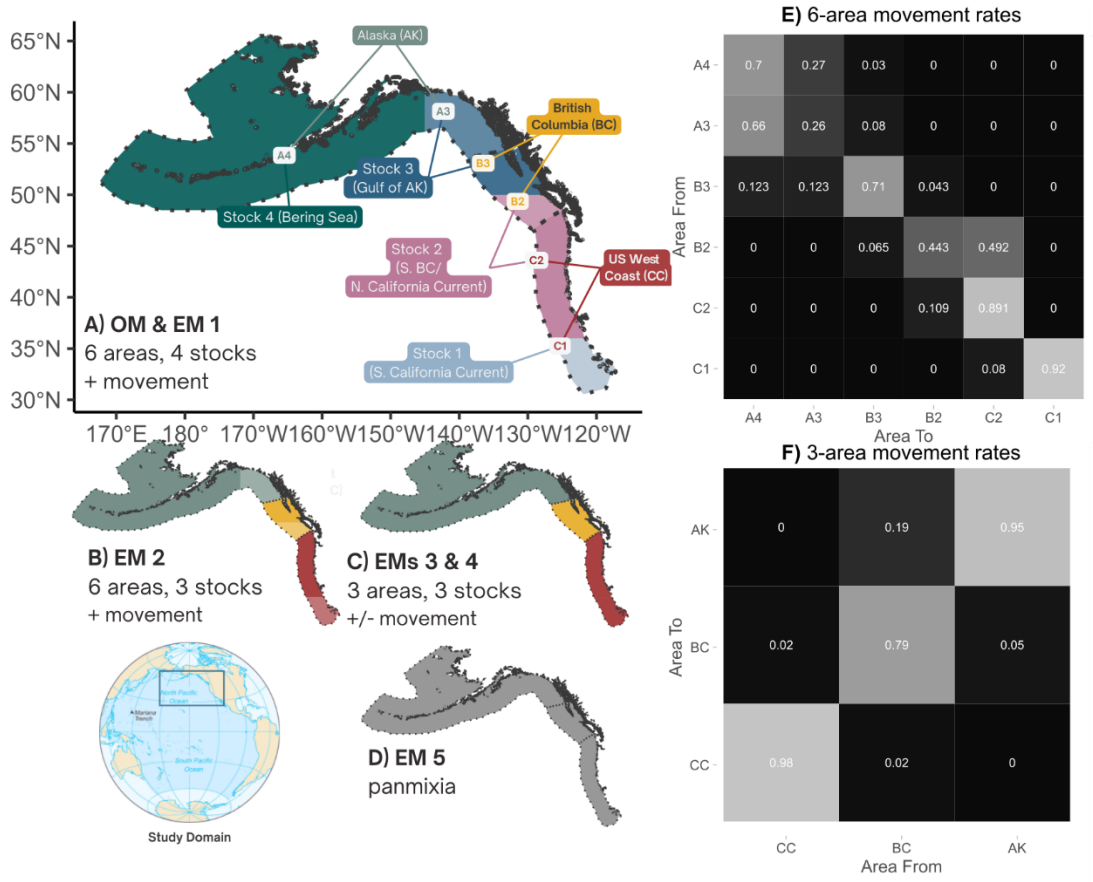


Figure 4.2: A-D) Maps depicting spatial structure in the MSE scenarios. In all scenarios, management actions occur at the three political levels (dotted lines). A) Spatial structure in the operating model and *SixAreaFourStockMove*. There are six modeled spatial areas (individually colored areas) within which dynamics are tracked; fish move among spatial areas. These spatial areas are each nested within a single demographic area or "stock", indicated by shared colors (green, blue, pink and gray). B) Spatial structure for *SixAreaThreeStockMove*. There are six spatial areas with movement among them, and the three stocks in the EM coincide with the management regions. C) Spatial structure for *ThreeAreaThreeStockMove* and *ThreeAreaThreeStockNoMove*. There are three spatial areas, with (*ThreeAreaThreeStockMove*) or without (*ThreeAreaThreeStockNoMove*) movement among them, and the three stocks in the EM coincide with the management regions. D) Spatial structure for *Panmixia*. There is one modeled spatial area and hence no movement. The entire spatial area is assumed to consist of a single stock. E) movement rates for the 6-area estimation methods, and the OM. F) movement rates for *ThreeAreaThreeStockMove*.

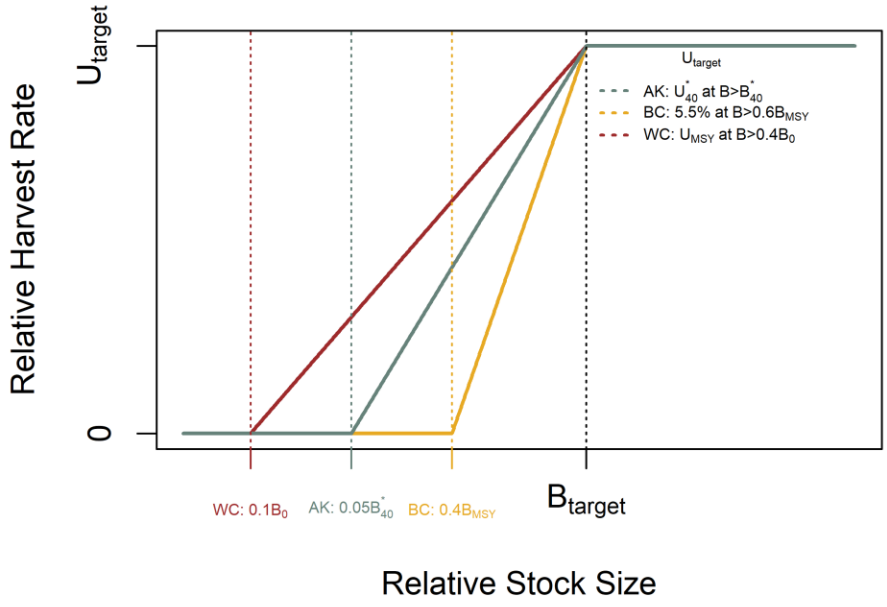


Figure 4.3: Schematic of 'status quo' harvest control rules for the three management regions, which specify the harvest rate *U* as a function of the stock size relative to reference points specific to each region. Vertical lines indicate the start of the ramp, below which *U* is zero. Note that the figure is not to scale and cannot be used to make conclusions about the relative conservatism of the given HCRs in each management region (as the measures on the x-axis differ among regions).

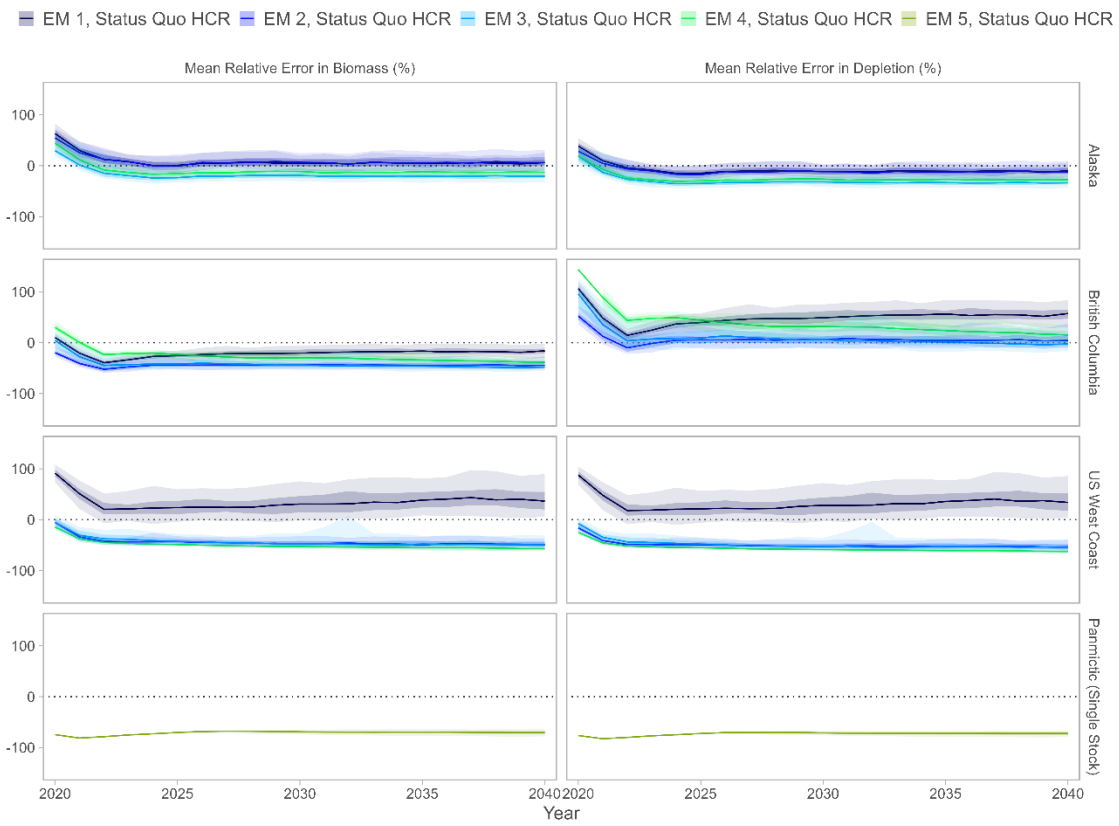


Figure 4.4: Mean Relative Error in total biomass from each of the five spatially-structured estimation methods (colors) by management area (rows) during the projection period (left column) and depletion (B/B_0, right column). The solid line is the median, the darker shaded area represents the 25th to 75th quantiles, and the lighter shaded area represents the 5th and 95th quantiles for 100 replicates. For the panmictic stock (yellow values), errors are calculated against the sum of OM biomass values for all spatial areas. Time-series of estimated quantities indicate the estimates for each year based on the assessment conducted in that year.

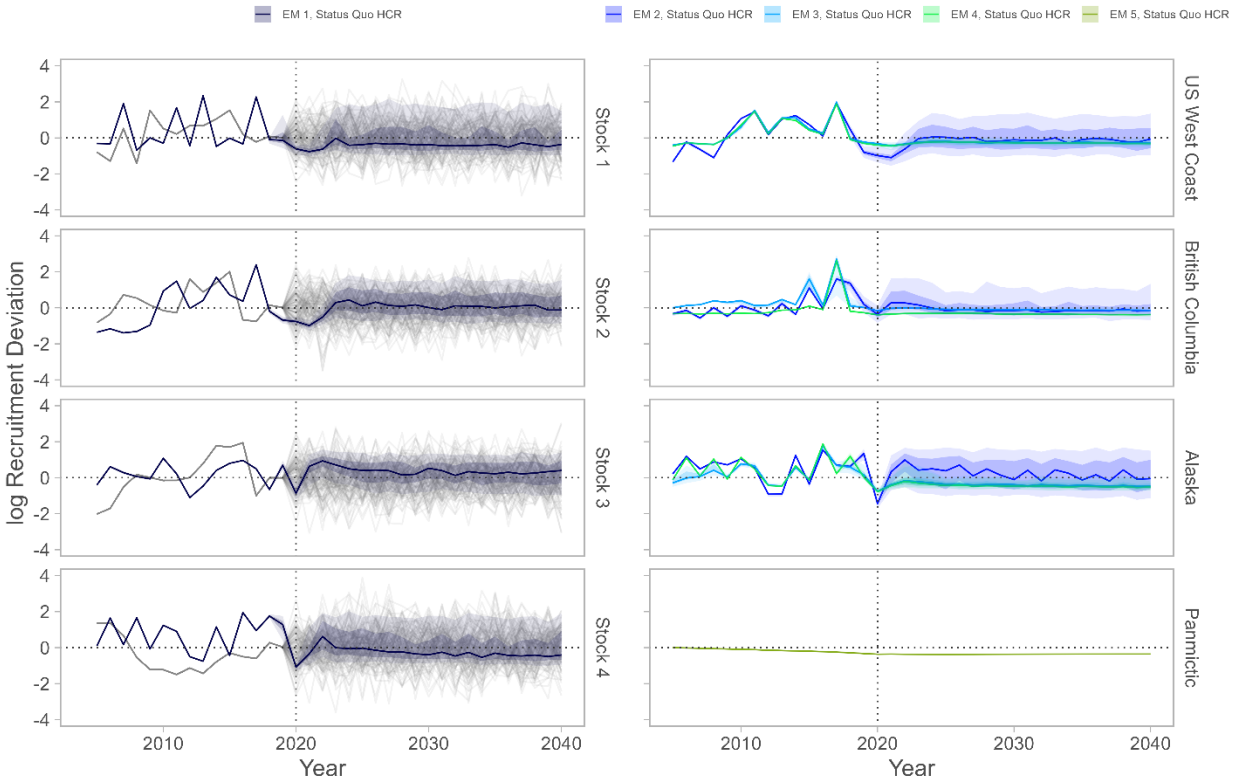


Figure 4.5: Estimated recruitment deviations by stock for the five estimation methods, 2005-2040 based on the assessment conducted in that year; historical values are from the 2020 EM. *SixAreaFourStockMove* (left column) has the same four stocks as the OM; historical OM recruitment deviations for individual replicates are shown as grey lines. All other EMs (right column) assume biological stocks coincide with management areas or that there is single panmictic stock (*Panmixia*). The solid line is the median, the darker shaded area represents the 25th to 75th quantiles, and the lighter shaded area represents the 5th and 95th quantiles for 100 replicates. The vertical dotted line indicates the start of the projection period.

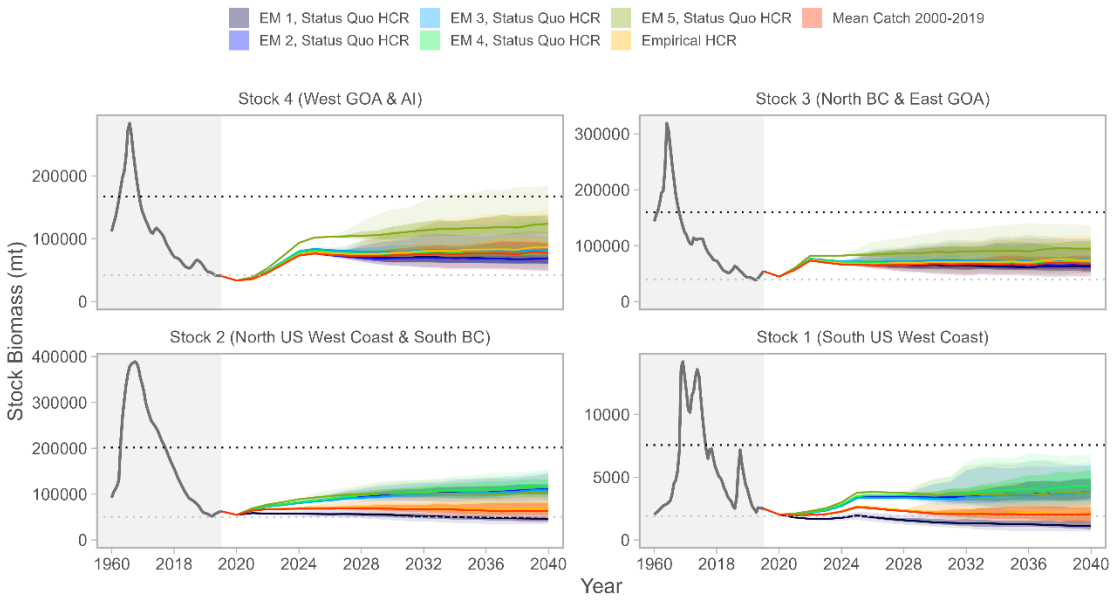


Figure 4.6: Operating model trajectories of stock biomass for four biological stocks (panels) for seven management strategies (colors). The x-axis is condensed during the historical period (1960-2019, grey rectangle). The solid line is the median, the darker shaded area is the 50% simulation interval, and the lighter shaded area is the 90% simulation interval. The horizontal dotted lines indicate the unfished biomass (black) and 25% of unfished biomass (grey).

Chapter 5. CONCLUSION

Introducing spatial structure into the data collection, assessment and management process is an immense undertaking for any fishery. Transboundary stocks present the additional challenge of formal, political barriers to constructing a mathematical representation of fish populations throughout their range. It is believed that transboundary stocks are declining more rapidly than those exploited by a single political entity (Palacios-Abrantes et al., 2020). This dissertation developed a management strategy evaluation to investigate the effects of accounting for or disregarding spatial structure in the management of sablefish in the northeast Pacific, a valuable transboundary stock with suspected connectivity throughout its range.

The first chapter presented a simulation experiment using a generalized fish species to investigate whether correctly accounting for spatial structure provided more benefit in terms of reduced estimation bias than correcting mis-specifications in other population or fishery processes. A "mis-specification" is a mismatch of certain model process(es) or parameter(s) between the assessment model and reality. The consequences of mis-specification that concern analysis are 1) bias, whereby the estimation method produces quantities of management interest that differ systematically from reality, and 2) precision, where the error terms associated with the estimated quantities are too large to be useful for stock status evaluation and/or management decision making. This chapter measured improvement in estimation method performance with sequential elimination of a series of known mis-specifications of observation and systems model processes. Correcting spatial structure led to large improvements in estimation accuracy, particularly in the case of population reproductive output. In addition, synergies between population productivity, natural mortality and size-specific mortality (due to fishing) precluded straightforward gains in estimation accuracy when any of these three processes remain uncorrected. This work encourages use of Integrated Population Models (IPMs), and emphasizes the importance of correctly specifying spatial structure for modeled populations, and how to prioritize the correct estimation of model parameters depending on the estimated quantity of interest.

The second chapter developed and applied a data-based method to detect significant breakpoints in fish size-at-age across space and time, and applied it to sablefish survey observations to determine major spatial regions and time periods among which sablefish growth varies the most.

This approach was validated as more sensitive than a pre-existing detection method (Rodionov, 2004), and has since been used for exploratory purposes in other western groundfish assessment and management applications (such as Pacific lingcod *Ophiodon elongatus*, L. Lam, Northwest Fisheries Science Center, *pers. comm.*). The spatial areas identified for sablefish in this chapter correspond with large oceanographic currents, and corroborate earlier, smaller-scale explorations of variation in the growth of sablefish through space (Coffin and Mueter, 2016; Gertseva et al., 2017). This study provided an empirical justification for the spatial units defined as "stocks" used in the transboundary operating model.

The third chapter presented a novel, transboundary statistical-catch-at-age model conditioned to historical observations of sablefish catch, discards, survey biomass and age compositions from 1960-2019. The model, and sensitivity analyses thereof, illustrated the plausibility of an interconnected transboundary sablefish stock that would result in similar biomass trends to those estimated in recent regional management models. The stock structure defined in the operating model (OM) followed from the analysis in Chapter 2, and the concept of the Bering Sea/Aleutian Islands/Western Gulf of Alaska (area west of 145°W) as a separate marine ecosystem from that of the eastern Gulf of Alaska (Cleaver and Evans, 2019). The US West Coast was modeled consistent with findings by Tolimieri and Haltuch (2018) that suggest that recruitment dynamics in the southern part of the US West Coast (south of 36 degrees N) are distinct from the rest of the coast, and not as strongly linked to stock-wide dynamics as recruitment in the northern regions; the recruitment paradigm in the northern region appears to be strongly linked to sea-level height (Tolimieri and Haltuch, 2023). Finally, the subpopulation off British Columbia was considered a highly connected component of the global population, generally dependent on subsidies from Alaska to meet catch limits and fit survey indices.

The fourth and final chapter presented a management strategy evaluation for the northeast Pacific transboundary sablefish stock, treating the OM from chapter three as the "true" population structure. A suite of estimation methods at various degrees of spatial complexity, coupled with the current harvest control rules from each management region, were tested as alternative management strategies and evaluated based upon their ability to realize performance metrics determined by stakeholders in a series of workshops (Kapur et al., 2021a). The study design of the MSE allowed for pairwise comparisons across several dimensions of spatial model specification (particularly

movement, number of areas modeled, and stock structure). The motivation for this effort, and others like it, often lies in the observation that population demography varies greatly throughout the stock's range (Berger et al., 2020), and it is the specification of demography (including the structure of stock units and movement among them) that most heavily dictates management strategy performance. Though this work implemented many of the "best practices" in spatial modeling for design and specification of the operating and estimation methods identified by Punt, (2023), it is clear the movement rates and the location and structure of density-dependence (recruitment) remain crucial uncertainties to evaluating the population status of long-lived, highly mobile species such as sablefish. These findings underscore the influence of movement rates on management performance (Goethel and Berger, 2016), and suggest that allowing for simultaneous estimation of movement rates (e.g., with the incorporation of tag-recapture data in an age-structured model) would be worthwhile for confirming the magnitude of movement effects, particularly for Alaska and British Columbia.

The software and data products produced for this dissertation form a toolset that can be used to explore additional questions about sablefish or other species with similar questions. While it is promising that a simplified model type can be used to capture the general trends in range wide population size, many estimators exhibited a negative bias compared to the operating model (likely due to the absence of age-composition data). A computationally tractable means to incorporate age-composition information into the MSE framework could resolve some issues related to model bias.

As of this writing, sablefish population dynamics in this region are experiencing a record "boom", with upswings in population size driven by unprecedented large recruitment events during the late 2010s in all regions. These events were captured by the simplified estimation methods used in this dissertation, and act as a protective cushion against overfished stock status for most scenarios. An important next step for this work would be to synthesize the drivers of recruitment dynamics for the sablefish population coastwide, to 1) confirm or modify the structural assumptions present in the operating model, and 2) allow for exploration of how oceanic conditions might lead to changes in the recruitment paradigm for sablefish, and effects on management performance. This analysis could inform the development of additional operating models and allow for investigation of the interaction of climate change and spatial uncertainty for high-value species such as sablefish.

Future phases of this work could also integrate economic models to investigate performance metrics such as fishery profits, or allocation questions, both of interest to stakeholders (Kapur et al., 2021a).

BIBLIOGRAPHY

A'Mar, Z.T., Punt, A.E., Dorn, M.W., 2009. The evaluation of two management strategies for the Gulf of Alaska walleye pollock fishery under climate change. ICES Journal of Marine Science 66, 1614–1632. https://doi.org/10.1093/icesjms/fsp044

Abadi, F., Gimenez, O., Arlettaz, R., Schaub, M., 2010. An assessment of integrated population models: Bias, accuracy, and violation of the assumption of independence 91, 7–14.

Adams, G.D., Leaf, R.T., Ballenger, J.C., Arnott, S.A., McDonough, C.J., 2018. Spatial variability in the growth of sheepshead (*Archosargus probatocephalus*) in the southeast US: Implications for assessment and management. Fisheries Research 206, 35–43. https://doi.org/10.1016/j.fishres.2018.04.023

Ahrestani, F.S., Saracco, J.F., Sauer, J.R., Pardieck, K.L., Royle, J.A., 2017. An integrated population model for bird monitoring in North America 27, 916–924. https://doi.org/10.1002/eap.1493

Austin, M.P., 2002. Spatial prediction of species distribution: An interface between ecological theory and statistical modelling. Ecological Modelling. https://doi.org/10.1016/S0304-3800(02)00205-3

Berger, A.M., Deroba, J.J., Bosley, K.M., Goethel, D.R., Langseth, B.J., Schueller, A.M., Hanselman, D.H., 2020. Incoherent dimensionality in fisheries management: Consequences of misaligned stock assessment and population boundaries. ICES Journal of Marine Science 78, 155–171. https://doi.org/10.1093/icesjms/fsaa203

Berger, A.M., Jones, M.L., Zhao, Y., Bence, J.R., 2012. Accounting for spatial population structure at scales relevant to life history improves stock assessment: The case for lake erie walleye sander vitreus. Fisheries Research 115-116, 44–59. https://doi.org/10.1016/j.fishres.2011.11.006

Bertalanffy, L.V., 1957. Quantitative laws in metabolism and growth. The Quarterly Review of Biology.

Bertalanffy, L.V., 1938. A quantitative theory of organic growth (inquiries on growth laws II). Human Biology 10, 181–213.

Beverton, R.J.H., Holt, S.J., 1957. On the dynamics of exploited fish populations. https://doi.org/10.1007/BF00044132

Booth, A.J., 2000. Incorporating the spatial component of fisheries data into stock assessment models. ICES Journal of Marine Science 57, 858–865. https://doi.org/10.1006/jmsc.2000.0816

Bosley, K.M., Goethel, D.R., Berger, A.M., Deroba, J.J., Fenske, K.H., Hanselman, D.H., Langseth, B.J., Schueller, A.M., 2019. Overcoming challenges of harvest quota allocation in spatially structured populations. Fisheries Research 220, 105344. https://doi.org/10.1016/j.fishres.2019.105344

Brodziak, J., Ianelli, J., Lorenzen, K., Jr, R.D.M., 2011. Estimating natural mortality in stock assessment applications (no. NMFS-f/SPO-119), NOAA technical memorandum.

Cadrin, S.X., 2020. Defining spatial structure for fishery stock assessment. Fisheries Research 221, 105397. https://doi.org/10.1016/j.fishres.2019.105397

Cadrin, S.X., Goethel, D.R., Morse, M.R., Fay, G., Kerr, L.A., 2019. "So, where do you come from?" The impact of assumed spatial population structure on estimates of recruitment. Fisheries Research 217, 156–168. https://doi.org/10.1016/j.fishres.2018.11.030

Carruthers, T.R., Walters, C.J., Mcallister, M.K., 2012. Evaluating methods that classify fisheries stock status using only fisheries catch data. Fisheries Research 119-120, 66–79. https://doi.org/10.1016/j.fishres.2011.12.011

Carvalho, F., Punt, A.E., Chang, Y., Maunder, M.N., Piner, K.R., 2017. Can diagnostic tests help identify model misspecification in integrated stock assessments ? Fisheries Research 192, 28–40. https://doi.org/10.1016/j.fishres.2016.09.018

Carvalho, F., Winker, H., Courtney, D., Kapur, M., Kell, L., Cardinale, M., Schirripag, M., Kitakado, T., Yemane, D., Piner, K.R., Maunder, M.N., Taylor, I., Wetzel, C.R., Doering, K., Johnsonm, K.F., Methot, R.D., 2021. A Cookbook for Using Model Diagnostics in Integrated Stock Assessments. Fisheries. https://doi.org/10.1016/j.fishres.2021.105959

Chandler, R.B., Clark, J.D., 2014. Spatially explicit integrated population models. Methods in Ecology and Evolution 5, 1351–1360. https://doi.org/10.1111/2041-210X.12153

Cleaver, S., Evans, D., 2019. Gulf of Alaska groundfish fishery management plan amendment action summaries. North Pacific Fish. Manag. Counc. 605 W. 4th Ave. Suite 306 Anchorage, AK 99301.

Coffin, B., Mueter, F., 2016. Environmental covariates of sablefish (*Anoplopoma fimbria*) and pacific ocean perch (*sebastes alutus*) recruitment in the Gulf of Alaska. Deep-Sea Research Part II: Topical Studies in Oceanography 132, 194–209. https://doi.org/10.1016/j.dsr2.2015.02.016

Conn, P.B., Williams, E.H., Shertzer, K.W., 2010. When can we reliably estimate the productivity of fish stocks? Canadian Journal of Fisheries and Aquatic Sciences. https://doi.org/10.1139/F09-194

Cope, J.M., Punt, A.E., 2008. Admitting ageing error when fitting growth curves: An example using the von bertalanffy growth function with random effects. Canadian Journal of Fisheries and Aquatic Sciences. https://doi.org/10.1139/F06-179

Cornthwaite, M., Granum, L., Haggarty, D., 2020. British Columbia groundfish fisheries and their investigations in 2019. Technical Sub-Committee of the Canada-United States Groundfish Committee 2507, 1–9.

Cummins, P.F., Freeland, H.J., 2007. Variability of the north pacific current and its bifurcation. Progress in Oceanography 75, 253–265. https://doi.org/10.1016/j.pocean.2007.08.006

Denson, L.T.S., Sampson, D.B., Stephens, A., 2016. Data needs and spatial structure considerations in stock assessments with regional differences in recruitment and exploitation. Canadian Journal of Fisheries and Aquatic Sciences 74, 1918–1929. https://doi.org/10.1139/cjfas-2016-0277

Eacker, D.R., Lukacs, P.M., Proffitt, K.M., Hebblewhite, M., 2017. Assessing the importance of demographic parameters for population dynamics using bayesian integrated population modeling. Ecological Applications 27, 1280–1293. https://doi.org/10.1002/eap.1521

Eaton, M.J., Link, W.A., 2011. Estimating age from recapture data: Integrating incremental growth measures with ancillary data to infer age-at-length. Ecological Applications. https://doi.org/10.1890/10-0626.1

Echave, K.B., Hanselman, D.H., Adkison, M.D., Sigler, M.F., 2012. Interdecadal change in growth of sablefish (*Anoplopoma fimbria*) in the northeast Pacific Ocean. Fishery Bulletin 110, 361–374.

Fay, G., Punt, A.E., Smith, A.D.M., 2011. Impacts of spatial uncertainty on performance of age structure-based harvest strategies for blue eye trevalla (*Hyperoglyphe antarctica*). Fisheries Research 110, 391–407. https://doi.org/10.1016/j.fishres.2011.04.015

Fenske, K.H., Berger, A.M., Connors, B., Cope, J.M., Cox, S.P., Haltuch, M.A., Hanselman, D.H., Kapur, M., Lacko, L., Lunsford, C., Rodgveller, C., Williams, B., 2019. Report on the 2018 international sablefish workshop (No. NFMS-AFSC-387). US Department of Commerce, National Marine Fisheries Service, Juneau, AK.

Fournier, D., Archibald, C.P., 1982. A general theory for analyzing catch at age data. Canadian Journal of Fisheries and Aquatic Sciences 39, 1195–1207. https://doi.org/10.1139/f82-157

Gabriel, W.L., Mace, P.M., 1999. A review of biological reference points in the context of the precautionary approach (No. NMFS-F/SPO-40). National Marine Fisheries Service.

Gauthier, G., Besbeas, P., Lebreton, J.-D., Morgan, B.J.T., 2007. Population growth in snow geese: A modeling approach integrating demographic and survey information. Ecology 88, 1420–1429. https://doi.org/10.1890/06-0953

Gertseva, V., Matson, S.E., Cope, J., 2017. Spatial growth variability in marine fish: Example from northeast Pacific groundfish. ICES Journal of Marine Science 74, 1602–1613. https://doi.org/10.1093/icesjms/fsx016

Goethel, D.R., Berger, A.M., 2016. Accounting for spatial complexities in the calculation of biological reference points: Effects of misdiagnosing population structure for stock status indicators. Canadian Journal of Fisheries and Aquatic Sciences 74, 1878–1894. https://doi.org/10.1139/cjfas-2016-0290

Goethel, D.R., Hanselman, D.H., Rodgveller, C.J., Echave, K.B., Williams, B.C., Shotwell, S.K., Sullivan, J.Y., Hulson, P.F., Malecha, P.W., Siwicke, K.A., Chris, R., 2021. Assessment of the sablefish stock in Alaska. North Pacific Fish. Manag. Counc. 605 W. 4th Ave. Suite 306 Anchorage, AK 99301.

Goethel, D.R., Ii, T.J.Q., Cadrin, S.X., 2011. Incorporating spatial structure in stock assessment: Movement modeling in marine fish population dynamics. Reviews in Fisheries Science, 19th series 1262, 119–136. https://doi.org/10.1080/10641262.2011.557451

Goetz, F.W., Jasonowicz, A.J., Roberts, S.B., 2018. What goes up must come down: Diel vertical migration in the deep-water sablefish (*Anoplopoma fimbria*) revealed by pop-up satellite archival tags. Fisheries Oceanography 27, 127–142. https://doi.org/10.1111/fog.12239

Guthery, F.S., Burnham, K.P., Anderson, D.R., 2003. Model selection and multimodel inference: A practical information-theoretic approach. The Journal of Wildlife Management. https://doi.org/10.2307/3802723

Haltuch, M.A., Johnson, K.F., Tolimieri, N., Kapur, M., Castillo-Jordan, C., 2019. Status of the sablefish stock in US waters in 2019. Pacific Fisheries Management Council, 7700 Ambassador Place NE, Suite 200, Portland, OR.

Hamel, O.S., 2014. A method for calculating a meta-analytical prior for the natural mortality rate using multiple life history correlates, in: ICES Journal of Marine Science. https://doi.org/10.1093/icesjms/fsu131

Hanselman, D., Heifetz, J., Echave, K.B., Dressel, S.C., 2014. Move it or lose it: Movement and mortality of sablefish tagged in Alaska. Canadian Journal of Fisheries and Aquatic Sciences 72, 238–251. https://doi.org/10.1139/cjfas-2014-0251

Hanselman, D.H., Lunsford, C.R., Rodgveller, C.J., 2017. Assessment of the sablefish stock in Alaska in 2017. National Marine Fisheries Service, Auke Bay Marine Station 11305 Glacier Highway Juneau, AK 99801.

Heller-Shipley, M.A., Stockhausen, W.T., Daly, B.J., Punt, A.E., Goodman, S.E., 2021. Should harvest control rules for male-only fisheries include reproductive buffers? A bering sea tanner crab (*Chionoecetes bairdi*) case study. Fisheries Research 243, 106049. https://doi.org/10.1016/j.fishres.2021.106049

Hilborn, R., Minte-Vera, C.V., 2008. Fisheries-induced changes in growth rates in marine fisheries: Are they significant? BULLETIN OF MARINE SCIENCE 83.

Ianelli, J., Holsman, K.K., Punt, A.E., Aydin, K., 2016. Multi-model inference for incorporating trophic and climate uncertainty into stock assessments. Deep-Sea Research Part II: Topical Studies in Oceanography. https://doi.org/10.1016/j.dsr2.2015.04.002

Jacobsen, N.S., Marshall, K.N., Berger, A.M., Grandin, C., Taylor, I.G., 2022. Climate-mediated stock redistribution causes increased risk and challenges for fisheries management. ICES Journal of Marine Science 79, 1120–1132. https://doi.org/10.1093/icesjms/fsac029

Jasonowicz, A.J., Goetz, F.W., Goetz, G.W., Nichols, K.M., 2016. Love the one you're with: Genomic evidence of panmixia in the sablefish (*Anoplopoma fimbria*). Canadian Journal of Fisheries and Aquatic Sciences 11, 1–11.

Johnson, C.J., Mumma, M.A., St-Laurent, M., 2019. Modeling multispecies predator–prey dynamics: Predicting the outcomes of conservation actions for woodland caribou. Ecosphere. https://doi.org/10.1002/ecs2.2622

Kapur, M., Connors, B., Devore, J.D., Fenske, K.H., Haltuch, M., Key, M., 2021a. Transboundary sablefish management strategy evaluation (MSE) workshop report.

Kapur, M., Haltuch, M., Connors, B., Rogers, L., Berger, A., Koontz, E., Cope, J., Echave, K., Fenske, K., Hanselman, D., Punt, A.E., 2020. Oceanographic features delineate growth zonation in northeast Pacific sablefish. Fisheries Research 222. https://doi.org/10.1016/j.fishres.2019.105414

Kapur, M., Lee, Q., Correa, G.M., Haltuch, M.A., Gertseva, V., Hamel, O.S., Sciences, F., Sciences, A., Fisheries, N., Oceanic, N., Marine, N., Service, F., 2021b. Status of sablefish (*Anoplopoma fimbria*) along the US west coast in 2021. Pacific Fisheries Management Council, 7700 Ambassador Place NE, Suite 200, Portland, OR.

Kapur, M., Siple, M.C., Olmos, M., Privitera-Johnson, K.M., Adams, G., Best, J., Castillo-Jordán, C., Cronin-Fine, L., Havron, A.M., Lee, Q., Methot, R.D., Punt, A.E., 2021c. Equilibrium reference point calculations for the next generation of spatial assessments. Fisheries Research 244, 106132. https://doi.org/10.1016/j.fishres.2021.106132

Kerr, L.A., Cadrin, S.X., Kovach, A.I., 2014. Consequences of a mismatch between biological and management units on our perception of atlantic cod off new england. ICES Journal of Marine Science 71, 1366–1381. https://doi.org/10.1093/icesjms/fsu113

Kim, H.J., Miller, A.J., McGowan, J., Carter, M.L., 2009. Coastal phytoplankton blooms in the southern California bight. Progress in Oceanography 82, 137–147. https://doi.org/10.1016/j.pocean.2009.05.002

Kristensen, K., Nielsen, A., Berg, C.W., Skaug, H., Bell, B.M., 2016. TMB: Automatic differentiation and laplace approximation. Journal of Statistical Software. https://doi.org/10.18637/jss.v070.i05

Kuykendall, K.M., Powell, E.N., Klinck, J.M., Moreno, P.T., Leaf, R.T., 2019. The effect of abundance changes on a management strategy evaluation for the atlantic surfclam (spisula solidissima) using a spatially explicit, vessel-based fisheries model. Ocean and Coastal Management 169, 68–85. https://doi.org/10.1016/j.ocecoaman.2018.11.008

Lacko, L., Acheson, S., Connors, B., 2020. Sablefish (*Anoplopoma fimbria*) trap surveys, october 9 - november 19, 2018 and october 8 - november 25, 2019. Canadian Technical Reports of Fisheries; Aquatic Sciences.

Langseth, B.J., Schueller, A.M., 2016. Calculation of population-level fishing mortality for single- versus multi-area models: Application to models with spatial structure. Canadian Journal of Fisheries and Aquatic Sciences 74, 1821–1831. https://doi.org/10.1139/cjfas-2016-0295

Larkin, P.A., 1977. Transactions of the american fisheries society an epitaph for the concept of maximum sustained yield. Transactions of the American Fisheries Society 106, 1–11. https://doi.org/10.1577/1548-8659(1977)106<1

Lee, H.H., Maunder, M.N., Piner, K.R., Methot, R.D., 2012. Can steepness of the stock-recruitment relationship be estimated in fishery stock assessment models? Fisheries Research. https://doi.org/10.1016/j.fishres.2012.03.001

Lee, H.H., Maunder, M.N., Piner, K.R., Methot, R.D., 2011. Estimating natural mortality within a fisheries stock assessment model: An evaluation using simulation analysis based on twelve stock assessments. Fisheries Research. https://doi.org/10.1016/j.fishres.2011.01.021

Lee, H., Piner, K.R., Taylor, I.G., Kitakado, T., 2019. On the use of conditional age at length data as a likelihood component in integrated population dynamics models. Fisheries Research 216, 204–211. https://doi.org/10.1016/j.fishres.2019.04.007

Mackas, D.L., Thomson, R.E., Galbraith, M., 2001. Changes in the zooplankton community of the British Columbia continental margin, 1985-1999, and their covariation with oceanographic conditions. Canadian Journal of Fisheries and Aquatic Sciences 58, 685–702. https://doi.org/10.1139/f01-009

Mason, J.C., Beamish, R.J., McFarlane, G.A., 1983. Sexual maturity, fecundity, spawning, and early life history of sablefish (*Anoplopoma fimbria*) off the Pacific coast of Canada. Canadian Journal of Fisheries and Aquatic Sciences. https://doi.org/10.1139/f83-247

Maunder, M.N., 2004. Standardizing catch and effort data : A review of recent approaches. Fisheries Research 70, 141–159. https://doi.org/10.1016/j.fishres.2004.08.002

Maunder, M.N., Piner, K.R., 2015. Contemporary fisheries stock assessment: Many issues still remain. ICES Journal of Marine Science 72, 7–18. https://doi.org/10.1093/icesjms/fsu015

Maunder, M.N., Punt, A.E., 2014. A review of integrated analysis in fisheries stock assessment. Fisheries Research 142, 61–74. https://doi.org/10.1016/j.fishres.2012.07.025

Maunder, M.N., Hamel, O.S., Lee, H.-H., Piner, K.R., Cope, J.M., Punt, A.E., Ianelli, J.N., Castillo-Jordán, C., Kapur, M.S., Methot, R.D., 2023. A review of estimation methods for natural mortality and their performance in the context of fishery stock assessment. Fisheries Research 257, 106489. https://doi.org/10.1016/j.fishres.2022.106489

Mazur, M.D., Jesse, J., Cadrin, S.X., Truesdell, S.B., Kerr, L., 2023. Consequences of ignoring climate impacts on new england groundfish stock assessment and management. Fisheries Research 262, 106652. https://doi.org/10.1016/j.fishres.2023.106652

McDevvitt, M., 1990. Growth analysis of sablefish from mark-recapture data from the northeast Pacific (PhD thesis). University of Washington.

McGarvey, R., Fowler, A.J., 2002. Seasonal growth of king george whiting (sillaginodes punctata) estimated from length-at-age samples of the legal-size harvest. Fishery Bulletin 100, 545–558.

Methot, R.D., Wetzel, C.R., 2013. Stock synthesis: A biological and statistical framework for fish stock assessment and fishery management. Fisheries Research 142, 86–99. https://doi.org/10.1016/j.fishres.2012.10.012

Okamoto, D.K., Hessing-Lewis, M., Samhouri, J.F., Shelton, A.O., Stier, A., Levin, P.S., Salomon, A.K., 2020. Spatial variation in exploited metapopulations obscures risk of collapse. Ecological Applications 30, 1–16. https://doi.org/10.1002/eap.2051

Pacific Fisheries Management Council, 2013. Pacific coast fishery ecosystem plan for the US portion of the California Current large marine ecosystem. Pacific Fisheries Management Council 7700 NE Ambassad. Place, Suite 101, Portland, Oregon, 97220.

Pacifici, K., Reich, B.J., Miller, D.A.W., Gardner, B., Stauffer, G., Singh, S., McKerrow, A., Collazo, J.A., 2017. Integrating multiple data sources in species distribution modeling: A framework for data fusion. Ecology. https://doi.org/10.1002/ecy.1710

Palacios-Abrantes, J., Reygondeau, G., Wabnitz, C.C.C., Cheung, W.W.L., 2020. The transboundary nature of the world's exploited marine species. Sci Rep 10, 17668. https://doi.org/10.1038/s41598-020-74644-2

Piner, K.R., Lee, H.H., Maunder, M.N., Methot, R.D., 2011. A simulation-based method to determine model misspecification: Examples using natural mortality and population dynamics models. Marine and Coastal Fisheries. https://doi.org/10.1080/19425120.2011.611005

Pörtner, H.O., Knust, R., 2007. Climate change affects marine fishes through the oxygen limitation of thermal tolerance. Science. https://doi.org/10.1126/science.1135471

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

Punt, A.E., 2023. Those who fail to learn from history are condemned to repeat it: A perspective on current stock assessment good practices and the consequences of not following them. Fisheries Research 261, 106642. https://doi.org/10.1016/j.fishres.2023.106642

Punt, A.E., 2019a. Modelling recruitment in a spatial context: A review of current approaches, simulation evaluation of options, and suggestions for best practices. Fisheries Research 217, 140–155. https://doi.org/10.1016/j.fishres.2017.08.021

Punt, A.E., 2019b. Spatial stock assessment methods: A viewpoint on current issues and assumptions. Fisheries Research 213, 132–143. https://doi.org/10.1016/j.fishres.2019.01.014

Punt, A.E., Allen Akselrud, C., Cronin-Fine, L., 2017. The effects of applying mis-specified age- and size-structured models. Fisheries Research 188, 58–73. https://doi.org/10.1016/j.fishres.2016.11.017

Punt, A.E., Butterworth, D.S., Moor, C.L. de, De Oliveira, J.A.A., Haddon, M., 2016a. Management strategy evaluation: Best practices. Fish and Fisheries 17, 303–334. https://doi.org/10.1111/faf.12104

Punt, A.E., Donovan, G.P., 2007. Developing management procedures that are robust to uncertainty: Lessons from the international whaling commission. ICES Journal of Marine Science 64, 603–612. https://doi.org/10.1093/icesjms/fsm035

Punt, A.E., Haddon, M., Little, L.R., Tuck, G.N., 2016b. The effect of marine closures on a feedback control management strategy used in a spatially aggregated stock assessment: A case study based on pink ling in australia. Canadian Journal of Fisheries and Aquatic Sciences 74, 1960–1973. https://doi.org/10.1139/cjfas-2016-0017

Punt, A.E., Haddon, M., Tuck, G.N., 2015. Which assessment configurations perform best in the face of spatial heterogeneity in fishing mortality, growth and recruitment? A case study based on pink ling in australia. Fisheries Research 168, 85–99. https://doi.org/10.1016/j.fishres.2015.04.002

Punt, A.E., Smith, D.C., KrusicGolub, K., Robertson, S., 2008. Quantifying age-reading error for use in fisheries stock assessments, with application to species in australia's southern and eastern scalefish and shark fishery. Can. J. Fish. Aquat. Sci. 65, 1991–2005. https://doi.org/10.1139/F08-111

R Core Team, 2019. R: A language and environment for statistical computing. Vienna, Austria.

Ralston, S., O'Farrell, M.R., 2008. Spatial variation in fishing intensity and its effect on yield. Canadian Journal of Fisheries and Aquatic Sciences 65, 588–599. https://doi.org/10.1139/F07-174

Regehr, E.V., Hostetter, N.J., Wilson, R.R., Rode, K.D., Martin, M.S., Converse, S.J., 2018. Integrated population modeling provides the first empirical estimates of vital rates and abundance for polar bears in the chukchi sea. Scientific Reports. https://doi.org/10.1038/s41598-018-34824-7

Rhodes, J.R., Ng, C.F., Villiers, D.L. de, Preece, H.J., McAlpine, C.A., Possingham, H.P., 2011. Using integrated population modelling to quantify the implications of multiple threatening processes for a rapidly declining population. Biological Conservation. https://doi.org/10.1016/j.biocon.2010.12.027

Riecke, T.V., Williams, P.J., Behnke, T.L., Gibson, D., Leach, A.G., Sedinger, B.S., Street, P.A., Sedinger, J.S., 2019. Integrated population models: Model assumptions and inference. Methods in Ecology and Evolution. https://doi.org/10.1111/2041-210X.13195

Rodionov, S.N., 2004. A sequential algorithm for testing climate regime shifts. Geophysical Research Letters 31, 2–5. https://doi.org/10.1029/2004GL019448

Rodionov, S., Overland, J.E., 2005. Application of a sequential regime shift detection method to the bering sea ecosystem. ICES Journal of Marine Science 62, 328–332. https://doi.org/10.1016/j.icesjms.2005.01.013

Rogers, L., Anderson, S.C., Aulthouse, B., Bosley, K.M., Burton, Connors, B., Cox, S.P., Echave, K.B., Fenske, K.H., Goethel, D.R., Haltuch, M.A., Hanselman, D.H., Kapur, M.S.,

Lacko, L., Lunsford, C., Punt, A.E., Rodgveller, C.J., Sogard, S.M., Sullivan, J.Y., Williams, Ben., In Prep. Sablefish movement reveals transboundary stock structure in the northeast Pacific.

Rudolph, T.D., Drapeau, P., Imbeau, L., Brodeur, V., Légaré, S., St-Laurent, M.H., 2017. Demographic responses of boreal caribou to cumulative disturbances highlight elasticity of range-specific tolerance thresholds. Biodiversity and Conservation. https://doi.org/10.1007/s10531-017-1292-1

Rutecki, L., Rodgveller, C.J., Lunsford, C.R., 2016. National marine fisheries service longline survey data report and survey history , 1990-2014 1990–2014. https://doi.org/10.7289/V5/TM-AFSC-324

Sampson, D.B., 2014. Fishery selection and its relevance to stock assessment and fishery management. Fisheries Research. https://doi.org/10.1016/j.fishres.2013.10.004

Schaefer, M.B., 1968. Methods of estimating effects of fishing on fish populations. Transactions of the American Fisheries Society 97, 231–241. https://doi.org/10.1577/1548-8659(1968)97[231:moeeof]2.0.co;2

Schaub, M., Abadi, F., 2011. Integrated population models: A novel analysis framework for deeper insights into population dynamics. J Ornithol 152, 227–237. https://doi.org/10.1007/s10336-010-0632-7

Schaub, M., Gimenez, O., Sierro, A., Arlettaz, R., 2007. Use of integrated modeling to enhance estimates of population dynamics obtained from limited data. Conservation Biology 21, 945–955. https://doi.org/10.1111/j.1523-1739.2007.00743.x

Shackell, N.L., Frank, K.T., Nye, J.A., Heyer, C.E. den, 2016. A transboundary dilemma: Dichotomous designations of atlantic halibut status in the northwest atlantic. ICES Journal of Marine Science 73, 1798–1805. https://doi.org/f8w25x

Shelton, A.O., Mangel, M., 2012. Estimating von Bertalanffy parameters with individual and environmental variations in growth. Journal of Biological Dynamics 6, 3–30. https://doi.org/10.1080/17513758.2012.697195

Sogard, S.M., Berkeley, S.A., 2017. Patterns of movement, growth, and survival of adult sablefish (*Anoplopoma fimbria*) at contrasting depths in slope waters off oregon. Fishery Bulletin 115(2):233–251. https://doi.org/10.7755/FB.115.2.10

Somers, K.A., Richerson, K., Tuttle, V., McVeigh, J., 2022. Estimated discard and catch of groundfish species in the 2021 u.s. West coast fisheries (No. 1). National Marine Fisheries Service.

Stachura, M.M., Essington, T.E., Mantua, N.J., Hollowed, A.B., Haltuch, M.A., Spencer, P.D., Branch, T.A., Doyle, M.J., 2014. Linking northeast Pacific recruitment synchrony to environmental variability. Fisheries Oceanography 23, 389–408. https://doi.org/10.1111/fog.12066

Stawitz, C.C., Essington, T.E., Branch, T.A., Haltuch, M.A., Hollowed, A.B., Spencer, P.D., 2015. A state-space approach for detecting growth variation and application to north pacific

groundfish. Canadian Journal of Fisheries and Aquatic Sciences 72, 1316–1328. https://doi.org/10.1139/cjfas-2014-0558

Szuwalski, C.S., Punt, A.E., 2015. Can an aggregate assessment reflect the dynamics of a spatially structured stock? Snow crab in the eastern bering sea as a case study. Fisheries Research 164, 135–142. https://doi.org/10.1016/j.fishres.2014.10.020

Taylor, B.M., Brandl, S.J., Kapur, M., Robbins, W.D., Johnson, G., Huveneers, C., Renaud, P., Choat, J.H., 2018. Bottom-up processes mediated by social systems drive demographic traits of coral-reef fishes. Ecology. https://doi.org/10.1002/ecy.2127

Taylor, I.G., Grandin, C., Hicks, A.C., Taylor, N., Cox, S., 2015. Status of the pacific hake (whiting) stock in US and Canadian waters in 2015. National Marine Fishery Service; Canada Department of Fisheries; Oceans.

Taylor, I.G., Methot, R.D., 2013. Hiding or dead? A computationally efficient model of selective fisheries mortality. Fisheries Research 142, 75–85. https://doi.org/10.1016/j.fishres.2012.08.021

Tempel, D.J., Peery, M.Z., Gutiérrez, R.J., 2014. Using integrated population models to improve conservation monitoring: California spotted owls as a case study. Ecological Modelling 289, 86–95. https://doi.org/10.1016/j.ecolmodel.2014.07.005

Then, A.Y., Hoenig, J.M., Hall, N.G., Hewitt, D.A., 2015. Evaluating the predictive performance of empirical estimators of natural mortality rate using information on over 200 fish species. ICES Journal of Marine Science 72, 82–92. https://doi.org/10.1093/icesjms/fsu136

Thorson, J.T., 2019. Guidance for decisions using the vector autoregressive spatio-temporal (VAST) package in stock, ecosystem, habitat and climate assessments. Fisheries Research 210, 143–161. https://doi.org/10.1016/j.fishres.2018.10.013

Thorson, J.T., Ianelli, J.N., Munch, S.B., Ono, K., Spencer, P.D., 2015a. Spatial delay-difference models for estimating spatiotemporal variation in juvenile production and population abundance. Canadian Journal of Fisheries and Aquatic Sciences. https://doi.org/10.1139/cjfas-2014-0543

Thorson, J.T., Munch, S.B., Cope, J.M., Gao, J., 2017. Predicting life history parameters for all fishes worldwide. Ecological Applications. https://doi.org/10.1002/eap.1606

Thorson, J.T., Skaug, H.J., Kristensen, K., Shelton, A.O., Ward, E.J., Harms, J.H., Benante, J.A., Inouye, B.D., 2015b. The importance of spatial models for estimating the strength of density dependence. Ecology 96, 1202–1212. https://doi.org/10.1890/14-0739.1

Tolimieri, N., Haltuch, M.A., 2023. Sea-level index of recruitment variability improves assessment model performance for sablefish (*Anoplopoma fimbria*). Can. J. Fish. Aquat. Sci. cjfas-2022-0238. https://doi.org/10.1139/cjfas-2022-0238

Tolimieri, N., Haltuch, M.A., Lee, Q., Jacox, M.G., Bograd, S.J., 2018. Oceanographic drivers of sablefish recruitment in the California current. Fisheries Oceanography 27, 458–474. https://doi.org/10.1111/fog.12266

Walter III, J.F., Peterson, C.D., Marshall, K., Deroba, J.J., Gaichas, S., Williams, B.C., Stohs, S., Tommasi, D., Ahrens, R., 2023. When to conduct, and when not to conduct, management strategy evaluations. ICES Journal of Marine Science fsad031. https://doi.org/10.1093/icesjms/fsad031

Waterhouse, L., Sampson, D., Maunder, M., Semmens, B., 2014. Using areas-as-fleets selectivity to model spatial fishing: Asymptotic curves are unlikely under equilibrium conditions. Fisheries Research 158. https://doi.org/10.1016/j.fishres.2014.01.009

Williams, A.J., Farley, J.H., Hoyle, S.D., Davies, C.R., Nicol, S.J., 2012. Spatial and sex-specific variation in growth of albacore tuna (thunnus alalunga) across the south pacific ocean. PLoS ONE 7, 1–10. https://doi.org/10.1371/journal.pone.0039318

Williams, B., In Prep. Spatiotemporal trends in sablefish (*Anoplopoma fimbria*) maturity curves in the northeast Pacific. National Marine Fisheries Service.

Winton, M.V., Wuenschel, M.J., McBride, R.S., 2014. Investigating spatial variation and temperature effects on maturity of female winter flounder (*Pseudopleuronectes americanus*) using generalized additive models. Canadian Journal of Fisheries and Aquatic Sciences. https://doi.org/10.1139/cjfas-2013-0617

Wood, S.N., 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. Journal of the Royal Statistical Society. Series B: Statistical Methodology. https://doi.org/10.1111/j.1467-9868.2010.00749.x

Wyeth, K., M. R., 2005. Summary of the 2005 British Columbia sablefish (*Anoplopoma fimbria*) research and assessment survey: Science Branch, Pacific Region Pacific Biological Station, Nanaimo, British Columbia V9T 6N7.

Ying, Y., Chen, Y., Lin, L., Gao, T., 2011. Risks of ignoring fish population spatial structure in fisheries management. Canadian Journal of Fisheries and Aquatic Sciences. https://doi.org/10.1139/F2011-116

Appendix A. Supplementary Information for Chapter 1

A.1 Figures

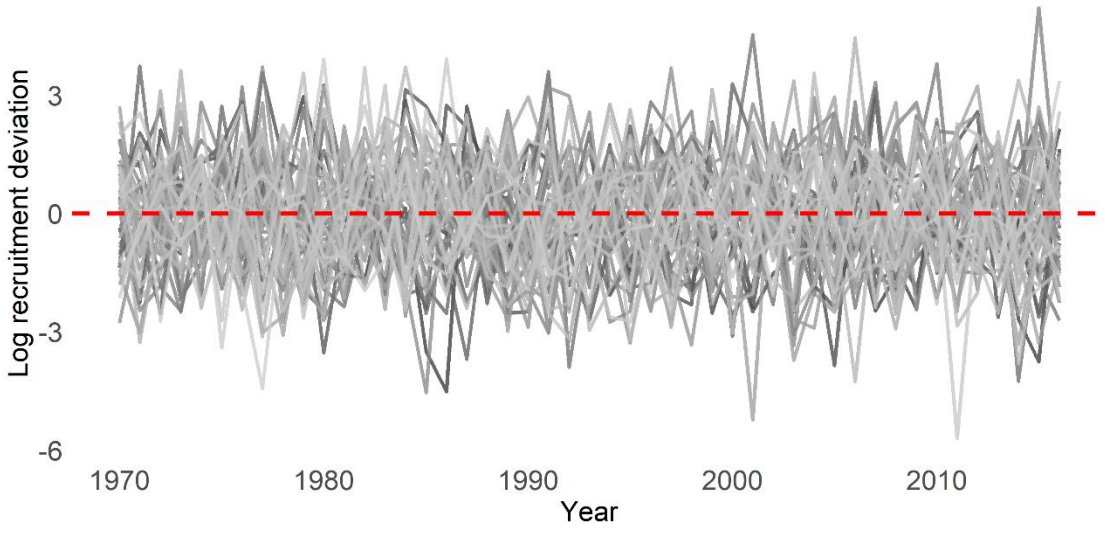


Figure A.1: Recruitment deviates used in unique OM replicates (grey lines) for the main recruitment period, 1980-2016.

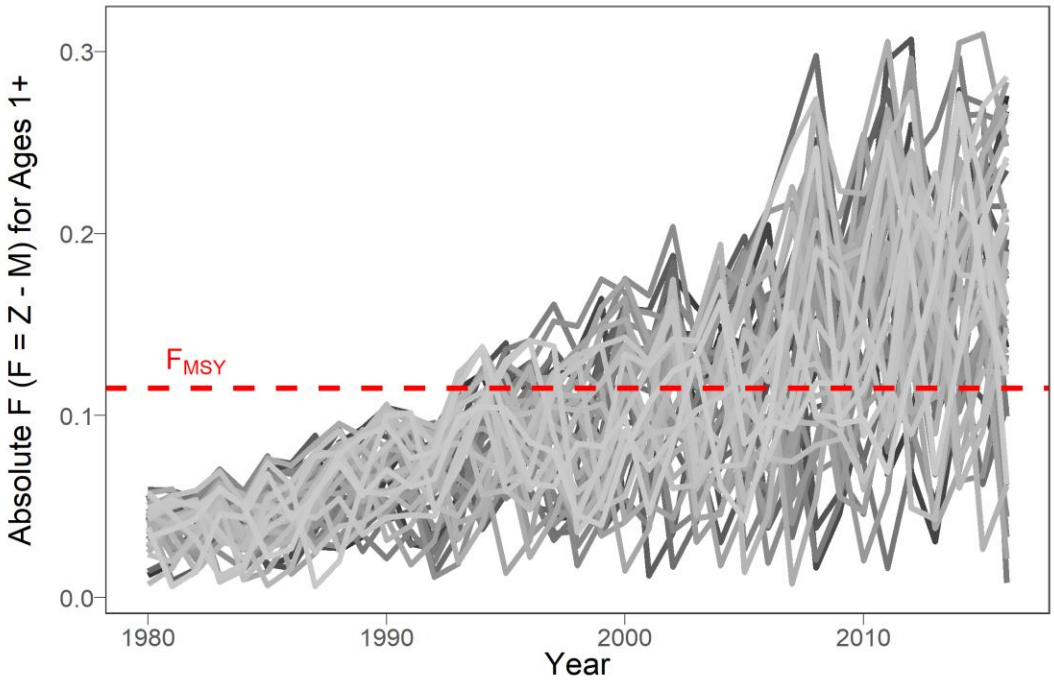


Figure A.2: Fully-selected fishing mortality rates for ages 1+ used in each operating model replicate (grey lines) and the fishing mortality rate corresponding to MSY (red dashed line).

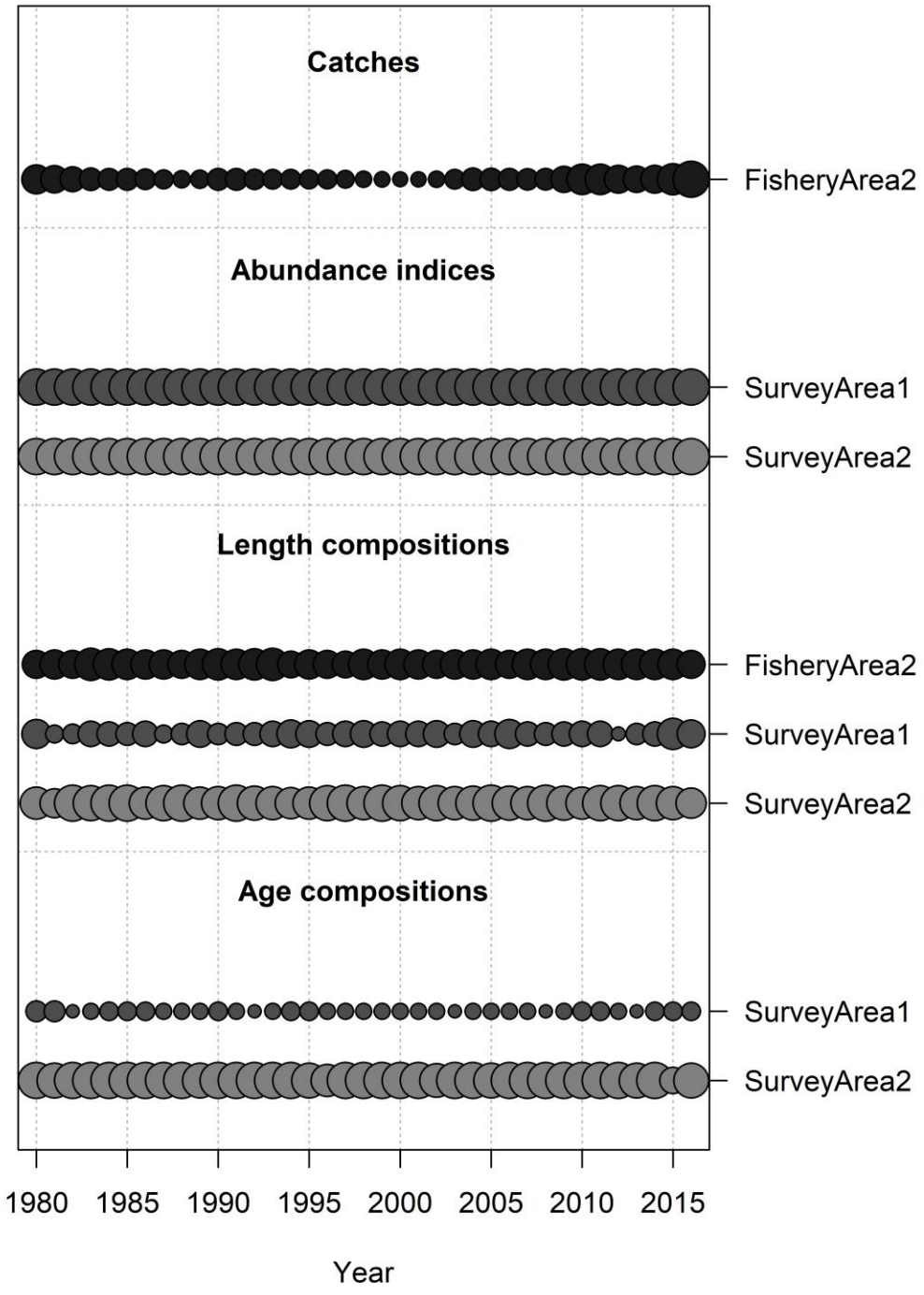


Figure A.3: Data used in the operating models. Circle size represents the relative amount of data available by year.

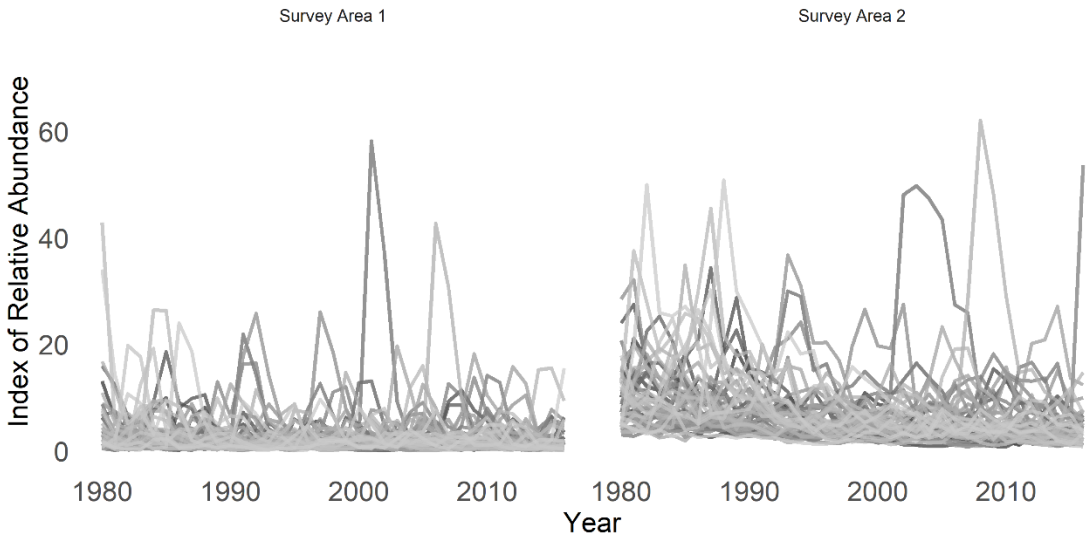


Figure A.4: Simulated survey indices used in unique OM replicates (grey lines) for each area.

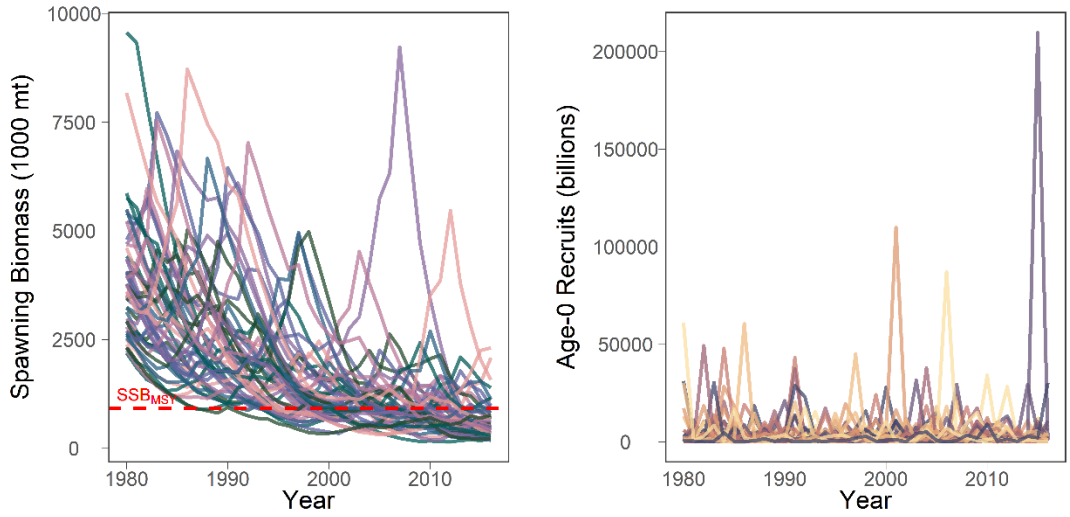


Figure A.5: Time series of spawning biomass (left panel) with biomass at maximum sustainable yield (dashed red line) and age-zero recruits (right panel) in replicates of operating models (colors).

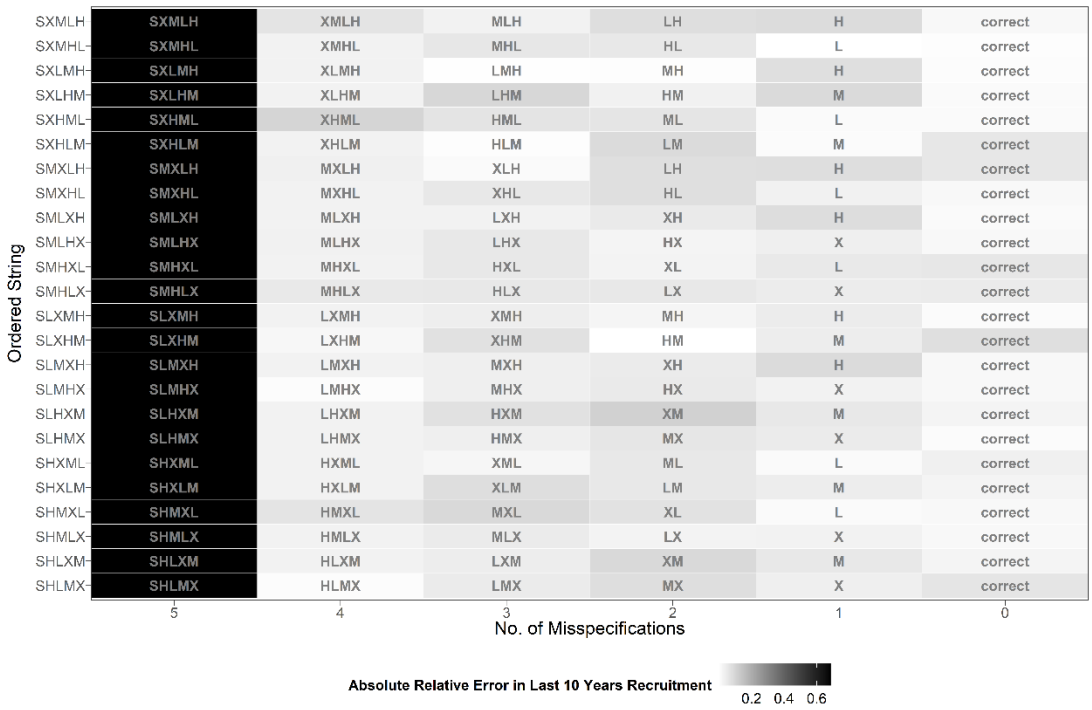


Figure A.6: As for Figure 1.4 (main text), except the results pertain to last 10 years of recruitment for Experiment 1.

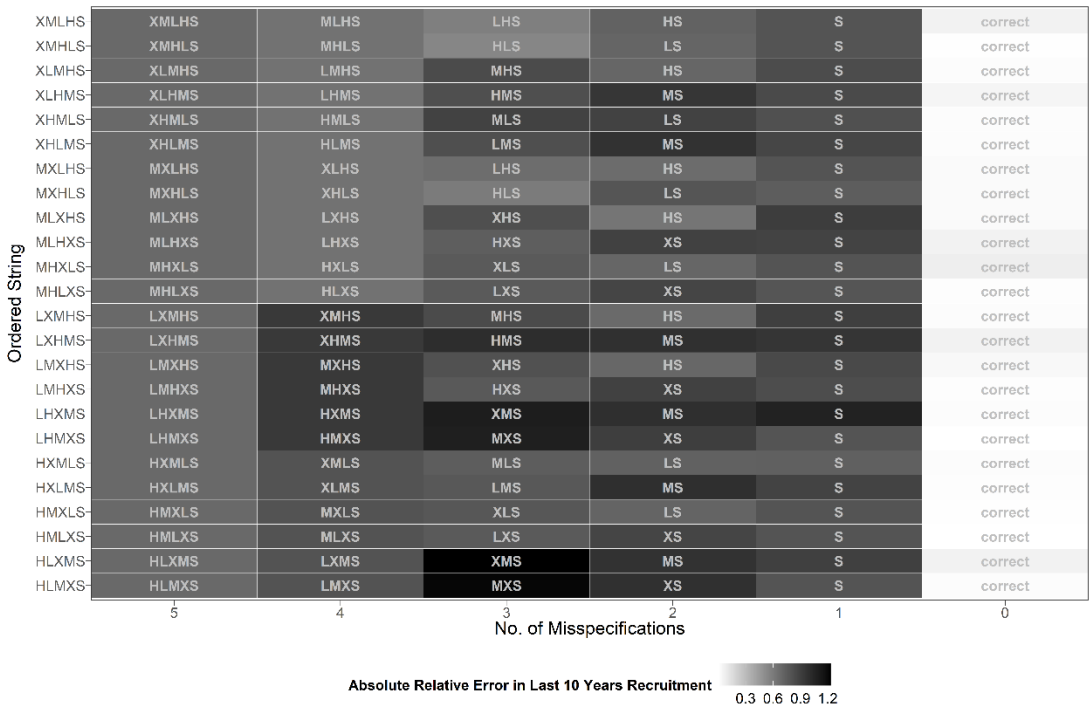


Figure A.7: As for previous figure, but results pertain to Experiment 2.

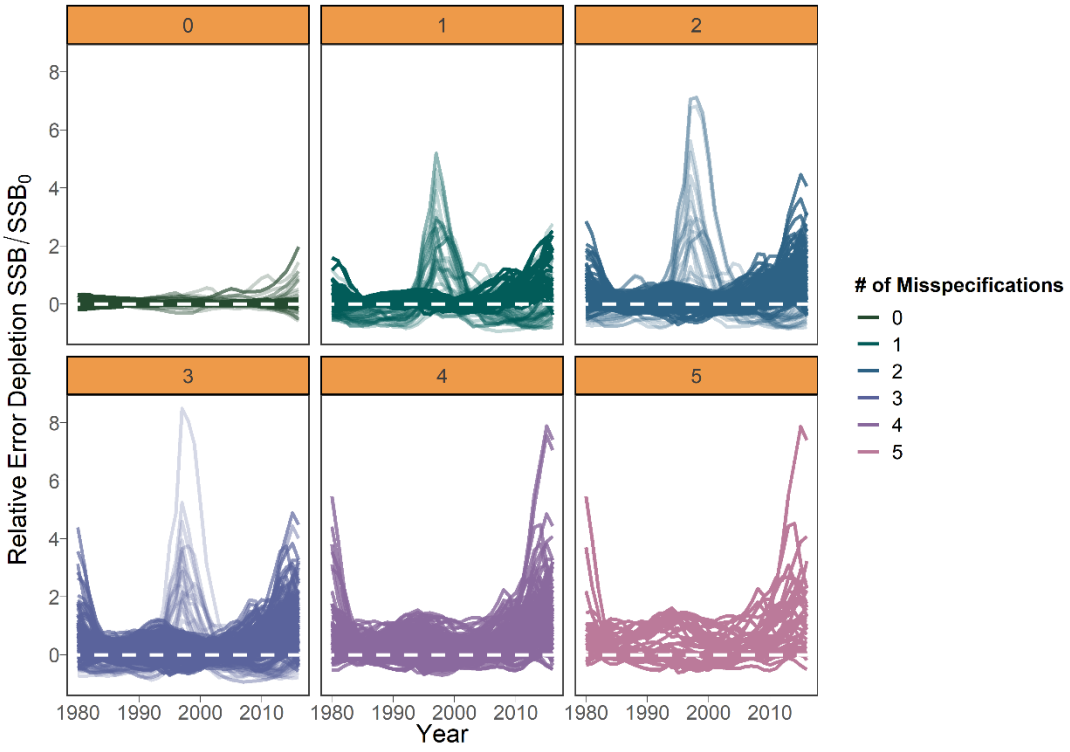


Figure A.8: Time series of relative error in depletion for Experiment 2. Panels and line colors correspond to the number of mis-specifications in the model. All models with one or fewer mis-specifications have space correctly specified.

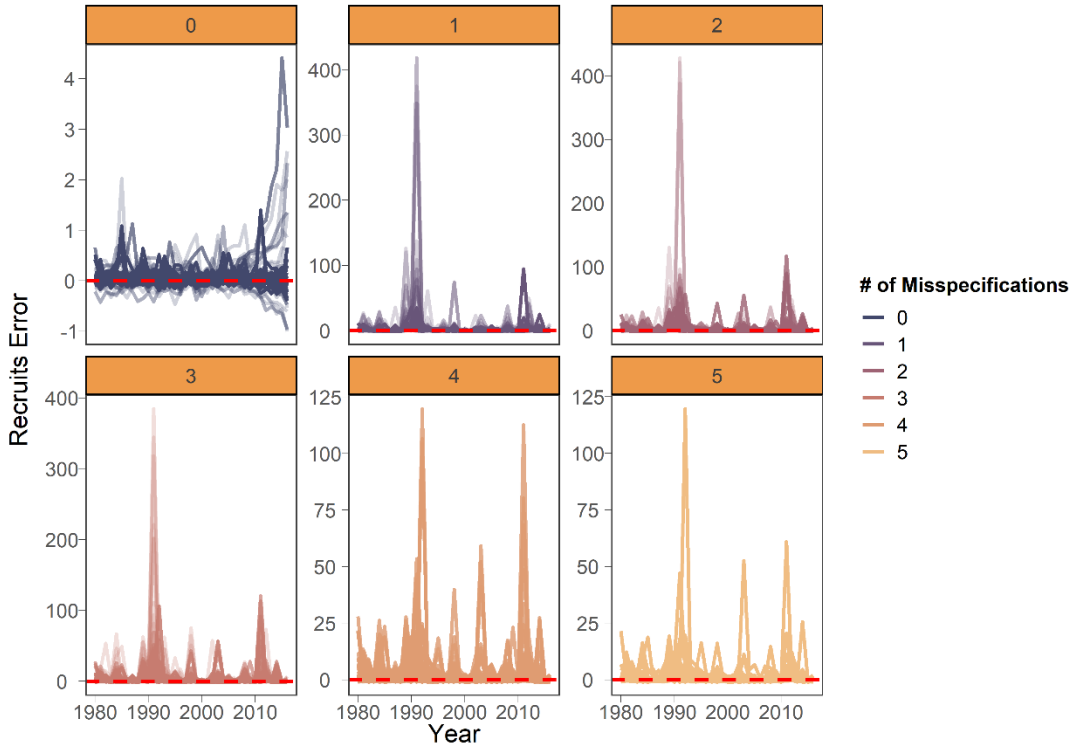


Figure A.9: Time series of relative error in recruitment for Experiment 2. Panels and line colors correspond to the number of mis-specifications in the model. All models with one or fewer mis-specifications have space correctly specified. Note that calculating relative error on very small numbers can result in high values.

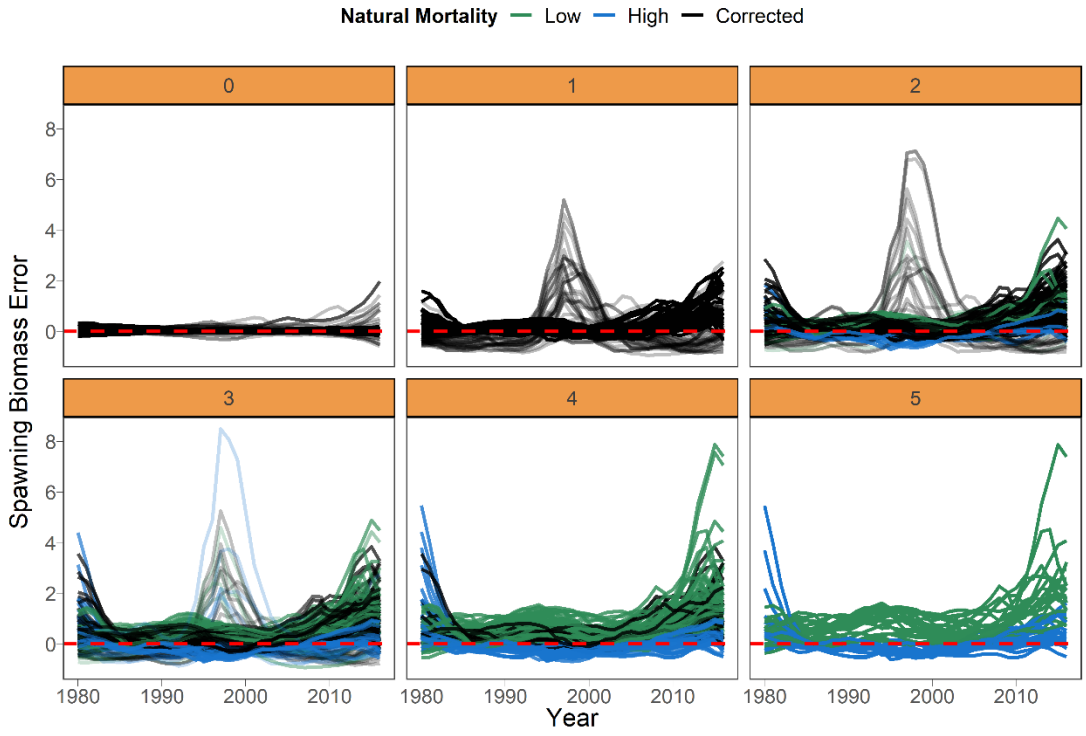


Figure A.10: Time series of relative error in SSB for Experiment 2. Line colors indicate whether natural mortality (*M*) was mis-specified high (blue) low (green) or correctly (black) in the EM. All models with one or fewer mis-specifications have space correctly specified.

Appendix B. Supplementary Information for Chapter 2

B.1 Generation of age-length data

The IBM is designed to mimic individual variation in growth for an unexploited fishery. The model runs for 100 years. Generally, all fish within each simulation are subject to the same baseline life history parameters, with three growth "regimes" (defined by distinct values for the parameters of the growth equation, see below) assigned spatial ranges accordingly (see below). Here we detail the growth component of the IBM, code to execute the simulations is available here: https://github.com/mkapur/sab-growth/blob/master/IBM_master.R

B.1.1 Growth

The growth module of the IBM is a von Bertalanffy growth function parameterized in terms of L_1 and L_2:

$$L_\infty = L_1 + \frac{L_2 - L_1}{1 - exp\big(-k(a_2 - a_1)\big)} \qquad (B.1)$$

where $L_{1,2}$ represent the lengths of a fish at ages $a_{1,2}$, and k is the growth coefficient. The size of individual i at age a is defined by its length in the previous year and a growth increment i that is lognormal:

$$L_{i,a} = \begin{Bmatrix} L_1\varepsilon & for\ a = 1 \\ L_{i,a-1} + (I_{i,a-1})\varepsilon_{i,a} & for\ a > 1 \end{Bmatrix} \quad (B.2)$$

where $I_{i,a} = (L_\infty - L_{i,a})(1 - e^K)$ and $\varepsilon_{i,a} = \exp\left(N(0, \sigma_\varepsilon) - \frac{{\sigma_\varepsilon}^2}{2}\right)$; $\sigma_\varepsilon = 0.025$ for all ages and growth regimes (Table B.1). The value for σ_ε was selected so that the growth increments were similar to those for sablefish (Figure B.1).

B.1.2 Survival

The composition of the fishery during year *y* includes all surviving fish from recruitment to a maximum age (represented here as a plus group ${a_2}^+$). After recruitment, all fish are subject to mortality, which consists only of natural mortality (*M*; set to 0.25yr^{-1} for all ages and years) as

there is no fishery, thus fishing mortality (typically denoted F) and selectivity are ignored. Because no fishing pressure nor selectivity acted upon the simulated population, we are unconcerned about variation in growth that can either be engendered (over time) or misrepresented by differences in selectivity[4]. Whether an individual survives the year is simulated by randomly drawing a number u from $U[0,1]$ and allowing the individual to survive if this number is less than e^{-M}, i.e.:

$$S_{i,y,a} = \{\begin{matrix} 1 \; if \; u_i < e^{-M} \\ 0 \; if \; u_i > e^{-M} \end{matrix} \quad for \; 0 \leq a \leq a_2{}^{+}\} \qquad (B.3)$$

where $S_{i,y,a}$ is a value to indicate whether individual i is alive (1) or dead (0) during year y when it would be of age a.

We initialized the population in year zero at equilibrium as follows. The number of fish in the system is scaled by R_0, or the expected number of recruits for an unfished population. For computational efficiency we set R_0=12 (see sensitivities on sample size in later sections). Numbers at age are rounded to the nearest integer.

$$N_{y=0,a} = \{\begin{matrix} R_0 & for \; a = 0 \\ N_{0,a-1}e^{-M} & for \; 1 \leq a < a_2{}^{+} \\ \frac{N_{0,a_{2+}-1}}{1-e^{-M}} & for \; a = a_2{}^{+} \end{matrix}\} \qquad (B.4)$$

Lengths and weights for each individual in year 0 are calculated as in Equations (B.1) and (B.2). The starting stock spawning biomass (SSB_0) is calculated using the sum of expected weights, maturities, and numbers of individuals at all ages in year 0. Maturity at age $x_{0,a}$ is given by:

$$SSB_0 = \sum_{1}^{a_2{}^{+}} N_{0,a} W_{0,a} x_{0,a} \qquad (B.5)$$

Otherwise, the total number of individuals of age a in the population in simulation year y (where $y > 0$) is given by:

$$N_{y,a} = \sum_{1}^{i} S_{i,y,a} \qquad (B.6)$$

[4] This was not assumed, however, for the sablefish application.

B.1.3 Recruitment

Recruitment in the IBM is governed by a Beverton-Holt stock-recruitment function (Beverton and Holt, 1957), and a size-based maturity ogive that determines the probability of individual i maturing during a given year *y*, $p_{i,y}$. The maturity ogives were fixed for all regimes, with L_{50} (the length at 50% maturity) at 75 cm, and the slope of the ogive at -0.1034. The probability of an individual maturing in a given year $p_{i,y}$ is conditional on the probability in the current and previous year of being mature $x_{i,y}$, and $x_{i,y-1}$, which is then converted to $m_{i,y}$ (0 for immature; 1 for mature) by randomly drawing a number u_i from $U[0,1]$ and defining the animal as mature if ui is less than $p_{i,y}$.

Recruitment during a given year R_y is the sum of the sum of empirical weights of each individual that is mature in that year, which is governed by a deterministic exponential length-weight relationship. The parameters of this relationship were the same for all regimes. Recruitment is subject to variation via a bias-corrected lognormal recruitment deviation $\partial_{r,y}$.

$$x_{iy} = \frac{1}{1 + exp - 0.1034\left(L_{iy} - L_{50}\right)} \qquad (B.7)$$

$$p_{iy} = \frac{x_{iy} - x_{i,y-1}}{1 - x_{i,y-1}} \qquad (B.8)$$

$$W_{i,y} = 1.35 \times 10^{-6} {L_{i,y}}^{3.42} \qquad (B.10)$$

$$SSB_y = \sum_{1}^{i} S_{i,y} W_{i,y} m_{i,y} \qquad (B.11)$$

$$R_y = \frac{4hR_0 SSB_{y-1}}{SSB_0(1-h) + SSB_{y-1}(5h-1)} exp\left(\partial_{r,y}\right) \qquad (B.12)$$

$$\partial_{r,y} = N(0, \sigma_r) - 0.5{\sigma_r}^2 \qquad (6.13)$$

B.2 Assigning spatio-temporal variation to synthetic populations

The simulation testing component required generation of datasets that comprised variation in fish length-at-age across space and/or time. To obtain spatial variation in length-at-age, we conducted simulations using one of two growth "regimes". The synthetic populations were designed to mimic

the level of variation among L_1 and L_2 in the sablefish dataset, which ranged from 10% to 40% between regions (see main text, Table 2.3 and Figure 2.8; we used a slightly conservative difference in 20% for each of L_1 and L_2 to generate each synthetic population). Other parameters were held constant across regimes. Spatial scenarios tested are described in Table 2.1. To simulate spatial zones, fish locations were sampled from a uniform distribution with boundaries specific to a certain growth regime. In all except Scenario 4, where the break is located at 48° and non-spatial scenarios, the latitude and longitude of fish grown under regime 1 were sampled independently and at random from a uniform distribution between 0° and 25°; for simulations with spatial variation, fish grown under regime 2 have latitude and longitude sampled uniformly from 25° to 50°. In Scenario 4, all simulated fish were assigned latitudes sampled independently and at random from a uniform distribution from 0° to 50°. Fish simulated under regime 2 were assigned longitudes sampled randomly from 0° to 48° and fish simulated under regime 2 have longitudes sampled randomly from 48° to 50°, forming a vertical "band" of larger fish in higher longitudes.

Commentary on comparison of simulation study and STARS method

We observed decreased ability of the method to detect breakpoints near the edge of the range, with a true break at 48° inconsistently being assigned between 46° and 50°. This outcome, and the resultant low coverage probabilities for parameters L_1 and L_2 for this scenario were likely due to the smaller number of samples present in the 'edge' of the simulated space, and contrast in length-at-age between the two regions, which rendered estimates of aggregated data uninformative. This suggests that fishery scientists and managers may need alternative tools to detect and appropriately consider variation in growth at the extremes of a stock's spatial domain, or occurring at present. Such breakdown of detection methods at the margins of a series (at the edges of a study region, or at the end of a time-series) has been documented in Rodionov (2004), who developed a method using sequential t-tests (STARS) to perform edge-case detection, and applied it to detect ecosystem regime shifts in the Bering Sea (Rodionov and Overland, 2005). The t-test approach can be tuned by the researcher to control the level of significance that determines a regime shift (or breakpoint), presenting the same challenge of spurious and/or missed detections depending on the sensitivity of the statistical test applied. The comparison with the STARS method demonstrated that the GAM-based method performs better at detecting spatial-temporal breakpoints, except for scenarios where the break occurs at the edge of the study system, which was expected. In terms of the

coverage probabilities, both methods had a slightly reduced ability to correctly estimate L_2, with the STARS method performing slightly better. This outcome is likely due to two interacting processes: the GAM-based method's sensitivity to temporal variation, and bias in parameter estimates due to reduced sample sizes at high ages. The GAM appeared to be more sensitive to temporal signals in the datasets, and although it correctly detected (or correctly failed to detect) a temporal breakpoint in the majority of datasets, when it mis-detected a year break it did so seemingly at random, thus splitting the dataset into arbitrary groups and leading to lower accuracy of estimation. This phenomenon was more pronounced for L_2 since L_2 relies on fish near the terminal age, of which there are typically fewer, and can lead to bias in the resultant estimate when the already-small sample of fish at age a_2 is split further due to spurious year detections. Indeed, for all scenarios besides Scenarios 1 and 4 (no breaks and break-at-edge), the margin by which L_2 was missed was greater in simulations that mis-detected the year break (Table B.4). For assessment methods that estimate VBGF growth parameters within the assessment model, this low-data/low-accuracy issue for the terminal length may induce greater uncertainty (e.g., the need for priors with lower standard deviations) until targeted survey sampling can improve precision in less-represented management regions.

B.3 Tables

Table B.1: Parameter symbols, definitions and values used in the simulation study.

Module	**Parameter**	**Definition**	**Value**
Growth	L_1	Length at age a_1 (cm)	10 (regime 1) 70 (regime 2)
Growth	L_2	Length at age a_2 (cm)	12 (regime 1) 84 (regime 2)
Growth	k	Growth coefficient (year^{-1})	0.30 (regime 1) 0.30 (regime 2)
Growth	a_1	Age at L_1 (years)	3
Growth	a_2	Age at L_2 (years)	30
Growth	σ_ε	Lognormal growth error term	0.1
Growth	a	Multiplier of length-weight function (g/cm)	1.35e-6
Growth	b	Exponent of length-weight function	3.427
Survival	M	Natural mortality (yr^{-1})	0.25
Recruitment	r	Slope of maturity ogive	-0.1034
Recruitment	L_{50}	Length at 50% maturity (cm)	44
Recruitment	h	Steepness of Beverton-Holt SRR	0.9
Recruitment	R_0	Maximum number of recruits per year	12
Recruitment	σ_r	Variation in recruitment	0.1

Table B.2: Summary of true break points, coverage probabilities of the endpoints of the post-aggregation growth curves, and the proportion of simulations that detected the exact breakpoints each or all of the three smoothers. For the overlapping scenario (Scenario 3), spatial breakpoints were considered a match if they fell within the true range. This analysis was repeated for the same datasets with the number of age-six fish reduced by either 50% or 25%.

			Original sample size (average # age six fish = 530)		**With sample size halved**		**With sample size reduced by 25%**	
Scenario Number	**Scenario Description**	**True Break Points**	**Coverage probability for L_1, L_2**	**Proportion correct latitude, longitude, year**	**Coverage probability for L_1, L_2**	**Proportion correct latitude, longitude, year**	**Coverage probability for L_1, L_2**	**Proportion correct latitude, longitude, year**
1	No spatial breaks	None	0.96, 0.74	0.86, 0.86, 0.82	0.96, 0.02	0.95, 0.93, 0.75	0.95, 0.01	0.95, 0.91, 0.83
2	Single, spatial break in middle of range, with no overlap and strong contrast	25° Latitude and 25° Longitude	0.48, 0.43	0.92, 0.99, 0.89	0.47, 0.02	0.78, 0.8, 0.87	0.48, 0	0.82, 0.92, 0.86
3	Some overlap between regions	Between 20° and 25° Latitude	0.85, 0.58	0.84, 0.11, 0.91	0.86, 0	0.75, 0.21, 0.8	0.87, 0	0.79, 0.14, 0.88
4	Single spatial break at edge of range with no overlap	48° Longitude	0.27, 0.16	1, 1, 0.85	0.3, 0.02	0.99, 0.97, 0.92	0.29, 0.02	0.99, 0.97, 0.79
5	Single temporal break at year 50 (of 100); no spatial variability	None for latitude or longitude; all fish under regime 1 from years 0 to 49 and regime 2 thereafter	0.97, 0.78	0.92, 0.96, 0.5	0.9, 0.13	0.95, 0.96, 0.25	0.95, 0.08	0.92, 0.96, 0.33

Table B.3: Number of sablefish at key ages by sex and used in VBGF estimation for application study.

Age	Sex	n
4	F	4,366
4	M	3,204
6	F	4,413
6	M	3,404
10	F	2,064
10	M	1,765
30	F	168
30	M	231

Table B.4: Mean absolute error in estimated L_2 across simulated scenarios with a 95% CI for L_2 which did not contain the true value, depending on whether the temporal breakpoint was accurately detected (TRUE or FALSE). Average Error is computed as the mean of the absolute difference between the end of the estimated confidence interval and the true value; if the true value was higher than the confidence interval, the difference is measured from the upper end of the interval and vice versa.

Scenario Number	Scenario Description	Year accurately detected	Mean absolute error in L_2 (cm)
1	No spatial breaks	FALSE	0.403857
1	No spatial breaks	TRUE	0.464984
2	Single, spatial break in middle of range, with no overlap	FALSE	7.019933
2	Single, spatial break in middle of range, with no overlap	TRUE	6.975099
4	Single spatial break at edge of range with no overlap	FALSE	0.95649
4	Single spatial break at edge of range with no overlap	TRUE	1.165541
3	Some overlap between regions	FALSE	7.031669
3	Some overlap between regions	TRUE	6.415294
5	Single temporal break at year 50 (of 100); no spatial variability	FALSE	0.668129
5	Single temporal break at year 50 (of 100); no spatial variability	TRUE	0.463615

B.4 Figures

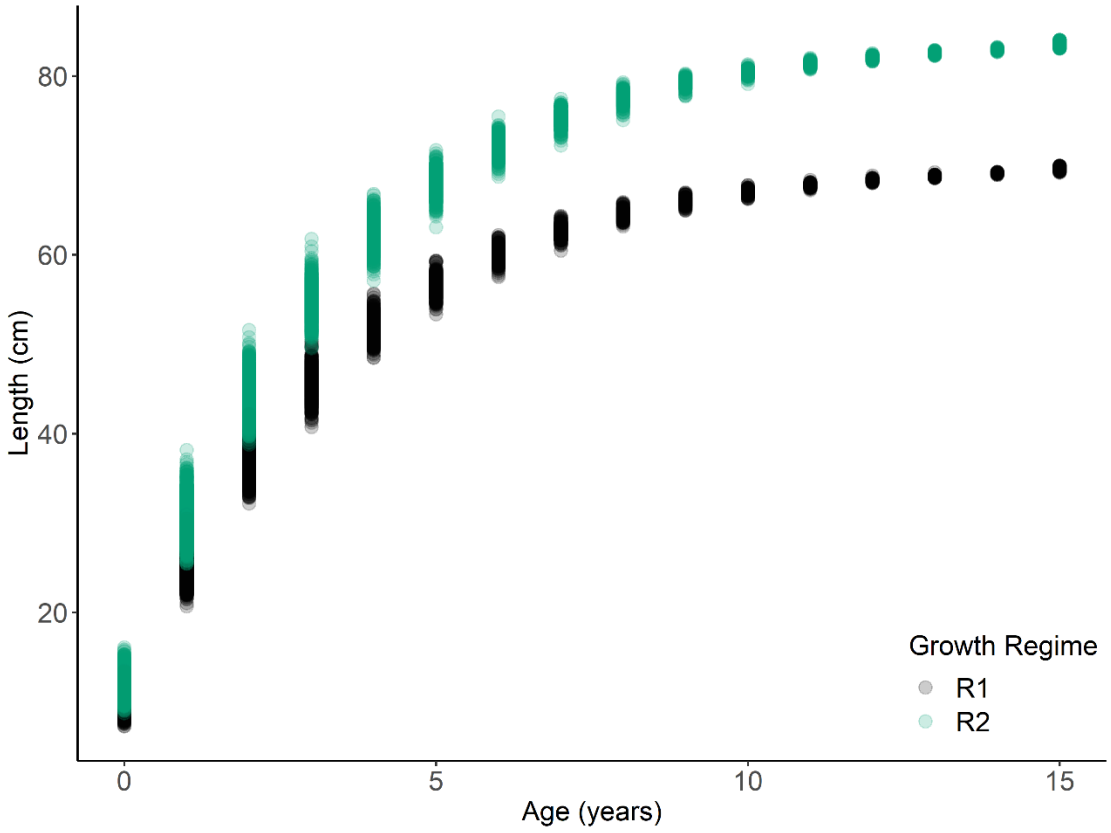


Figure B.1: Example growth trajectories from simulated populations. Each circle represents a simulated individual fish's length and age; colors correspond to the growth regime (i.e., growth curve) under which that fish was generated.

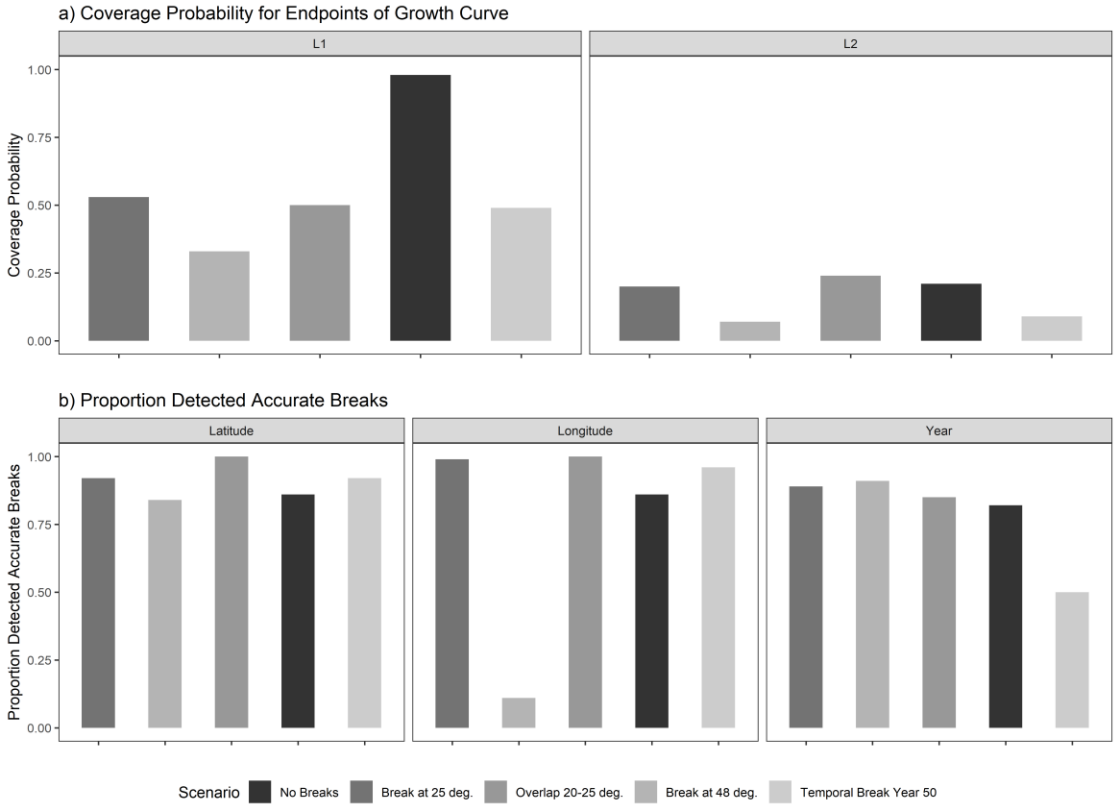


Figure B.2: Using the STARS method (Rodionov, 2004) a) coverage probabilities for the endpoints of the growth curve, L_1 (left) and L_2 (right), and b) proportion of 100 simulations for each spatial scenario wherein the correct latitudinal breaks (left), or longitudinal breaks (center) or yearly break (right) were detected.

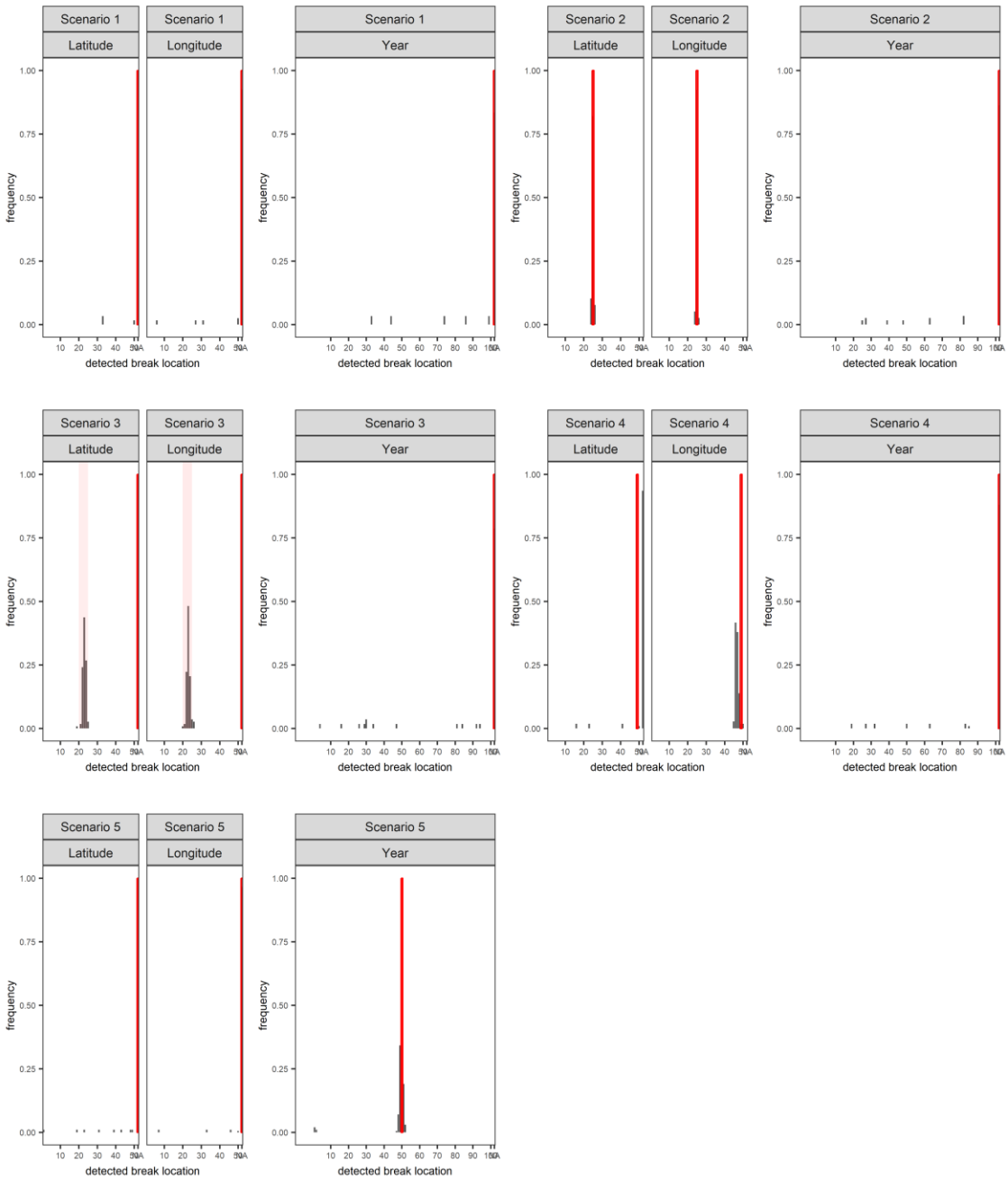


Figure B.3: Histogram of detected breakpoints (grey bars) from the GAM analysis by scenario. Vertical red bars indicate true breakpoints used to generate synthetic populations. For Scenario 3, the synthetic population overlapped between 20 and 25 degrees latitude and longitude.

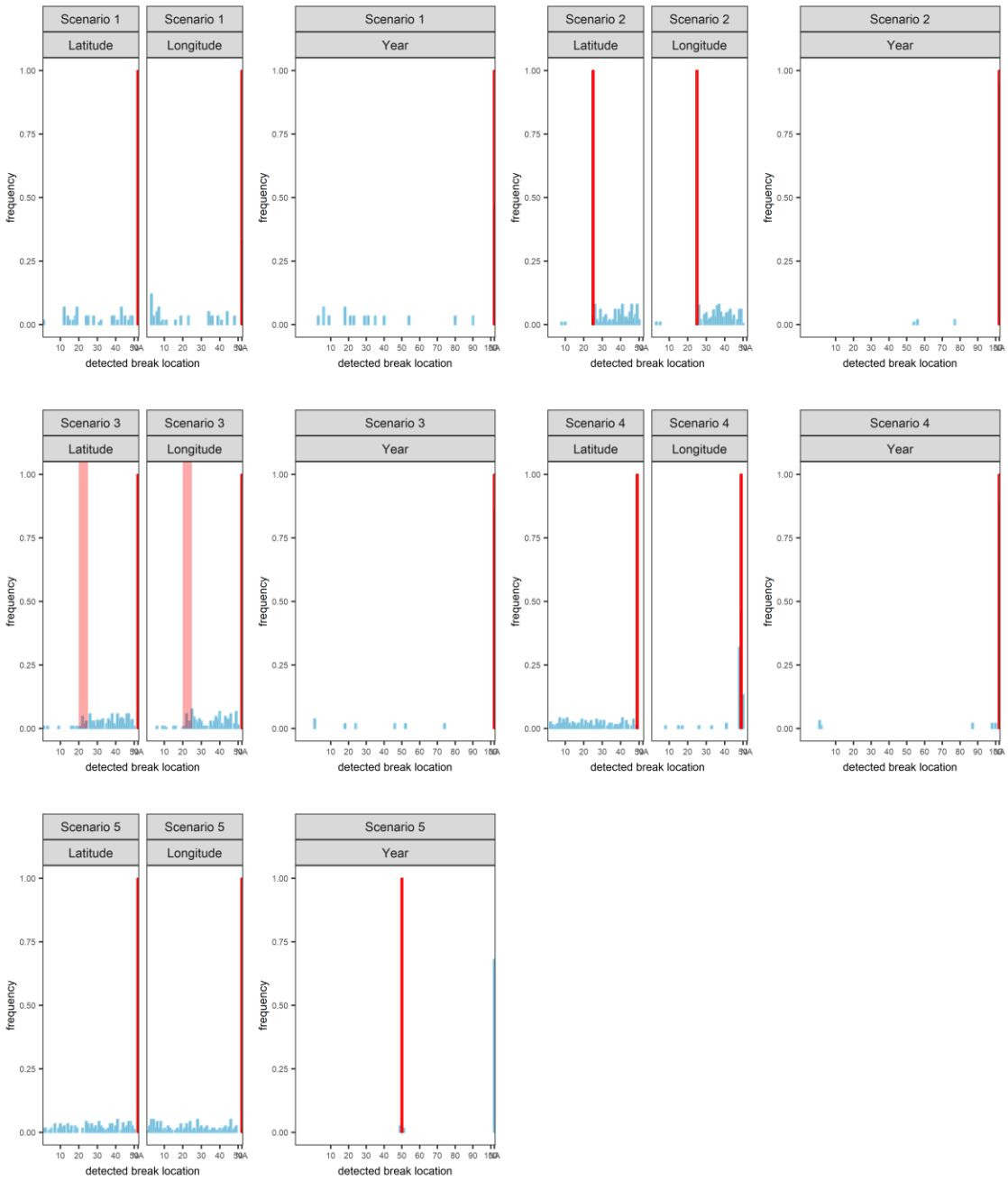


Figure B.4: Histogram of detected breakpoints (blue bars) from the STARS analysis by scenario. Vertical red bars indicate true breakpoints used to generate synthetic populations. For Scenario 3, the synthetic population overlapped between 20 and 25 degrees latitude and longitude.

Figures B.5 to B.14 are identical in form to Figure 2.5 and 2.6 in main text, which presented results for age six female sablefish. These plots contain results for ages four and thirty for males and females, and age-four males.

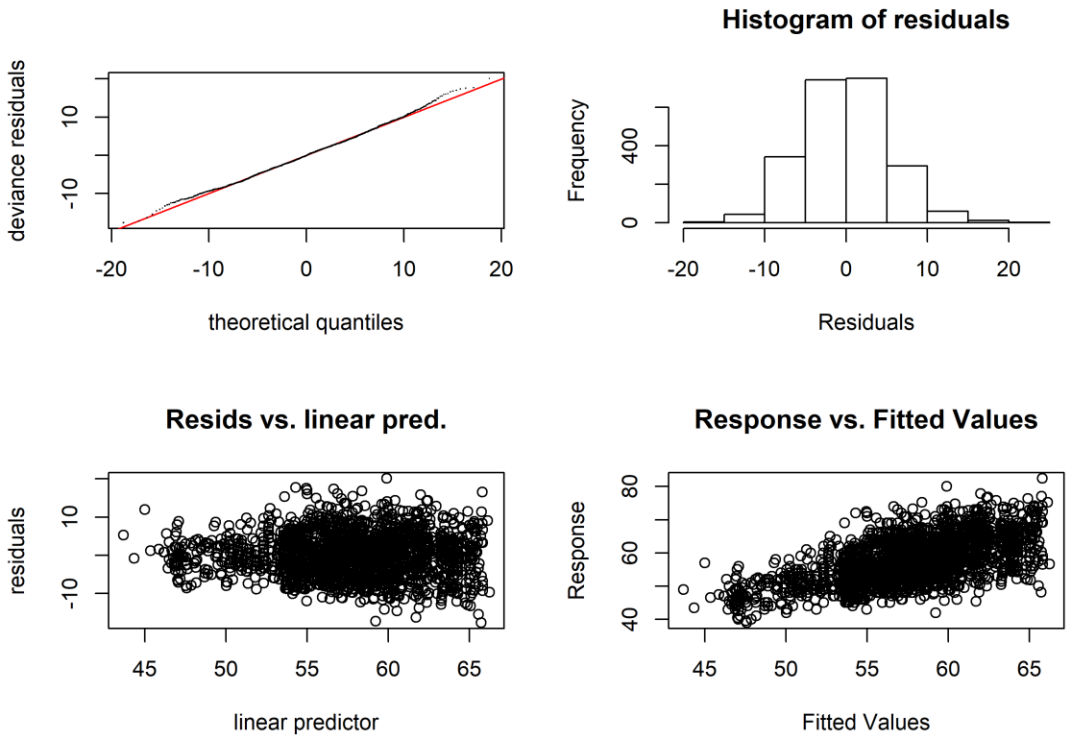


Figure B.5: Diagnostic plots of best-fit GAM model for female age four sablefish. Clockwise from top left: quantile-quantile plot of deviance residuals; histogram of residuals; observed response values (lengths, in cm) vs predicted values, and model-predicted residuals vs linear predictor.

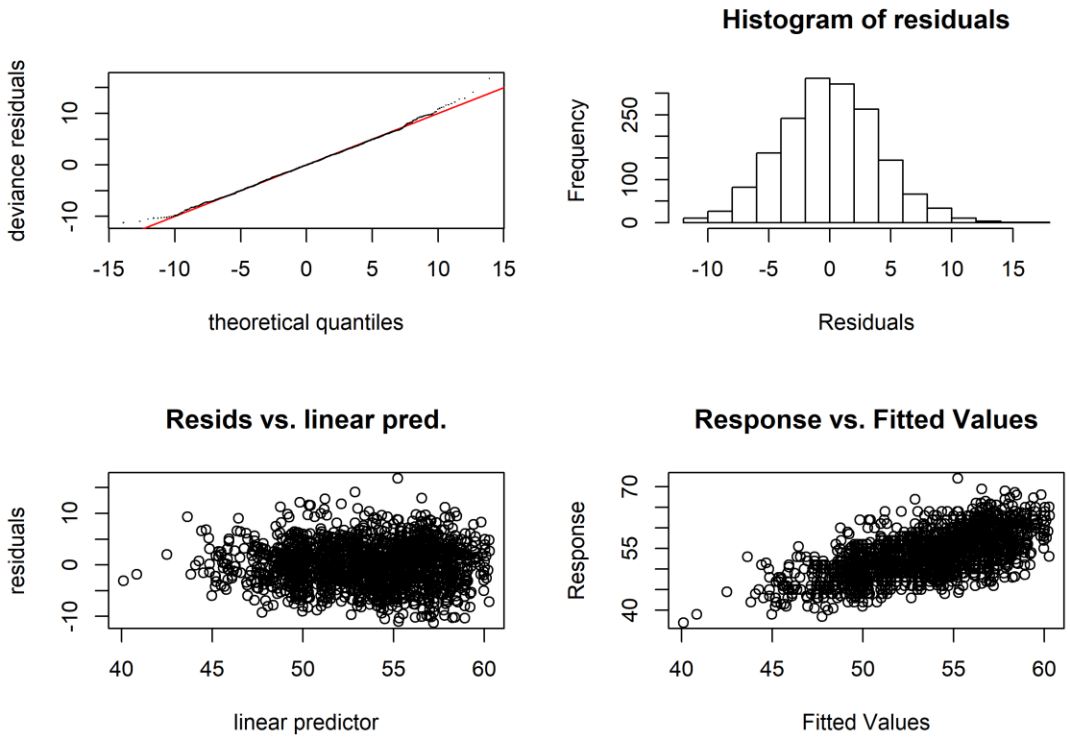


Figure B.6: Diagnostic plots of best-fit GAM model for male age four sablefish. Clockwise from top left: quantile-quantile plot of deviance residuals; histogram of residuals; observed response values (lengths, in cm) vs predicted values, and model-predicted residuals vs linear predictor.

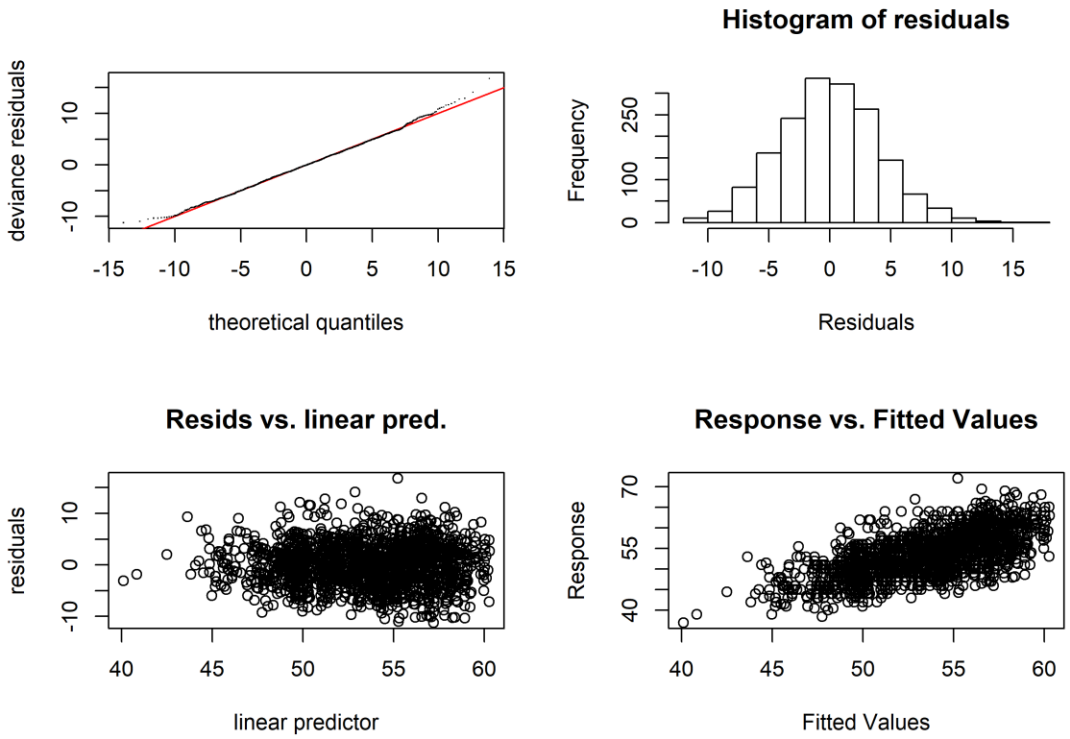


Figure B.7: Plots of smoothers for Year, Latitude, and Longitude, and first derivatives thereof for age-four male sablefish (b,d,f). Red lines indicate latitudes or longitudes that produced the highest first derivative and had a confidence interval that did not include zero. g) map with model-detected breakpoints (red lines).

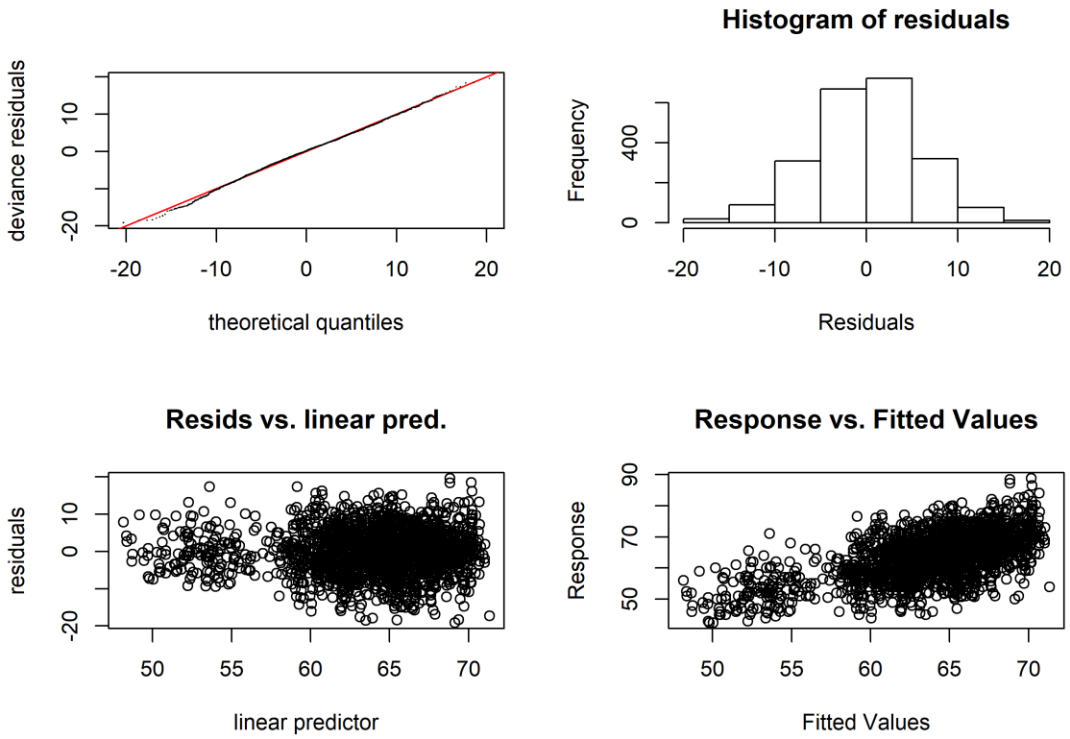


Figure B.8: Diagnostic plots of best-fit GAM model for female age six sablefish. Clockwise from top left: quantile-quantile plot of deviance residuals; histogram of residuals; observed response values (lengths, in cm) vs predicted values, and model-predicted residuals vs linear predictor.

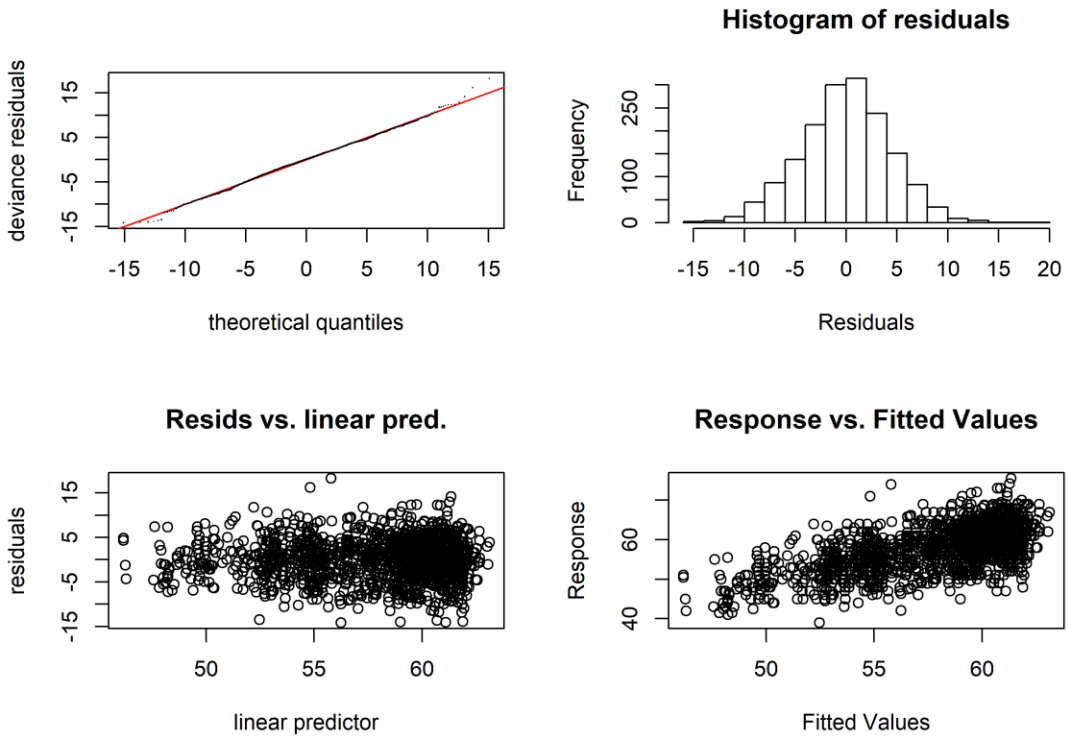


Figure B.9: Diagnostic plots of best-fit GAM model for male age six sablefish. Clockwise from top left: quantile-quantile plot of deviance residuals; histogram of residuals; observed response values (lengths, in cm) vs predicted values, and model-predicted residuals vs linear predictor.

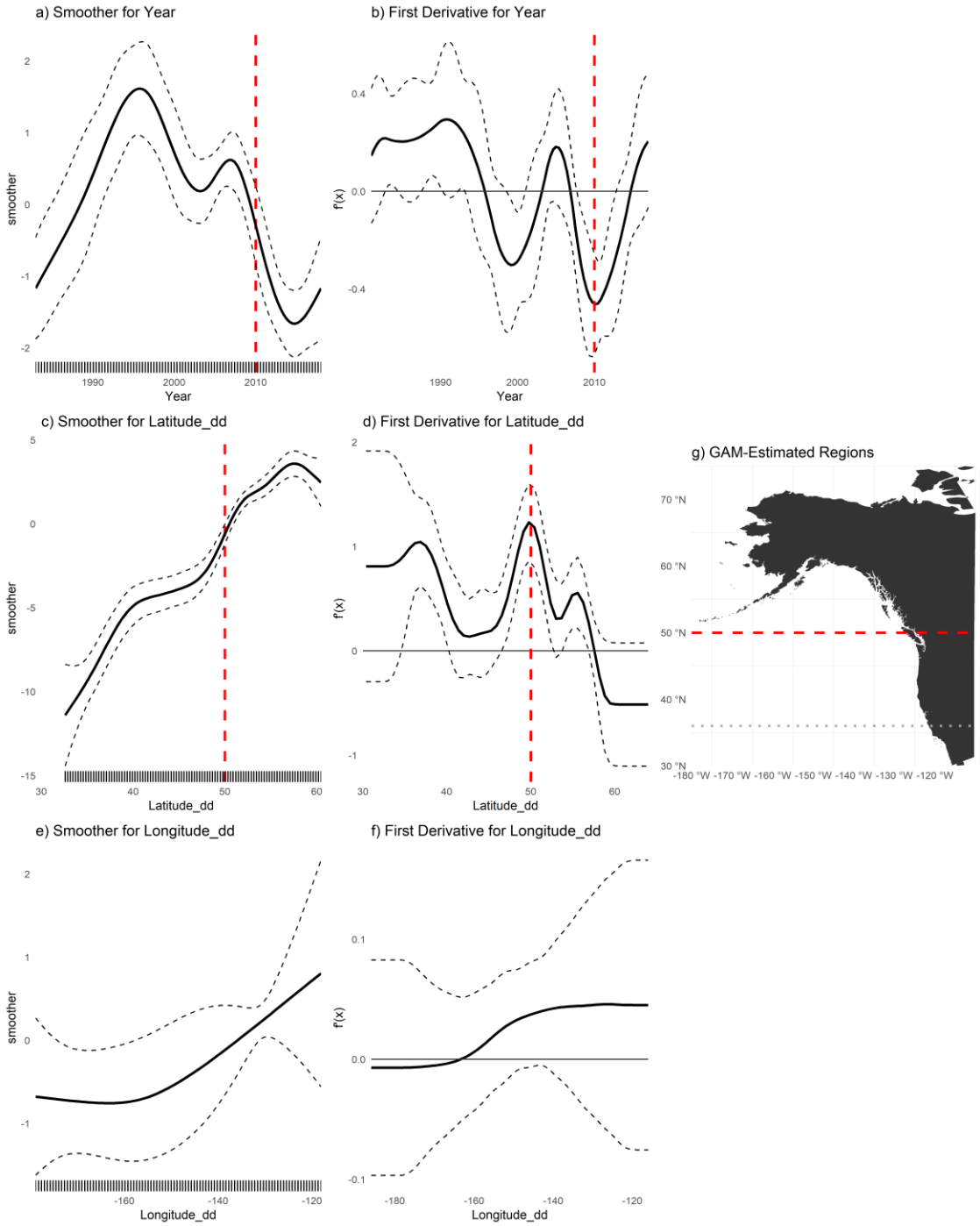


Figure B.10: (a,c,e) Plots of smoothers for Year, Latitude, and Longitude, and first derivatives thereof for male age six sablefish (b,d,f). Red lines indicate latitudes or longitudes that produced the highest first derivative and had a confidence interval that did not include zero.g) map with model-detected breakpoints (red lines).

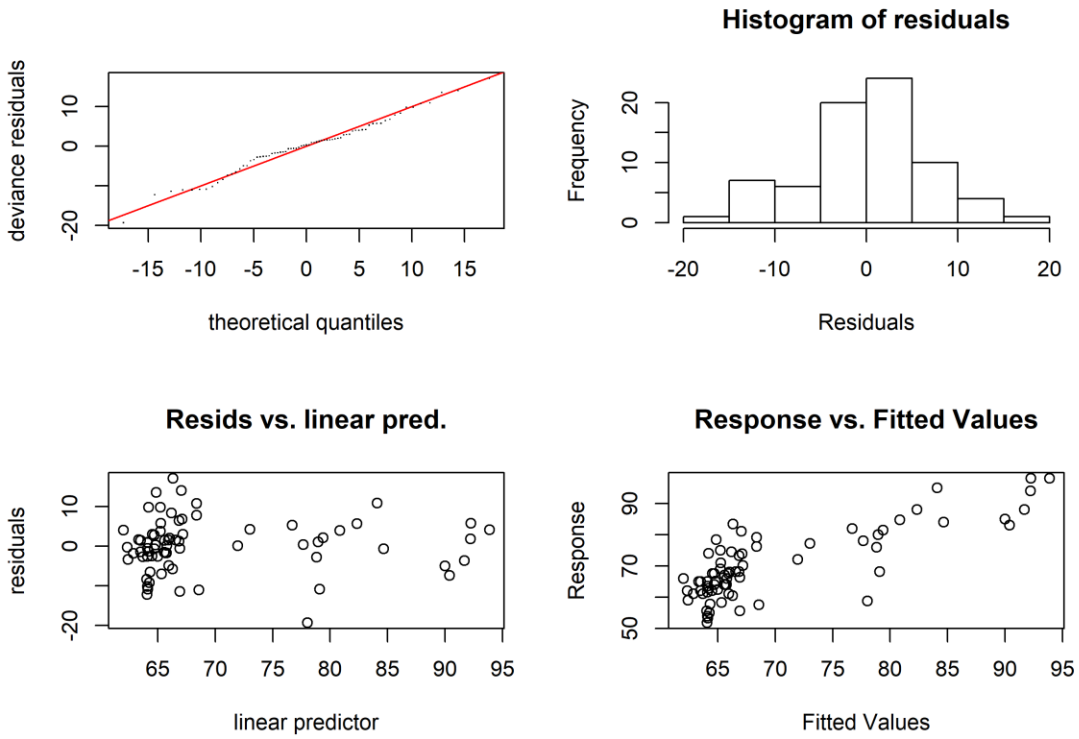


Figure B.11: Diagnostic plots of best-fit GAM model for female age thirty sablefish. Clockwise from top left: quantile-quantile plot of deviance residuals; histogram of residuals; observed response values (lengths, in cm) vs predicted values, and model-predicted residuals vs linear predictor.

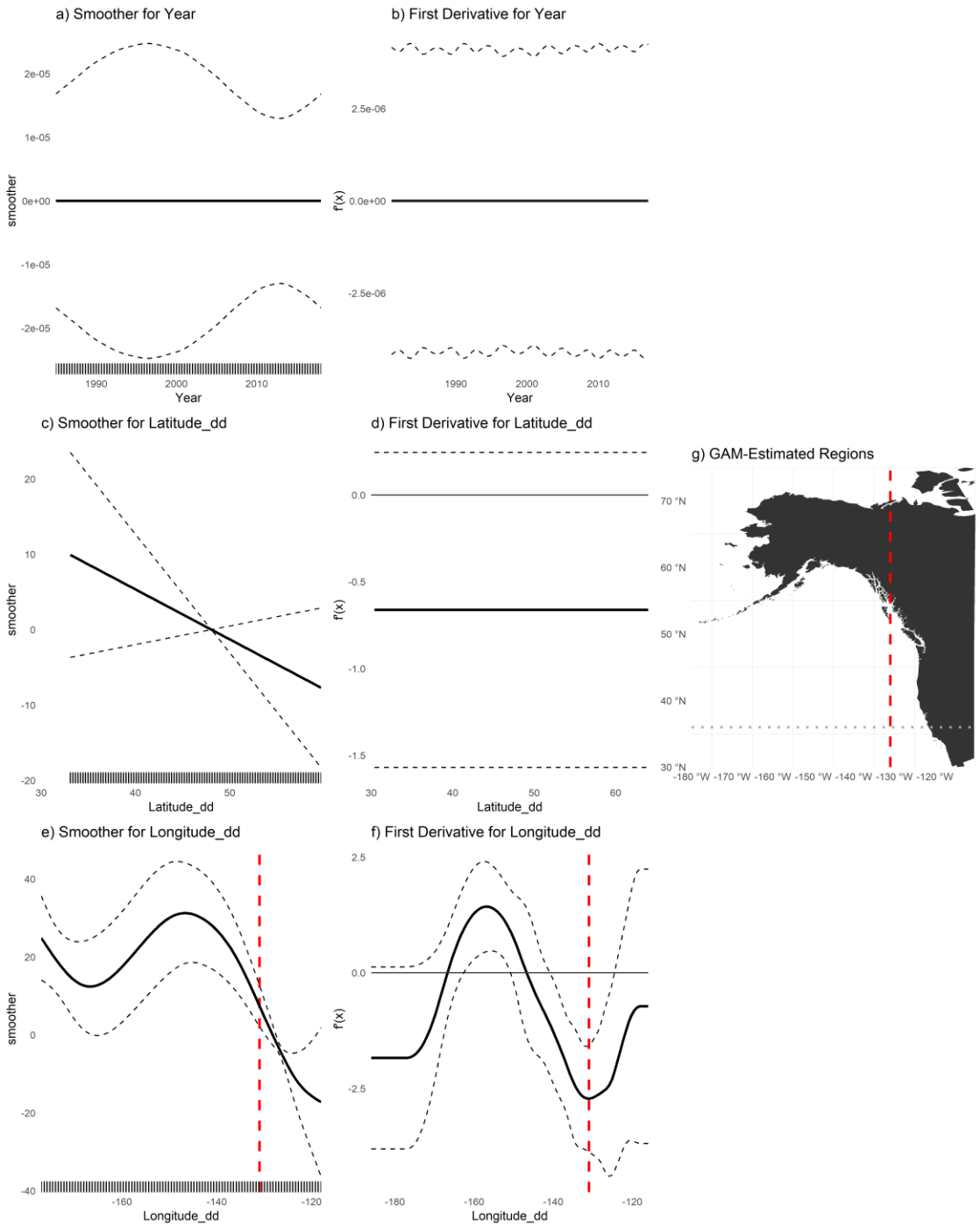


Figure B.12: (a,c,e) Plots of smoothers for Year, Latitude, and Longitude, and first derivatives thereof for female age thirty sablefish (b,d,f). Red lines indicate latitudes or longitudes that produced the highest first derivative and had a confidence interval that did not include zero.g) map with model-detected breakpoints (red lines).

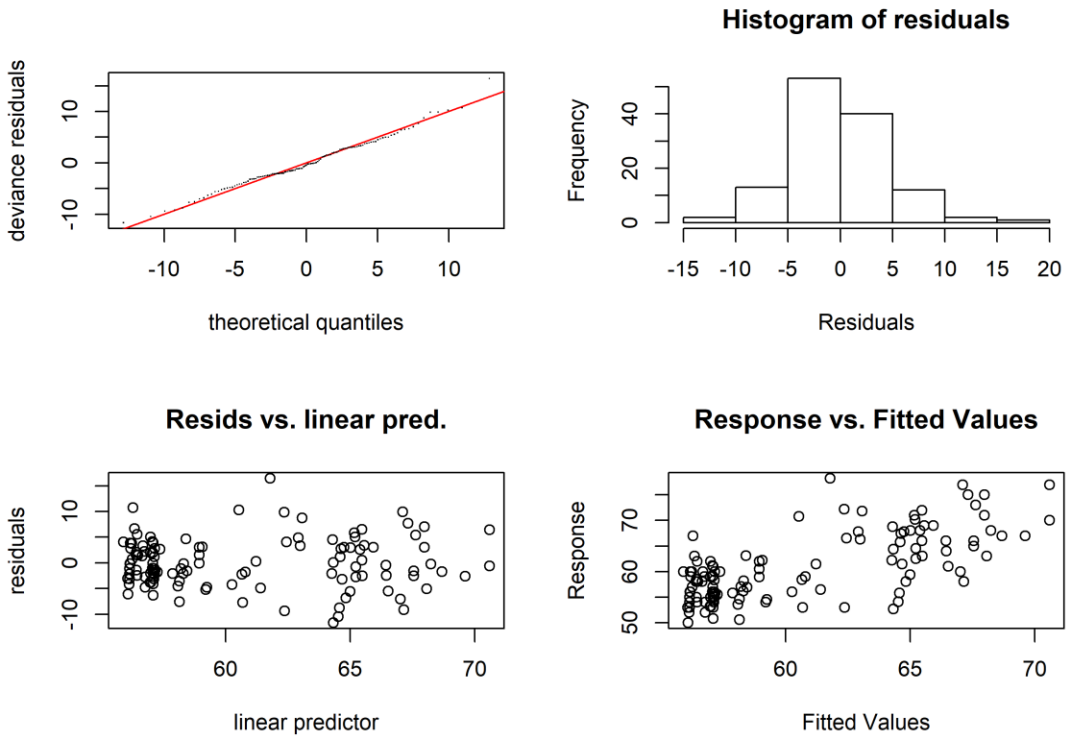


Figure B.13: Diagnostic plots of best-fit GAM model for male age thirty sablefish. Clockwise from top left: quantile-quantile plot of deviance residuals; histogram of residuals; observed response values (lengths, in cm) vs predicted values, and model-predicted residuals vs linear predictor.

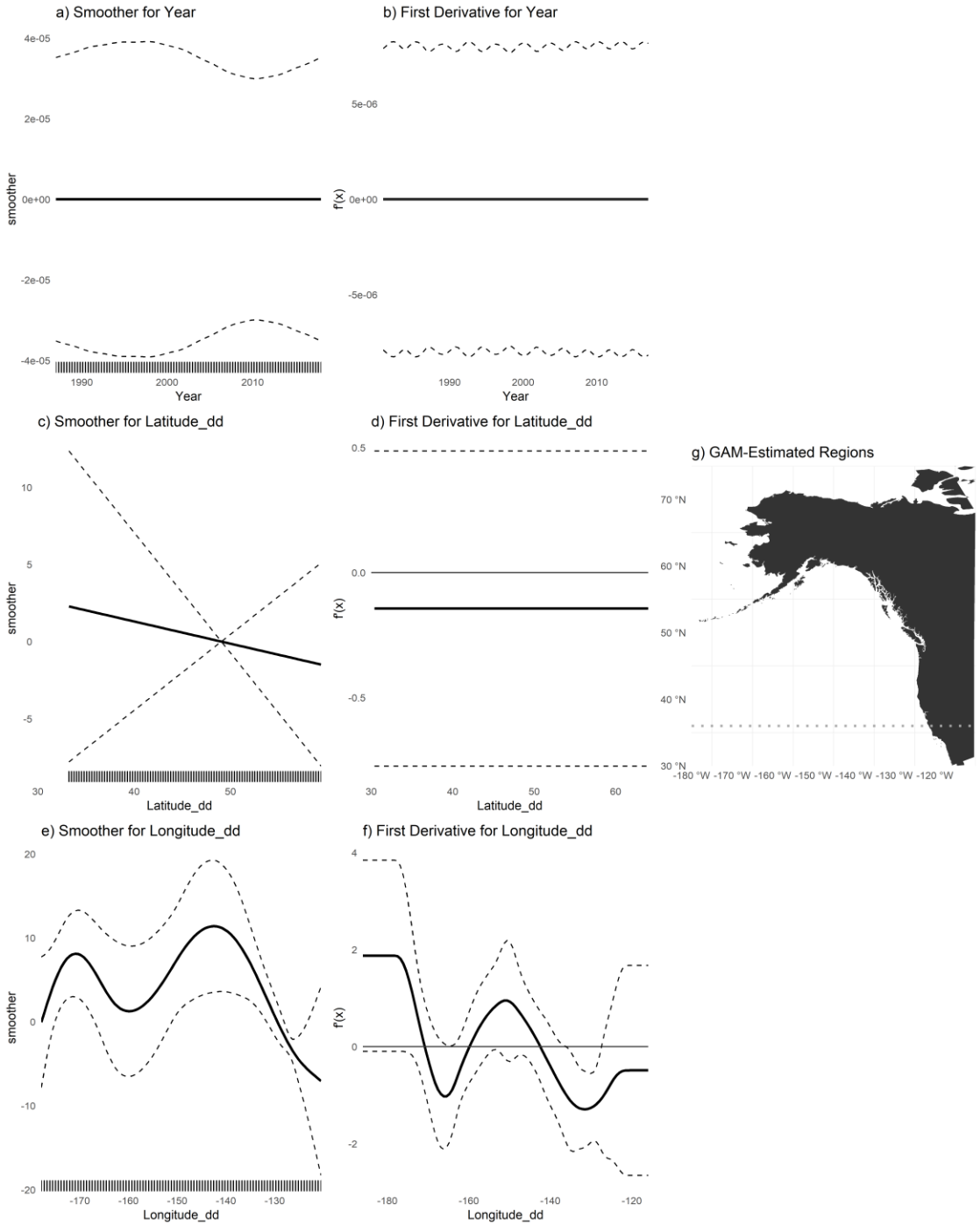


Figure B.14: (a,c,e) Plots of smoothers for Year, Latitude, and Longitude, and first derivatives thereof for male age thirty sablefish (b,d,f). Red lines indicate latitudes or longitudes that produced the highest first derivative and had a confidence interval that did not include zero. g) map with model-detected breakpoints (red lines).

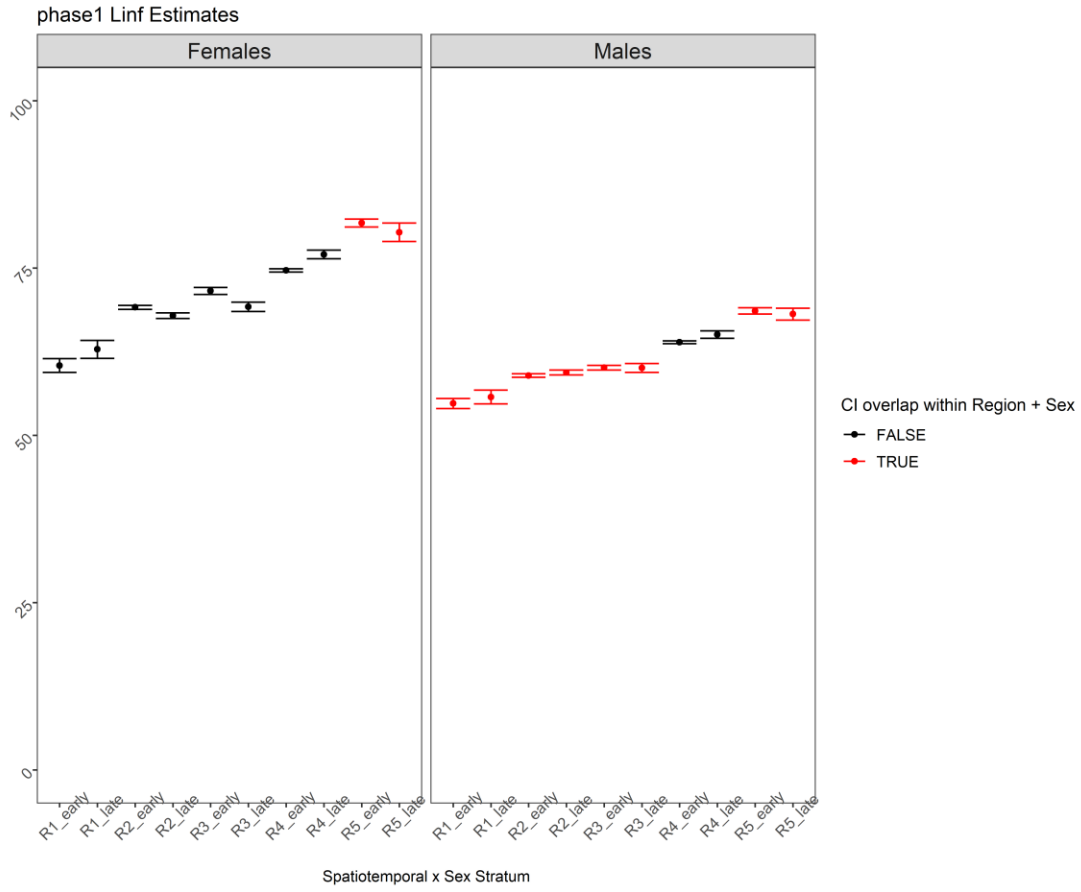


Figure B.15: L_∞ estimates for the fully stratified, 5-region, 2-period (during and after 2010, and before) and 2-sex model. Bars represent 95% confidence intervals. Strata from the same spatial region and sex that shared overlapping ranges for L_8 are colored in red and early and late periods were combined within their respective regions and sexes for the subsequent analysis.

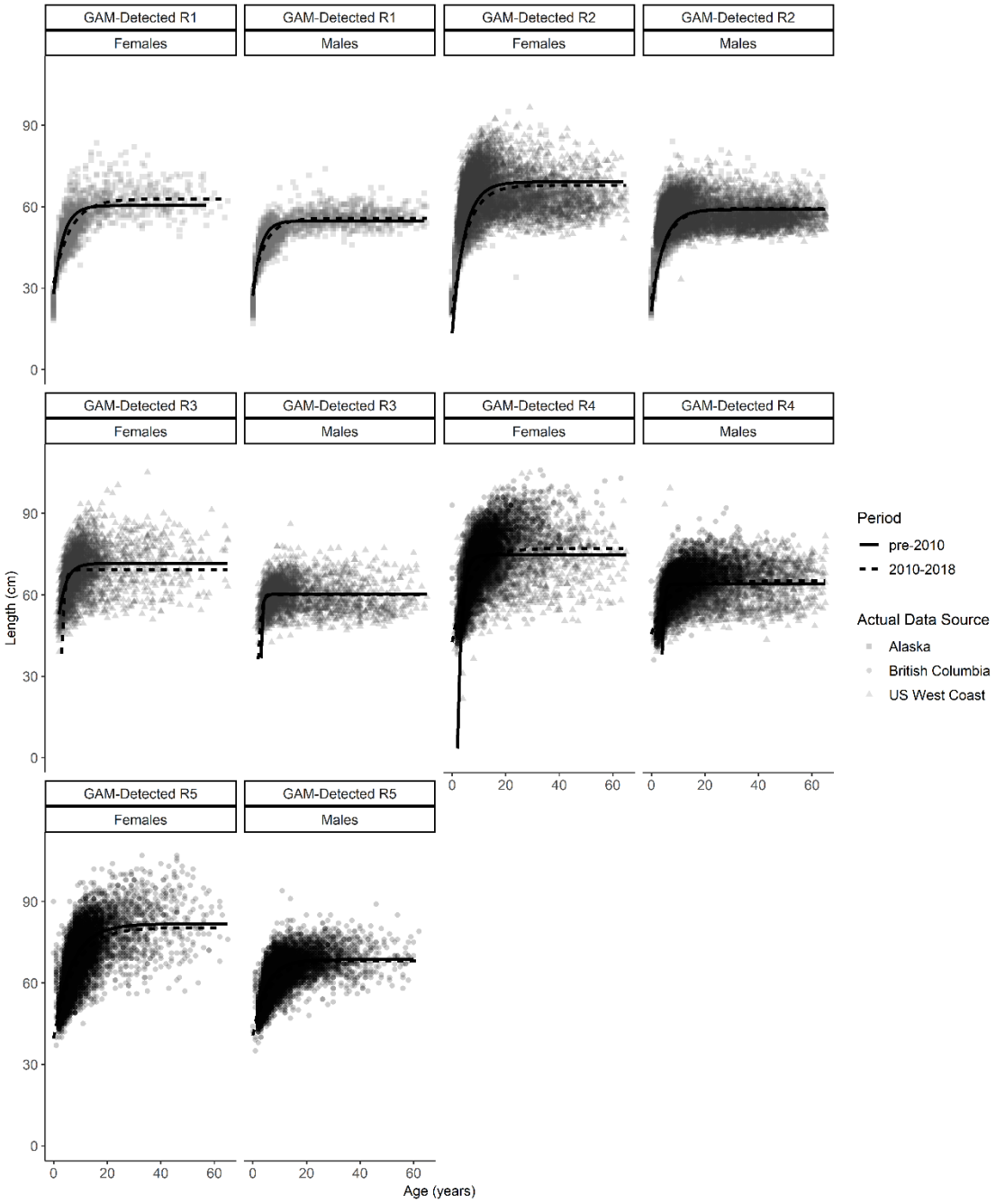


Figure B.16: Fits of von Bertalanffy growth function (black lines) to data for Phase 1 spatio-temporal aggregation. Points are raw survey data colored by their source. Line types denote whether fit is for early, late or pooled time period.

Appendix C. Supplementary Information for Chapter 3

C.1 Appendix C.1 Movement Rates at Age

C.1.1 Background

Rogers et al. (In Prep) used a hidden Markov movement model and forty years of tagging data to quantify sablefish movement rates among management regions in the northeast Pacific Ocean. The data inputs comprised tag release and recovery observations in the northeast Pacific from 1979 to 2018. Tagged sablefish were released from National Marine Fisheries Service (NMFS) longline surveys along the continental slopes of Alaska and the California Current (Rutecki et al., 2016), and from Fisheries and Oceans Canada (DFO) sablefish surveys along the continental slope and in four mainland inlets of British Columbia (Lacko et al., 2020; Wyeth, 2005). Tagged sablefish were also released by the State of Oregon during 1996–2004 (Sogard and Berkeley, 2017). Sablefish tags were recovered via surveys or commercial fishery catches.

The sub-area boundaries represented in the movement model were the same as the six sub-areas in the OM (Figure 3.1). Tag reporting rates were provided for the corresponding regions for Alaska and Canada; Alaskan reporting rates were used for the US West Coast as specific estimates of reporting rate were unavailable for that region. Fishing mortality rates for each management region were set to the values estimated in the recent regional stock assessments. The original analysis produced pooled-sex movement rates for fish than will become larger than 400 mm (Figure C.1), approximating the transition from immature to mature sablefish. Fish smaller than 400 mm are assumed to not move.

C.1.2 Converting to movement rates-at-age given spatiotemporal variability in growth

The OM requires a matrix **X** that specifies the proportion of fish of age a and sex γ in sub-area i moving to sub-area j. Growth and movement parameters are pre-specified within the OM (not estimated during the conditioning process), although growth varies spatially, among sexes, and through time. Thus, we first establish the expected distribution of length-at-age ψ for each year, sex and sub-area, noting that 81cm is the plus group length bin (Figure C.2).

$$\psi^{i}_{y\gamma al} = \begin{cases} \Phi\left(1, \tilde{l}^{i}_{y\gamma a}, \sigma^{i}_{G\,y\gamma}\right) & if\ l\ =\ 0cm \\ \Phi\left(l+1, \tilde{l}^{i}_{y\gamma a}, \sigma^{i}_{G\,y\gamma}\right) - \Phi\left(l, \tilde{l}^{i}_{y\gamma a}, \sigma^{i}_{G\,y\gamma}\right) & if\ 0\ <\ l\ <\ 81cm \\ 1 - \Phi\left(l, \tilde{l}^{i}_{y\gamma a}, \sigma^{i}_{G\,y\gamma}\right) & if\ l\ =\ 81cm \end{cases} \quad (C.1.1)$$

where Φ is the cumulative normal density function; $\tilde{l}^{i}_{y\gamma a}$ is deterministic length-at-age obtained by applying the von Bertalanffy function to growth parameters applicable to sub-area i; and $\sigma^{i}_{G\,y\gamma}$ is the variation in growth for sub-area i; σ^{i}_{G} are estimates of the variance of length-at-age in subarea i, as in *Chapter 2*.

Size-based movement estimates $\widetilde{\mathbf{X}}$ are converted into fixed movement rates by age and sex via the dot product of Ψ and $\widetilde{\mathbf{X}}$ ($\widetilde{\mathbf{X}}$ is the identity matrix for fish smaller than 400 mm).

$$\mathbf{X}^{ij}_{y\gamma a} = \sum_{l} \Psi^{i}_{y\gamma al} \tilde{X}^{ij}_{l} \quad (C.1.2)$$

where $\tilde{X}^{ij}_{l}$ is the estimated proportion of fish in length-class l and sub-area i moving to sub-area j, from the original movement analysis; and $\Psi^{i}_{y\gamma al}$ is length-at-age distribution defined above.

To reduce the complexity of the matrix, and avoid movement of fewer than a small number of fish between sub-areas, values of $\mathbf{X}^{ij}_{y\gamma a}$ less than 0.01 were rounded to zero. Final values were rescaled to ensure no aberrant loss or gain of fish (i.e., the rows of $\widetilde{\mathbf{X}}$ sum to 1).

C.1.3 Simplifying the movement matrix

After conditioning the operating model, an evaluation was conducted as to whether there were ages below which movement was nonexistent or negligible, and whether the effect of temporal variation in growth upon the most probable age-at-size through time could either be ignored or accounted for in a simpler manner. For the first analysis, it was determined that there was no movement between any sub-areas for any years nor sexes for animals below age 1 (as in no sub-areas have age-1 individuals of either sex that have reached 400mm). Therefore, $\mathbf{X}$ for age 0 is the identity matrix.

We wished to reduce computational cost and avoid introducing a complex time-varying component to movement that could confound with recruitment. The percent difference between

the movement estimates for each year/sex/sub-area/age combination and the mean movement estimates for the sex/sub-area/age combinations across all years was examined because the OM includes an estimated, time-varying process that allocates individuals to sub-areas (the down-scaling of recruitment deviations, τ_y^{ik}). This analysis showed that less than 1.2% of all unique movement rates differed from the all-year mean by greater than 5%, and none differed by more than 12.5%. Moreover, 50 out of 60 model years had no single movement rate that differed from the all-year sex/sub-area/age means by more than 5%. Therefore, mean movement rates across all years for unique sub-area-sex-age combinations were used as pre-specified values throughout the OM period. The final input movement matrix **X** is defined as:

$$\mathbf{X}_{\gamma a}^{ij} = \frac{\sum_y \sum_l \psi_{y\gamma al}^i \tilde{X}_l^{ij}}{60} \qquad (C.1.3)$$

The final movement matrix (Figure 3.6) retains the qualitative behavior of the original tag-recapture analysis, in that British Columbia experiences the lowest self-seeding rates and transports a large fraction of individuals southward to the California Current, and a slightly lesser fraction of individuals northward to the Gulf of Alaska. These dynamics are consistent with the bifurcation of the North Pacific Current, a major oceanographic feature that directs water masses north and south of approximately 50°N. As expected, the variation in growth rates across subregions results in larger fish at earlier ages in more northerly regions, so there is less variability in export rates among ages for fish from sub-areas A3 and A4, especially for females. Conversely, the large standard deviation in expected length-at-age in the US West Coast/Southern British Columbia stocks (sub-areas C2 and B2) leads to export rates that change more markedly among ages, increasing from 42% of age-5 males leaving sub-area C2 to 52% of males females leaving sub-area C2 by age 10.

C.1.4 Movement sensitivity analyses

The exercise described above was repeated for the lower and upper 95% confidence interval for each length class-based movement rate estimated from the original tag-recapture model (i.e. $\tilde{X}_l^{ij} \pm 1.96 \times SE\left(\tilde{X}_l^{ij}\right)$) to sensitivity to the uncertainty in movement rates. These represent varying degrees of self-seeding, with the "high" scenario representing greater exchange of individuals between adjacent areas, and the "low" scenario representing more retention within each sub-area.

It was assumed that these low (Figure C.3) and high (Figure C.4) movement rates were fixed through time, and that age-0 fish do not move in the sensitivity analyses.

C.1.5 Figures

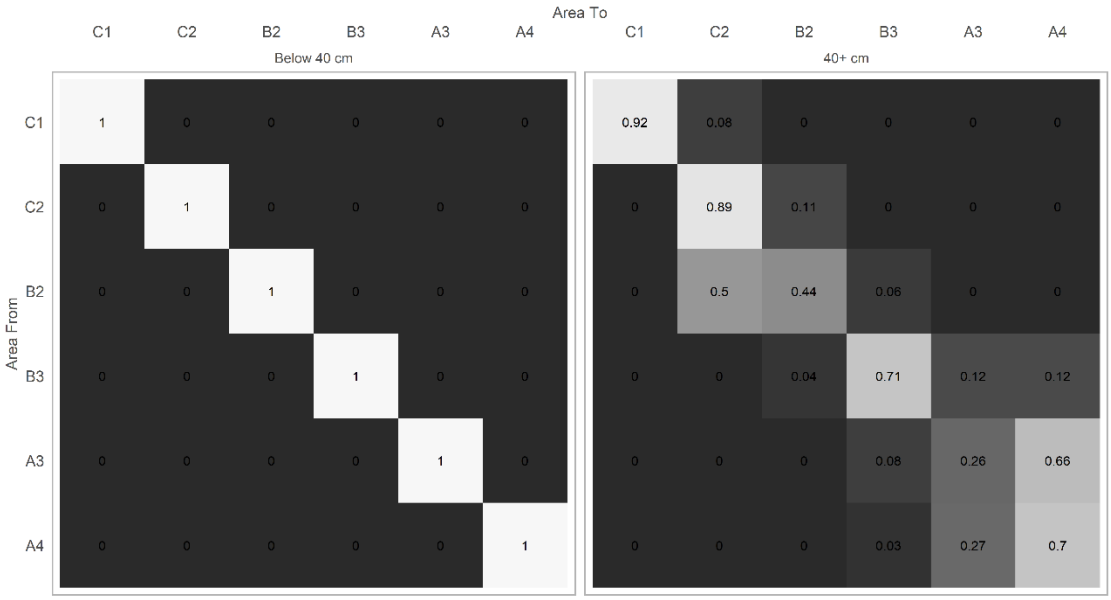


Figure C.1: Input movement rates among sub-areas determined from the tag-recapture analysis. These values are assumed to apply to both sexes for all years to animals within the length-classes indicated at the top of each panel.

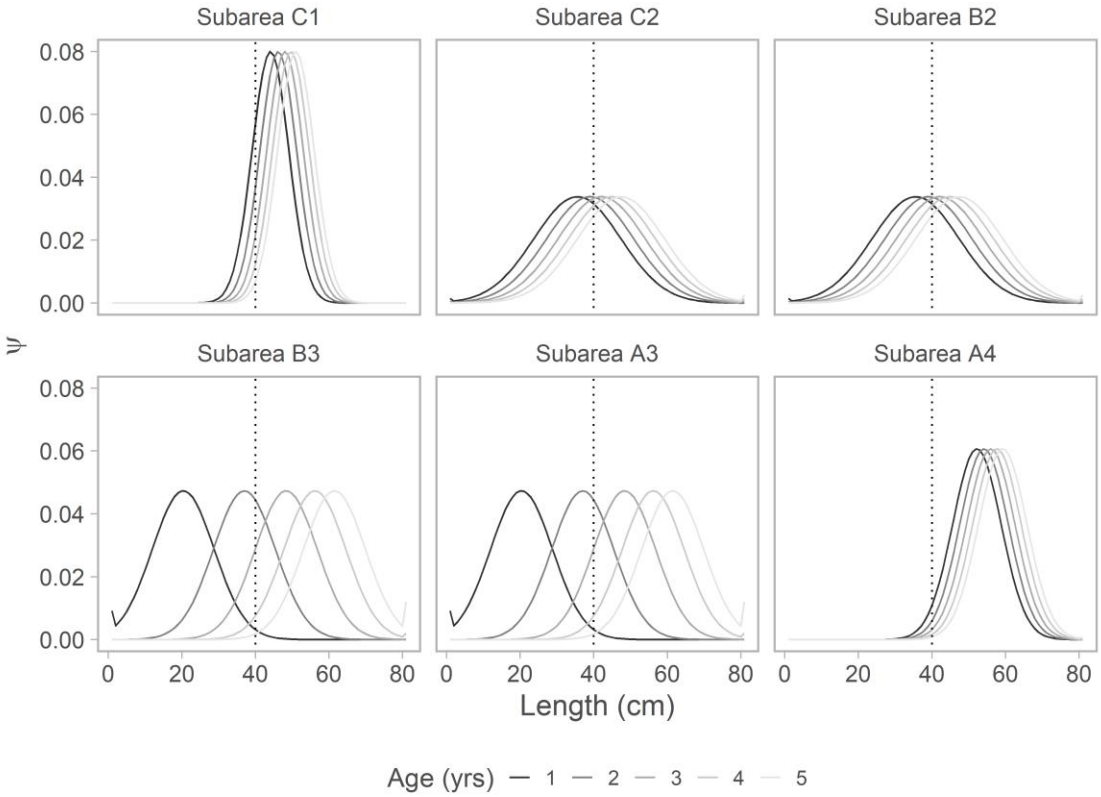


Figure C.2: Length-at-age distributions for female sablefish all sub-areas in year 2000 for a subset of age classes. Vertical dashed lines indicate 40 cm, the lower limit of the size class used in the original movement analysis. The probability of being length 40 cm is maximized at either age one or age two depending on the growth regime.

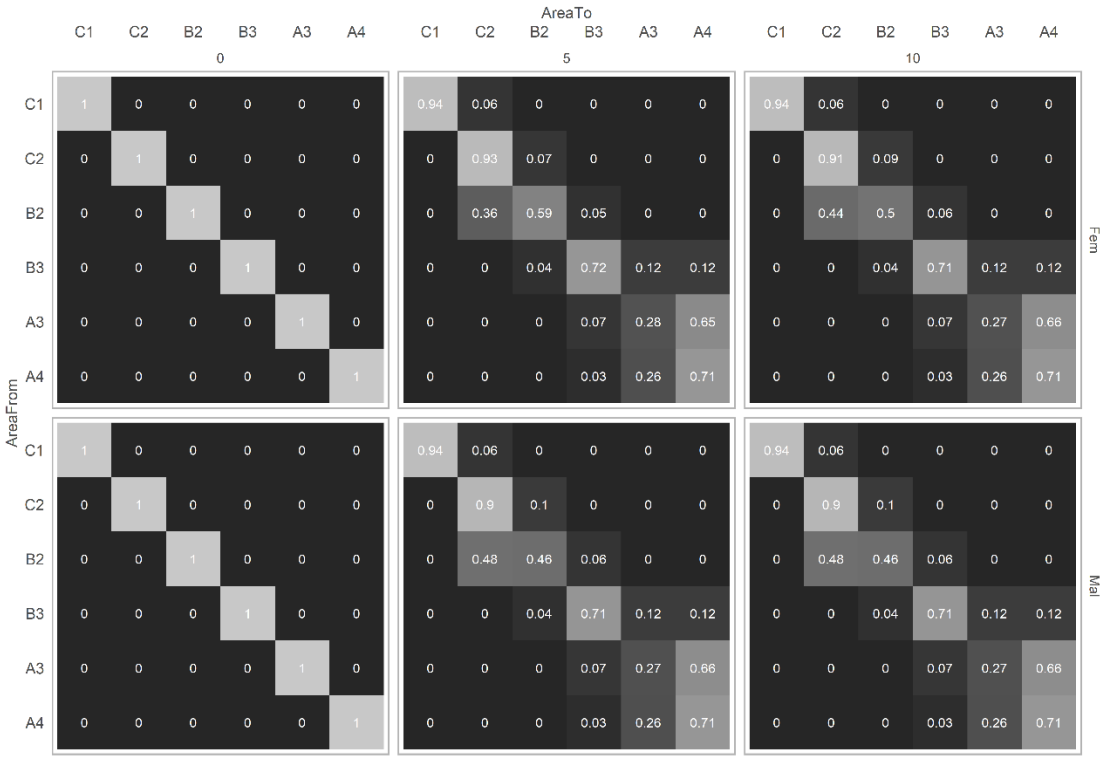


Figure C.3: Input for the "low" movement scenario. These represent the lower 95% confidence interval corresponding to the size-based movement rates from the tag-recapture analysis, recalculated to represent movement rates at age. These values are assumed to apply for all years for animals within the age classes indicated at the top of each panel.

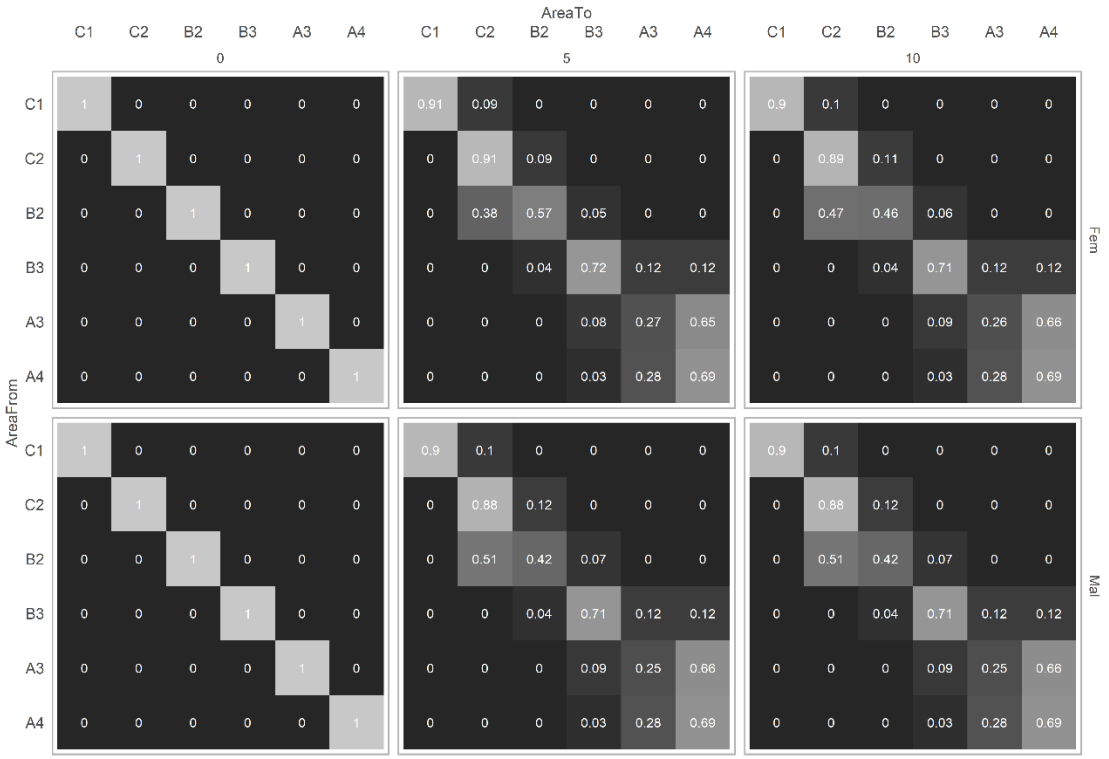


Figure C.4: Input for the "high" movement scenario. These represent the upper 95% confidence interval of the size-based movement rates from the tag-recapture analysis, recalculated to represent movement rates at age. These values are assumed to apply for all years for animals within the age classes indicated at the top of each panel.

C.2 Spatiotemporal trends in the relative abundance of northeast Pacific Sablefish

C.2.1 Introduction

Stock assessments typically require the use of fish abundance surveys, which are often conducted by government agencies at regional scales relevant to management. For sablefish, a highly valuable groundfish species in the northeast Pacific, management is structured by three broad jurisdictions that are sampled by multiple, non-overlapping surveys, which differ in survey design and gear types, sampling rates and catchability. Due to synchrony in sablefish abundance trends from both surveys and assessment estimates across all three regions, and movement of sablefish among regions, there is interest in developing a northeast Pacific-wide assessment model to support a management strategy evaluation for sablefish. This appendix presents the construction of four time-series of relative abundance for sablefish by applying a spatiotemporal standardization approach to data from the pre-existing fishery -independent surveys conducted by the United States government, which overlap in time but not space. The indices suggest that biomass is greatest in Alaska and presents regional trends similar to those used in separate assessments. While relative biomass has declined from a historical high over the past several decades throughout the northeast Pacific, the model suggests an increase in abundance throughout all regions in recent years. Parameter estimates indicate variation in both encounter probabilities and capture rates among regions, the latter of which is likely a result of selectivity differences among surveys.

C.2.1.1 Indices of abundance in stock assessment

Indices of abundance, ideally from fishery-independent surveys, are often used in fishery stock assessments as information on the relative size of the fish population. Such indices are then used within an assessment framework to estimate quantities such as current abundance that are of interest to decision makers. Fished stocks with adequate resources and managerial capacity, such as the valuable groundfish of the northeast Pacific, are monitored via government-sponsored surveys, or joint government-industry surveys as is the case in British Columbia, and in Alaska for species other than sablefish. These surveys typically occur on at least a biennial basis and use statistical survey designs to produce representative sampling of age and size (length or weight) compositions, and relative indices of abundance for the stocks surveyed. Stock assessment scientists implement a variety of methods to ‘standardize’ raw observations, with the intent of reducing erroneous signals caused by survey implementation (i.e., time of year, vessel) while

retaining “true” fluctuations in population abundance (Maunder, 2004). Occasionally, a given management region may have more than one survey that overlaps in space and/or time, and standardization of the resulting survey data sets are typically modeled separately. For example, the 2019 benchmark assessment for sablefish off the West Coast of the U.S. conducted a spatiotemporal standardization of the recent groundfish bottom trawl survey, years 2003-2018, but separately conducted that same standardization for the Triennial survey that overlapped geographically (Haltuch et al., 2019).

C.2.1.2 Spatial concerns

Many commercially valuable stocks, including sablefish, are broadly distributed and exhibit spatial demographic patterns that can be temporally inconsistent with the static management boundaries defining current stock assessments. This has the potential to lead to biased estimates of management quantities and reference points (Goethel and Berger, 2016). Here, “region” pertains to the current management regions within the range at which stock assessments are currently conducted for sablefish at varying regularity: the U.S. California Current, coastal British Columbia, and the entirety of Alaska. Concurrent declines in sablefish abundance have led to a concerted effort to develop a range-wide population model for this species, which would benefit from the development of an index of abundance representative of the entire northeast Pacific. An obvious hurdle in this effort is differences in sampling efficiency, which varies across gear (e.g., trawl, longline hook or trap), size/age, and operational protocols. These differences can result in dissimilar estimated proportions of local biomass captured by a given unit of survey effort (i.e., catchability). In the absence of a range-wide index, assessment scientists could attempt to directly estimate separate catchabilities for individual surveys within a stock assessment model. However, concurrently estimating selectivity and catchability for multiple areas (surveys) can prove unwieldy, leading to model instability if differences among regions are high (Punt, 2019b).

The objective of this work was to develop indices of relative abundance for use in a spatially-stratified model of northeast Pacific sablefish, integrating information from two of three management regions. The Vectorized autoregressive spatio-temporal (VAST, Thorson (2019)) standardization framework (detailed in Section C.2.2) was applied to the results of fishery-independent surveys of sablefish from US waters within the northeast Pacific. The VAST-derived index of relative abundance can help to address two major research questions for this population:

1) are the synchronous trends in observed survey abundance indicative of range-wide population trends, and 2) does the implementation of a spatially-explicit range-wide management strategy evaluation using this index suggest there are better ways to manage sablefish fisheries in each region? The latter question would necessarily be dealt with in a management strategy evaluation (MSE) framework (Chapter 4), which motivates the development of these spatially-explicit relative indices of sablefish abundance.

C.2.2 Methods

C.2.2.1 VAST

The spatio-temporal analysis is implemented using the VAST modeling software package (Thorson, 2019) to predict biomass density ($kg/km2$) for individual combinations of location (s) and year (t) within a given region that has been surveyed. One advantage of VAST is that it enables unsampled locations and/or years to be informed by locations and years where there are data, thus creating a smooth estimate of fish abundance (density) and extent (spatial distribution) through time. The model calculates the probability of the observed biomass *b* at each unique location *i* using two linear predictors that represent encounter probability $r_1(i)$ and positive catch rate $r_2(i)$, whereby:

$$P(B = b_i) = \begin{cases} 1 - r_1(i) & B = 0 \\ r1(i) = b_i | r_2(i), \sigma_m^2(c_i) & B > 0 \end{cases} \quad (C.2.1)$$

where b_i is observed biomass, and $\sigma_m^2(c_i)$ is the residual variance in positive catch rates. The model also enables a random effect to be specified for individual survey vessels for both linear predictor components, which was possible in the analysis for all surveys. The model specified a spatiotemporal effect for both positive catch rates and encounter probabilities with 500 spatial knots. This was the minimum number of knots for which convergence was obtained. Convergence of the model was evaluated by the maximum gradient obtained and whether the Hessian matrix was invertible. Further details of the VAST model implementation is available in Thorson (2019). A principal assumption of the VAST approach is that true spatial variation in fish abundance is independent of survey sampling boundaries. Long-term tag-recapture experiments of sablefish throughout the northeast Pacific have demonstrated that sablefish move readily across the entire range, with an average point-to-point migration distance of 191 km (Hanselman et al., 2014). Additional genetic evidence supports the notion that northeast Pacific sablefish are genetically

homogenous, with no major detectable fine scale population structure that would suggest sub-stocks at the regional level (Jasonowicz et al., 2016)[5].

C.2.2.2 Data sources and pre-processing

The VAST standardization includes fishery-independent survey data provided by the National Oceanic and Atmospheric Administration (NOAA). The three survey datasets, with observations in kilograms, span two regions, with varied annual coverage from 1980 to 2020[6], and a depth range of 10 to 1,500m (Table C.1; Figures C.5 and C.6). These survey datasets are all currently used in stock assessments for the respective regions. The spatial extent of the surveys includes the California Current (CC), the Gulf of Alaska (GOA) and the Eastern Bering Sea (Figure C.5). A subset of surveys is time-blocked where changes in gear or sampling methodology are likely to have led to intra-survey changes in catchability. The domestic longline survey was filtered to only include deployed at depths less than 500 m and ignored the survey data for 1984 and 1987, when the surveys were conducted in cooperation with Japanese fleets and had gear-area combinations distinct from later years, in addition to 30-minute tows. Finally, the Triennial survey data were split before and after 1995 to reflect changes in survey timing (Table 3.6).

The VAST method enables input of a value proportional to effort (in this case, number of traps times minutes of soak time) and estimates two catchability coefficients (one for each linear predictor) as fixed effects. Under plausible generative processes, a linear transformation of units in the response (arising from mis-specifying area) are captured by a transformation of catchability coefficients. Despite this, a requirement for implementation of the encounter-rate linear predictor across all surveys is that input data contain "zero hauls", records made (or backfilled) of locations and years where sets were cast but no sablefish caught.

The VAST method estimates relative catchability parameters for each of the surveys, and aggregates biomass outputs to regions, which in this case were set to correspond to the sub-areas represented in the operating model; there are two such areas in each management region (Figure

[5]Although there is not genomic evidence for stock structure, there is clear evidence for environmentally-driven gradients in sablefish growth (Kapur et al., 2020) and the existence of lagged recruitment pulses among management regions (Fenske et al., 2019). Therefore, the operating model does not assume that sablefish reproduction is fully panmictic, as density dependence likely operates at a sub-regional scale due to differences in size-at-age, exploitation and oceanographic features. The implications of panmixia are explored in Chapter 4.

[6] The operating model only uses standardized data through 2019, but indices were standardized through 2020 where data availability permitted.

3.1). The Gulf of Alaska is stratified to the east and west of 145°W, while the California Current is stratified north and south of 36°N. Due to computational constraints, and the risk of confidence intervals crossing zero when interpolating across large geographic space, the VAST standardization was conducted separately for the California Current and Alaska. These analyses were conducted using `VAST` version `3.8.2` and the `FishStatsUtils` package version `2.10.2`. The model treats non-encounter years as zero, which precluded interpolation of biomass estimates for years where there was no survey. This is preferable as both regional models treat temporal covariates as fixed effects, meaning that the intercepts of the delta model are unidentifiable during years with no data, and thus should not influence nor be included in the final analysis. The estimates of surveyed biomass from each time series are divided by 1000 to match the scale of the input data (in tonns) and partitioned into the relevant sub-areas of the OM: C1 and C2 for US West Coast indices, and A3 and A4 for Alaskan indices (Figure 3.1).

C.2.3 Results

The purpose of this exercise was to estimate indices of abundance at the spatial strata appropriate to an operating model for northeast Pacific sablefish, and to investigate whether VAST-based indices of relative abundance were correlated with the existing independent regional indices used in current assessments. The final time-series was not expected to be at the exact scale of the previous indices not standardized with VAST (i.e., a summation of the regions' currently-used indices for each year). However, it was anticipated that the range-wide index was able to mimic the regional trends. However, because the 2021 West Coast sablefish assessment used VAST to standardize the West Coast Groundfish Bottom Trawl Survey (WCGBTS) data (2003 – 2019) and Triennial data (1980-2004, every three years) the index was anticipated to roughly match both the scale and trend from that assessment for those years, which it did (Figure C.7). The indices for sub-areas C1 and C2 suggest a decline from 1980-2000, an overall high point just after 2000, followed again by a steep decline. The index for sub-area C2 increases dramatically during the final 10 years of the time series, while the index for sub-area C1 remains stable (Figure C.7).

The VAST method estimates two indices (A3 and A4) for Alaska, both informed by the domestic longline survey. The Alaskan VAST estimates show a general decline in sablefish abundance in sub-area A4 starting around 2000, with a small increase around 2010 then declining a low point in 2015 (Figure C.7). The index for sub-area A4 is less variable over time and represents an overall

lower biomass than sub-area A3 (Figure C.7). As for the Domestic Longline survey data, the index of abundance has increased to an all-time high in the last ~8 years of the time series; this was more pronounced in sub-area A4 than A3.

Interpolated maps of relative abundance estimate that biomass density of the US West Coast similarly show a slight northerly shift (Figure C.8); the most recent regional assessment of survey biomass estimated roughly three-fourths of surveyed biomass is found to the north of Point Conception, which corresponds with sub-area C2 in the OM. Similarly, biomass density in the western Gulf of Alaska has increased dramatically during the last five years (Figure C.9). This region is sampled biennially by the Alaskan domestic longline survey and should be interpreted with caution as sample sizes (number of stations and high whale depredation) are low. There was no spatial patterning to the residuals for either region (Figures C.10 and C.11).

C.2.4 Discussion

This survey standardization for sablefish was undertaken with the goal of developing a range-wide index that could act as the index of relative abundance in a range-wide operating model for northeast Pacific sablefish.

In terms of geographic distribution, density maps (Figures C.8 and C.9) indicate a north and westerly shift in sablefish distribution in recent years. There are several potential factors that could lead to the observed biomass trends, none of which can be definitively ruled out by this analysis. The first, which is suggested by the density maps, is that fish have migrated from the Gulf of Alaska to the eastern Bering Sea over the last decade or so, leading to a sharp increase in the biomass in the western Gulf/Bering Sea. This hypothesis is unlikely, because the fish found in the Bering Sea are primarily younger and smaller fish resulting from the recent large recruitment. The second hypothesis is that proportionally larger recruitment pulses have occurred in the eastern Gulf of Alaska compared to other regions; these patterns are likely related to environmental changes, such sea surface height (Stachura et al., 2014), also an important driver of sablefish recruitment in the California Current (Tolimieri and Haltuch, 2023). These recruitment pulses would mean a positive change in the absolute number of individuals present in sub-area A4 at the expense of sub-area A3. The cold pool, which has provided more habitat for sablefish in the Aleutian Islands/Bering Sea, leading to large sablefish catches in recent decades, has only recently receded.

Alternatively, considering recent demographic investigations of sablefish growth across ecosystems (e.g., Chapter 2), it is possible that oceanographic and/or fishery pressures lead to a decline in maximum size obtained by individual sablefish in the eastern Gulf, depressing biomass estimates after 2010. The aforementioned growth work did not observe a similar temporal breakpoint for sablefish growth west of 145°W, although sablefish obtain their largest size in this area. The evidence from the VAST analysis is insufficient to confirm which hypothesis (migration, recruitment or reduction in size-at-age) is driving the observed changes, although the most recent Alaskan sablefish assessment documented a 47% increase in observed biomass in the longline survey from 2018 to 2019 and anticipates a large cohort from 2016 (Goethel et al., 2021), suggesting that recruitment and movement could accelerate the increasing trend in the Alaska region. Similarly, the West Coast assessment estimated a large recruitment pulse in 2016 (Haltuch et al., 2019).

C.2.5 Tables

Table C.1: Datasets used in the VAST standardizations. *Indicates the minimum and maximum year used, though coverage may be intermittent.

Survey Code	**Years***	**Name**	**Region**	**Treatment**	**Design**	**Depth Range**
WCGBTS	2003-2019	West Coast Groundfish Bottom Trawl Survey	West coast of US	Single Index	Random stratified Design	55-1300m
Triennial	Intermittent; 1980-2004	Triennial West Coast groundfish survey	West coast of US	Two Indices: 1. Early: 1980-1995 2. Late: 1998-2004	Transect based design	<500m
Alaska Domestic Longline Survey	1990-2019	Alaska Domestic Longline Survey	Gulf of Alaska, Aleutian Islands	Single Index	Random stratified design	10-1350m

Table C.2: Summary of VAST model configuration used in this analysis.

Region	**Number of Knots**	**Treatment of Intercepts**	**Treatment of Spatiotemporal Variation**	**Observation Model**	**Rationale**
California Current	500	Both intercepts as fixed effect	Each year and knot as a fixed effect; predictions north and south of 36°N	Catch rates: Gamma (option 2) Encounter probabilities: Conventional delta-model (option 0)	As is done in regional assessment VAST standardization, used for management
Alaska	500	Both intercepts as fixed effect	Each year and knot as a fixed effect; predictions east and west of 145°W	Catch rates: Gamma (option 2) Encounter probabilities: Poisson-link Delta-model (option 4)	As recommended by Groundfish Assessment Program survey team, to account for years with 100% encounter rates

Table C.3: Summary of parameter estimates from the California Current VAST standardization.

Parameter Name	Description	Estimate	Standard Error	Survey
ln_H_input	First anisotropy parameter	-0.529	0.080	California Current Surveys
ln_H_input.1	Second anisotropy parameter	-0.409	0.088	California Current Surveys
beta1_ft	Intercepts for year effect, capture probability, C1	0.290	0.614	California Current Surveys
beta1_ft.1	“	0.623	0.607	California Current Surveys
beta1_ft.2	"	1.598	0.617	California Current Surveys
beta1_ft.3	"	1.037	0.605	California Current Surveys
beta1_ft.4	"	0.638	0.604	California Current Surveys
beta1_ft.5	"	0.950	0.606	California Current Surveys
beta1_ft.6	"	-0.462	0.600	California Current Surveys
beta1_ft.7	"	1.232	0.607	California Current Surveys
beta1_ft.8	"	0.833	0.600	California Current Surveys
beta1_ft.9	"	0.792	0.594	California Current Surveys
beta1_ft.10	"	0.674	0.594	California Current Surveys
beta1_ft.11	"	-0.143	0.589	California Current

				Surveys
beta1_ft.12	"	-0.232	0.590	California Current Surveys
beta1_ft.13	"	-0.317	0.596	California Current Surveys
beta1_ft.14	"	-0.354	0.593	California Current Surveys
beta1_ft.15	"	-0.144	0.589	California Current Surveys
beta1_ft.16	"	0.049	0.591	California Current Surveys
beta1_ft.17	"	-0.253	0.592	California Current Surveys
beta1_ft.18	"	0.065	0.594	California Current Surveys
beta1_ft.19	"	0.366	0.591	California Current Surveys
beta1_ft.20	"	-0.117	0.588	California Current Surveys
beta1_ft.21	"	-0.098	0.591	California Current Surveys
beta1_ft.22	"	-0.066	0.592	California Current Surveys
beta1_ft.23	"	-0.311	0.591	California Current Surveys
beta1_ft.24	"	0.115	0.605	California Current Surveys
L_omega1_z	Trimmed cholesky pointwise variance in spatial variation, C1	3.692	0.341	California Current Surveys

L_epsilon1_z	Trimmed cholesky pointwise variance in spatio-temporal variation, C1	0.910	0.068	California Current Surveys
logkappa1	Decorrelation rate, C1	-3.248	0.126	California Current Surveys
beta2_ft	Intercepts for year effect, capture probability, C2	6.296	0.223	California Current Surveys
beta2_ft.1	"	5.620	0.204	California Current Surveys
beta2_ft.2	"	6.019	0.205	California Current Surveys
beta2_ft.3	"	5.391	0.195	California Current Surveys
beta2_ft.4	"	5.454	0.204	California Current Surveys
beta2_ft.5	"	5.299	0.191	California Current Surveys
beta2_ft.6	"	5.772	0.199	California Current Surveys
beta2_ft.7	"	6.586	0.189	California Current Surveys
beta2_ft.8	"	6.126	0.173	California Current Surveys
beta2_ft.9	"	6.316	0.171	California Current Surveys
beta2_ft.10	"	6.093	0.170	California Current Surveys
beta2_ft.11	"	6.058	0.171	California Current Surveys
beta2_ft.12	"	6.069	0.170	California Current

				Surveys
beta2_ft.13	"	5.778	0.170	California Current Surveys
beta2_ft.14	"	5.768	0.171	California Current Surveys
beta2_ft.15	"	5.822	0.166	California Current Surveys
beta2_ft.16	"	5.842	0.170	California Current Surveys
beta2_ft.17	"	5.727	0.173	California Current Surveys
beta2_ft.18	"	5.806	0.176	California Current Surveys
beta2_ft.19	"	5.817	0.168	California Current Surveys
beta2_ft.20	"	5.946	0.173	California Current Surveys
beta2_ft.21	"	5.740	0.170	California Current Surveys
beta2_ft.22	"	6.076	0.169	California Current Surveys
beta2_ft.23	"	6.140	0.173	California Current Surveys
beta2_ft.24	"	6.006	0.185	California Current Surveys
L_omega2_z	Trimmed cholesky pointwise variance in spatial variation, C2	1.288	0.064	California Current Surveys
L_epsilon2_z	Trimmed cholesky pointwise variance in spatio-temporal variation, C2	1.070	0.032	California Current Surveys

logkappa2	Decorrelation rate, C2	-2.695	0.059	California Current Surveys
logSigmaM	Variance parameter for positive catch rates	-0.006	0.008	California Current Surveys

Table C.4: Summary of parameter estimates from the Alaska VAST standardization.

Parameter	Description	Estimate	Standard Error	Survey
ln_H_input	First anisotropy parameter	0.15	0.15	AK Domestic Longline Survey
ln_H_input.1	Second anisotropy parameter	-0.01	0.13	AK Domestic Longline Survey
beta1_ft	Intercepts for year effect, capture probability, A4	6.96	0.74	AK Domestic Longline Survey
beta1_ft.1	"	6.86	0.74	AK Domestic Longline Survey
beta1_ft.2	"	6.64	0.73	AK Domestic Longline Survey
beta1_ft.3	"	7.09	0.73	AK Domestic Longline Survey
beta1_ft.4	"	6.84	0.73	AK Domestic Longline Survey
beta1_ft.5	"	5.51	0.72	AK Domestic Longline Survey
beta1_ft.6	"	7.24	0.77	AK Domestic Longline Survey
beta1_ft.7	"	7.52	0.76	AK Domestic Longline Survey
beta1_ft.8	"	6.67	0.76	AK Domestic Longline Survey
beta1_ft.9	"	6.77	0.75	AK Domestic Longline Survey
beta1_ft.10	"	7.09	0.81	AK Domestic Longline Survey
beta1_ft.11	"	7.2	0.78	AK Domestic Longline Survey
beta1_ft.12	"	6.66	0.78	AK Domestic Longline Survey
beta1_ft.13	"	7.88	0.81	AK Domestic Longline Survey
beta1_ft.14	"	7.17	0.81	AK Domestic Longline Survey
beta1_ft.15	"	7.38	0.79	AK Domestic Longline Survey
beta1_ft.16	"	7.14	0.81	AK Domestic Longline Survey
beta1_ft.17	"	7.3	0.76	AK Domestic Longline Survey
beta1_ft.18	"	7.06	0.81	AK Domestic Longline Survey
beta1_ft.19	"	7.22	0.75	AK Domestic Longline Survey
beta1_ft.20	"	6.69	0.77	AK Domestic Longline Survey
beta1_ft.21	"	7.27	0.76	AK Domestic Longline Survey

beta1_ft.22	"	6.91	0.82	AK Domestic Longline Survey
beta1_ft.23	"	7.35	0.76	AK Domestic Longline Survey
beta1_ft.24	"	6.4	0.76	AK Domestic Longline Survey
beta1_ft.25	"	7.46	0.78	AK Domestic Longline Survey
beta1_ft.26	"	7.66	0.8	AK Domestic Longline Survey
beta1_ft.27	"	6.39	0.78	AK Domestic Longline Survey
L_omega1_z	Trimmed cholesky pointwise variance in spatial variation, A4	-2.98	0.37	AK Domestic Longline Survey
logkappa1	Decorrelation rate, A4	-4.42	0.17	AK Domestic Longline Survey
beta2_ft	Intercepts for year effect, capture A3	-1.47	0.91	AK Domestic Longline Survey
beta2_ft.1	"	-1.6	0.91	AK Domestic Longline Survey
beta2_ft.2	"	-1.45	0.91	AK Domestic Longline Survey
beta2_ft.3	"	-2	0.91	AK Domestic Longline Survey
beta2_ft.4	"	-1.67	0.91	AK Domestic Longline Survey
beta2_ft.5	"	-14.61	1.05	AK Domestic Longline Survey
beta2_ft.6	"	-0.27	0.9	AK Domestic Longline Survey
beta2_ft.7	"	-1.84	0.94	AK Domestic Longline Survey
beta2_ft.8	"	-14.66	1.05	AK Domestic Longline Survey
beta2_ft.9	"	-2.3	0.94	AK Domestic Longline Survey
beta2_ft.10	"	-1.44	0.93	AK Domestic Longline Survey
beta2_ft.11	"	-1.33	0.93	AK Domestic Longline Survey
beta2_ft.12	"	-1.8	0.97	AK Domestic Longline Survey
beta2_ft.13	"	-1.87	0.94	AK Domestic Longline Survey
beta2_ft.14	"	-1.28	0.95	AK Domestic Longline Survey
beta2_ft.15	"	-2.51	0.98	AK Domestic Longline Survey
beta2_ft.16	"	-1.8	0.98	AK Domestic Longline Survey
beta2_ft.17	"	-1.99	0.96	AK Domestic Longline Survey

beta2_ft.18	"	-1.86	0.98	AK Domestic Longline Survey
beta2_ft.19	"	-2.2	0.94	AK Domestic Longline Survey
beta2_ft.20	"	-1.91	0.98	AK Domestic Longline Survey
beta2_ft.21	"	-2.09	0.93	AK Domestic Longline Survey
beta2_ft.22	"	-1.56	0.94	AK Domestic Longline Survey
beta2_ft.23	"	-2.37	0.93	AK Domestic Longline Survey
beta2_ft.24	"	-1.89	0.98	AK Domestic Longline Survey
beta2_ft.25	"	-2.68	0.93	AK Domestic Longline Survey
beta2_ft.26	"	-1.09	0.93	AK Domestic Longline Survey
beta2_ft.27	"	-2.05	0.95	AK Domestic Longline Survey
beta2_ft.28	"	-14.54	1.05	AK Domestic Longline Survey
beta2_ft.29	"	-1.67	0.97	AK Domestic Longline Survey
beta2_ft.30	"	-0.36	0.95	AK Domestic Longline Survey
L_omega2_z	Trimmed cholesky pointwise variance in spatial variation, A3	1.26	0.38	AK Domestic Longline Survey
L_epsilon2_z	Trimmed cholesky pointwise variance in spatio-temporal variation	0.23	0.02	AK Domestic Longline Survey
logkappa2	Decorrelation rate, A3	-5.59	0.14	AK Domestic Longline Survey
logSigmaM	Variance parameter for positive catch rates	-0.44	0.01	AK Domestic Longline Survey

7.2.6 Figures

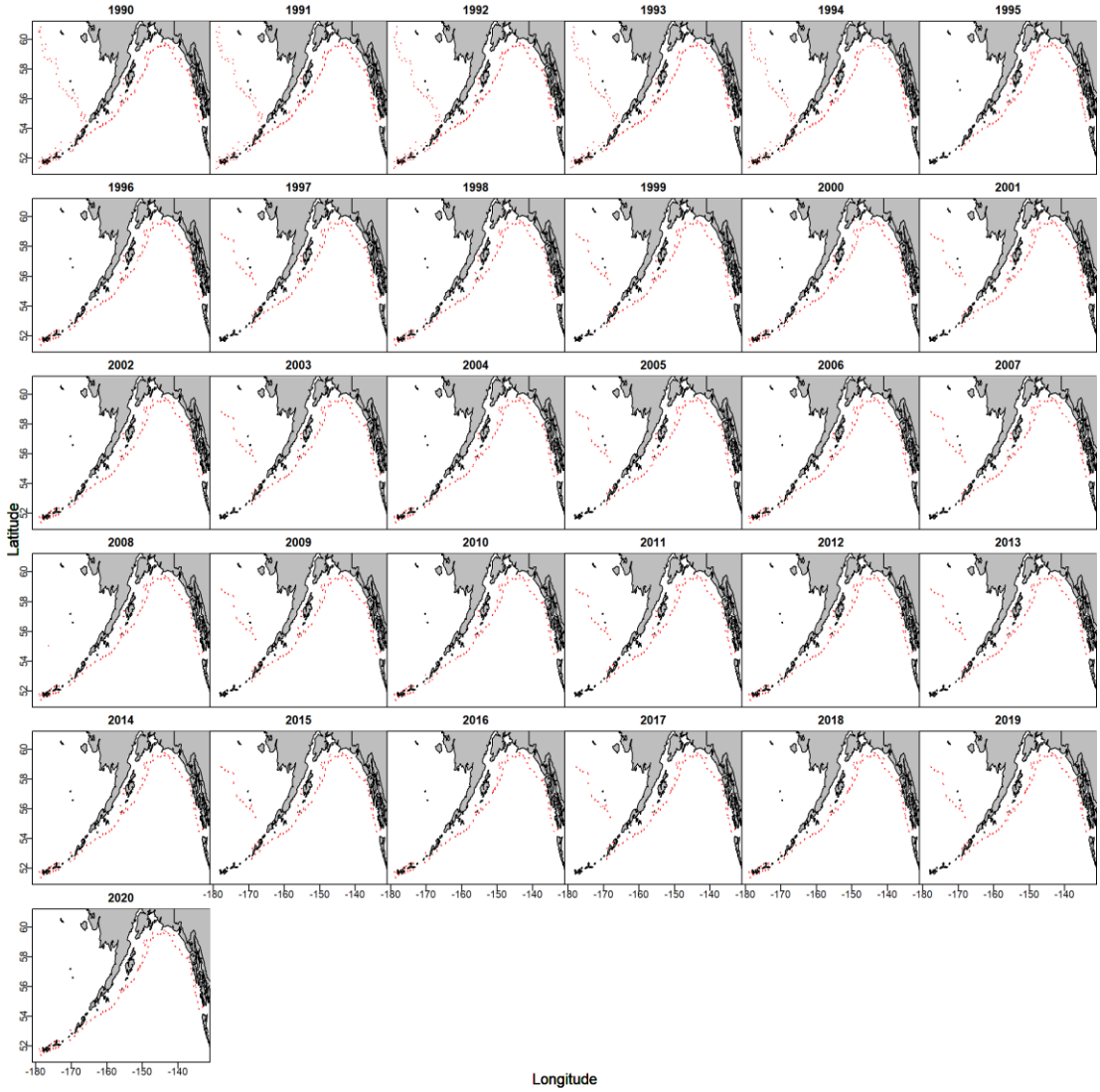


Figure C.5: Map of sampling data from Alaska Domestic Longline survey.

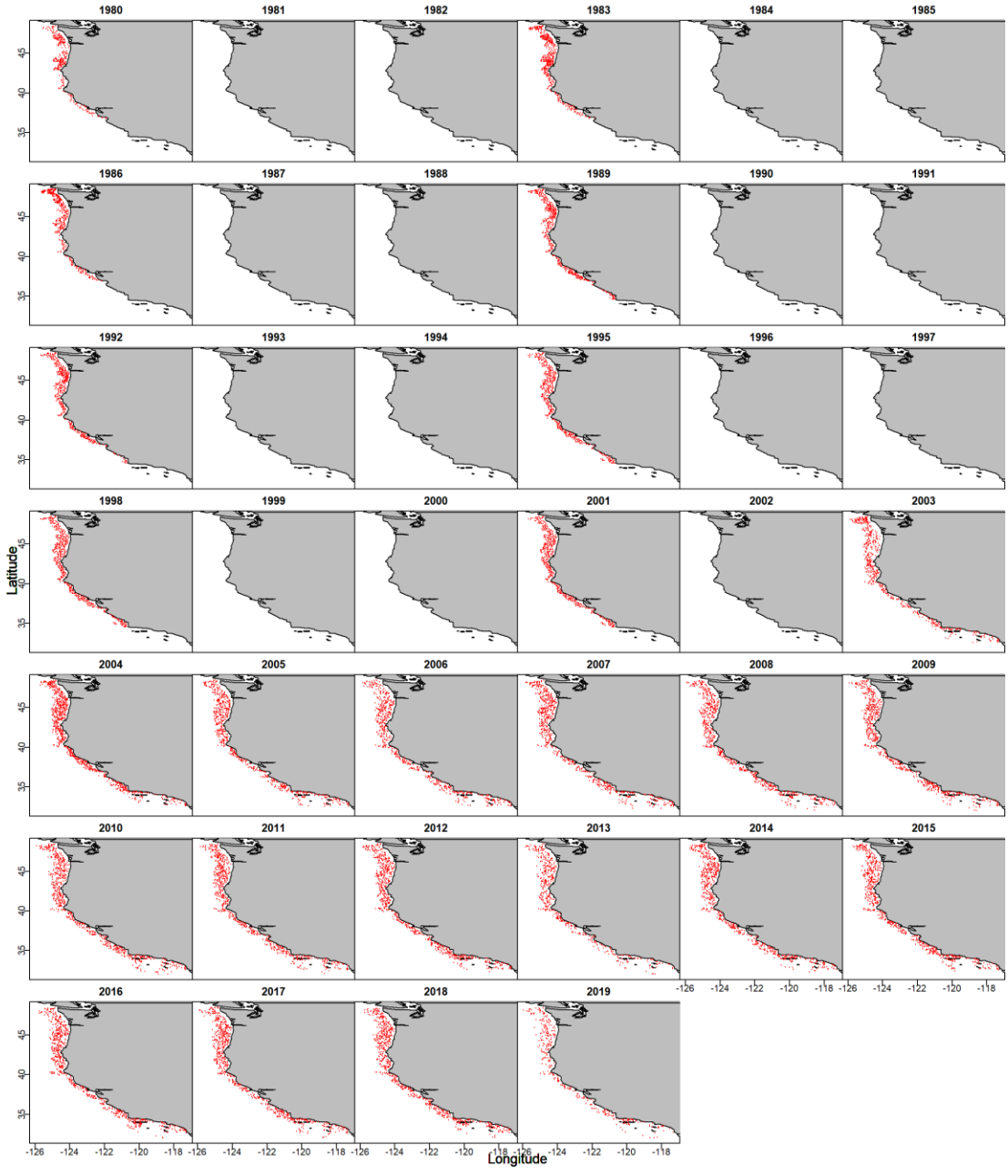


Figure C.6: Map of survey data from the Triennial (1980-2004) and West Coast Groundfish Bottom Trawl surveys (2003-2019).

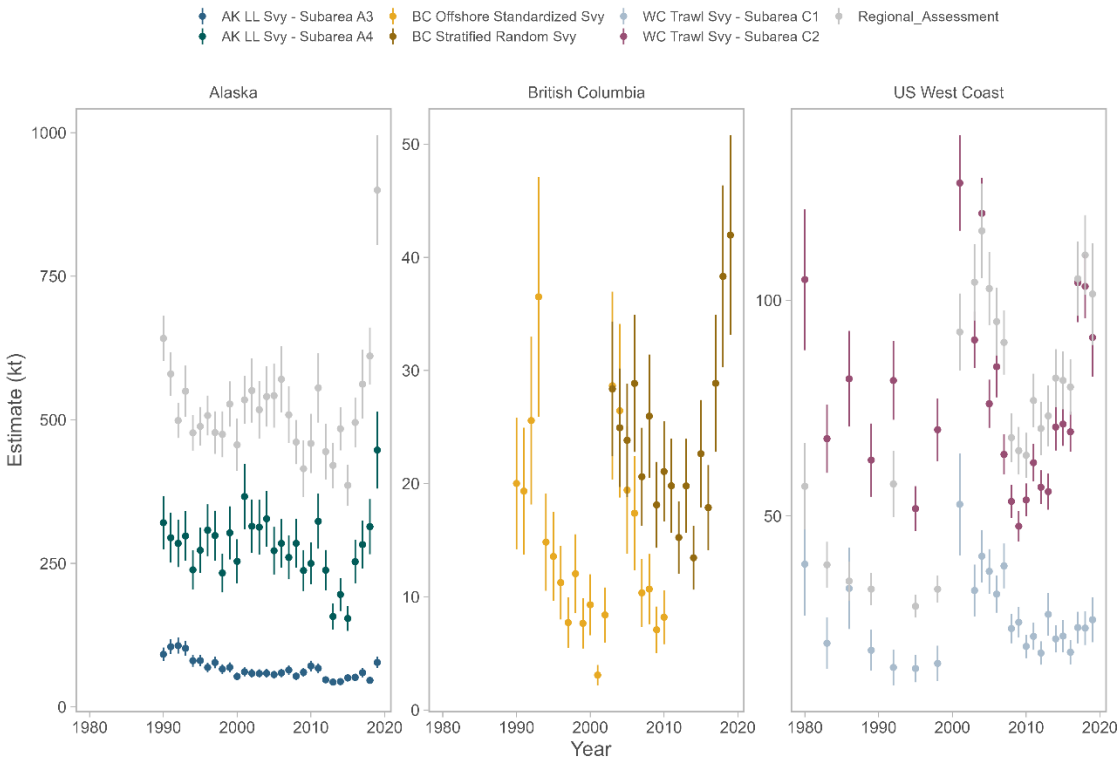


Figure C.7: VAST model estimates (colored points) for the standardizations performed separately for Alaska (leftmost panel) and the California Current (rightmost panel). Vertical bars indicate ±1 standard error. Indices used in recent regional assessments are shown in grey.

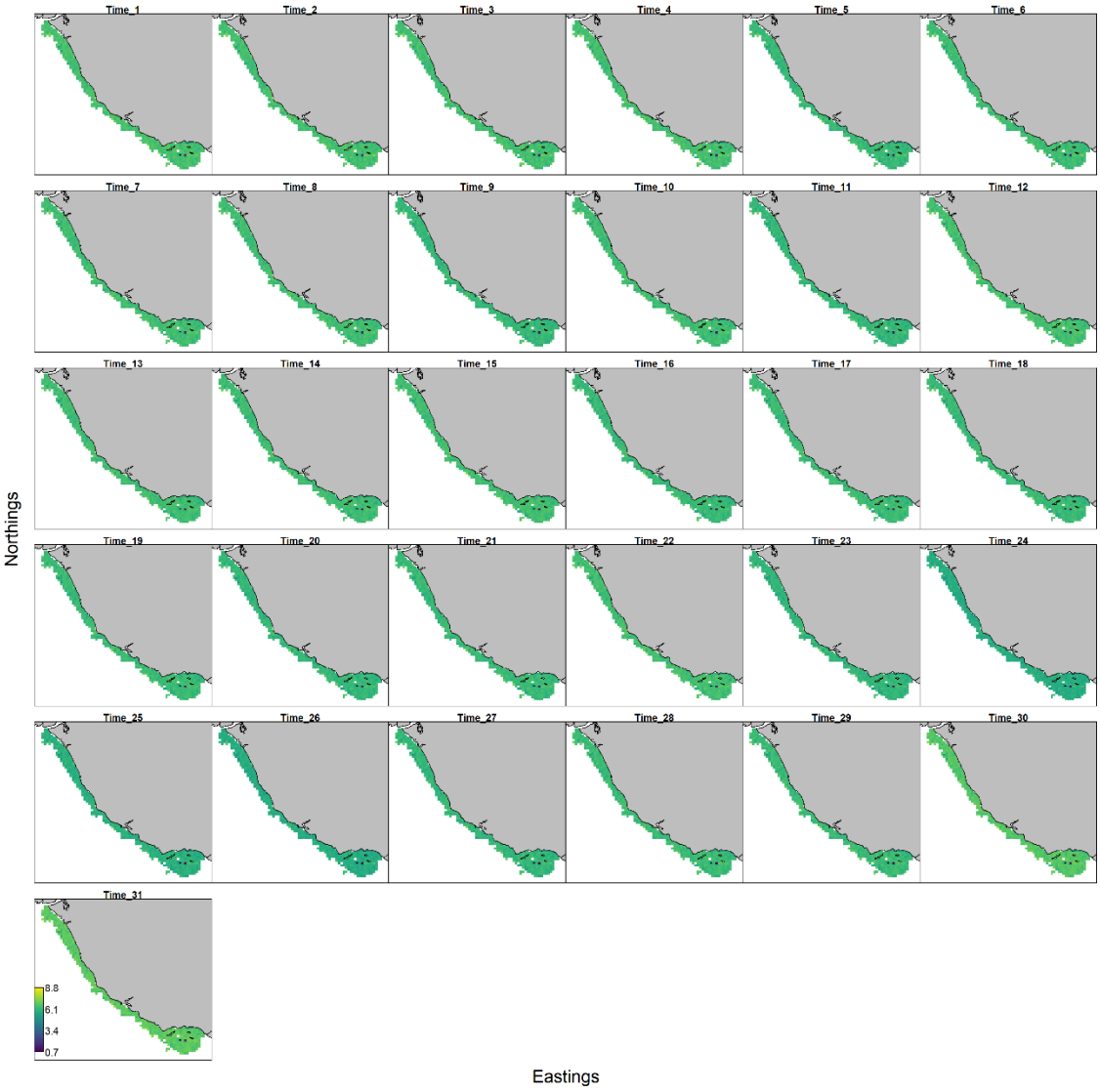


Figure C.8: Spatial distribution of estimated log-relative abundance for years with survey data for US West Coast VAST standardization.

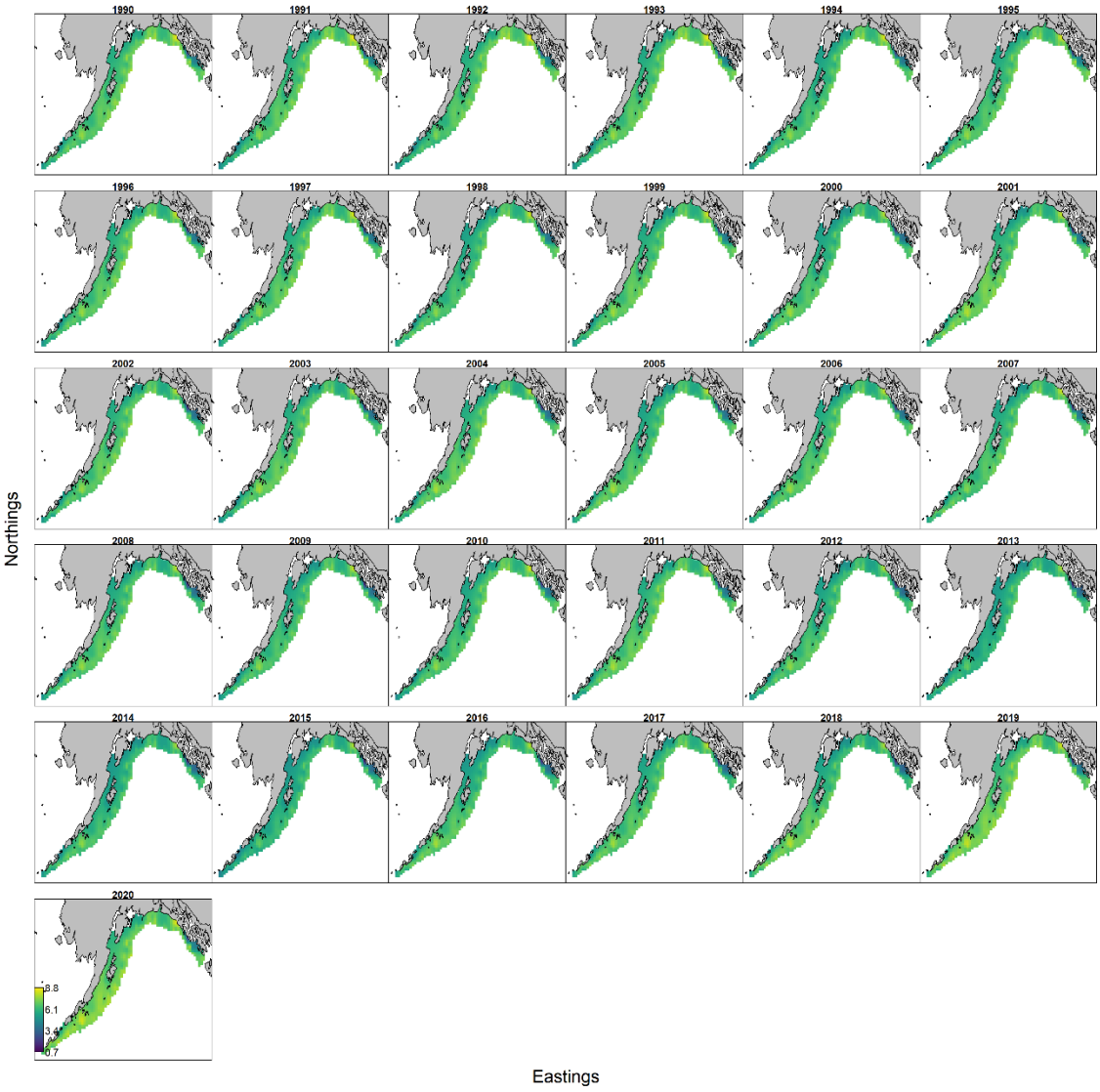


Figure C.9: Spatial distribution of estimated log-relative abundance for Alaska VAST standardization.

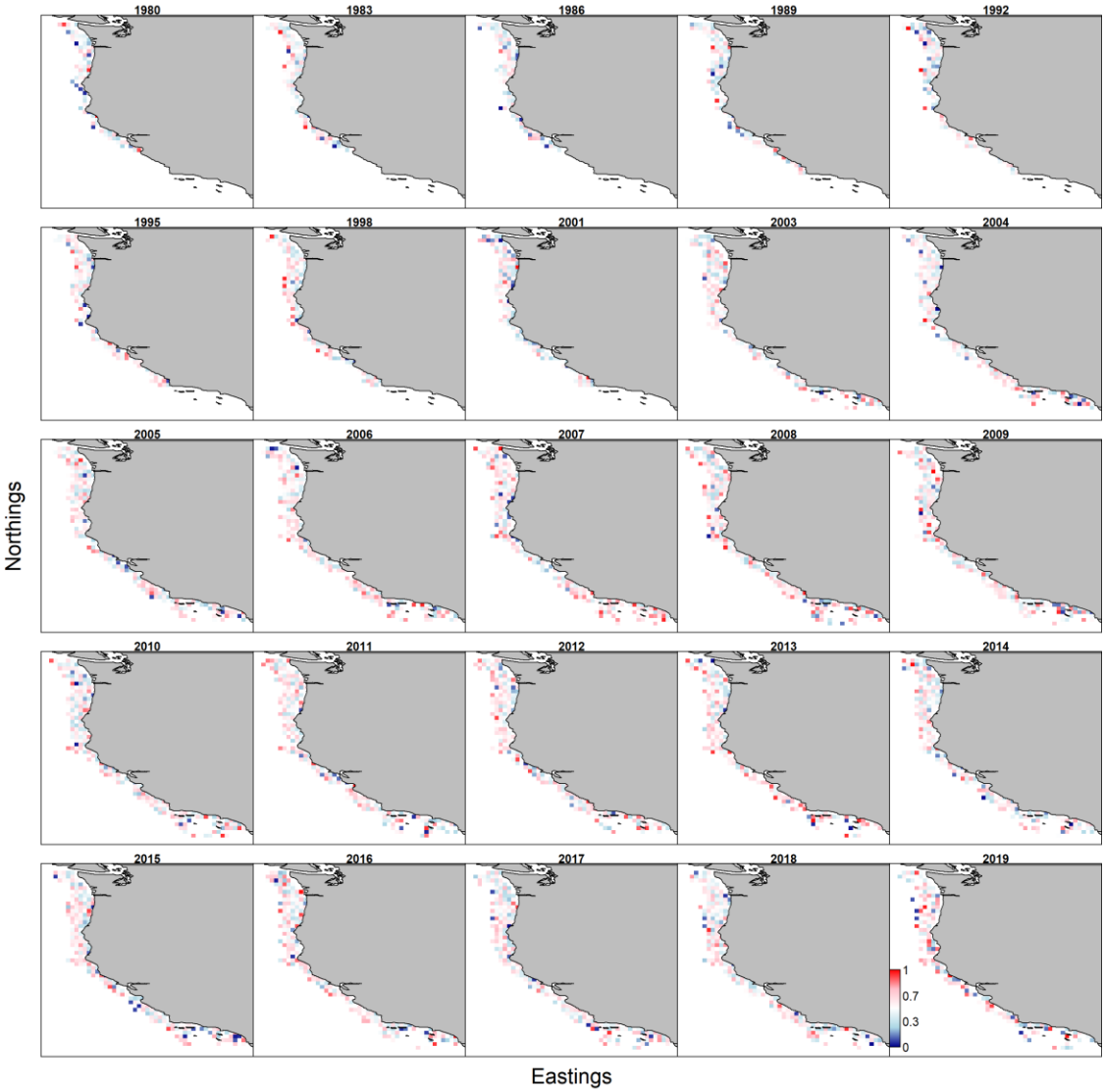


Figure C.10: Pearson residuals of estimated log-relative abundance for years with survey data for California Current standardization. The residuals are scaled between 0 and 1 to eliminate the influence of outlier points on visualization.

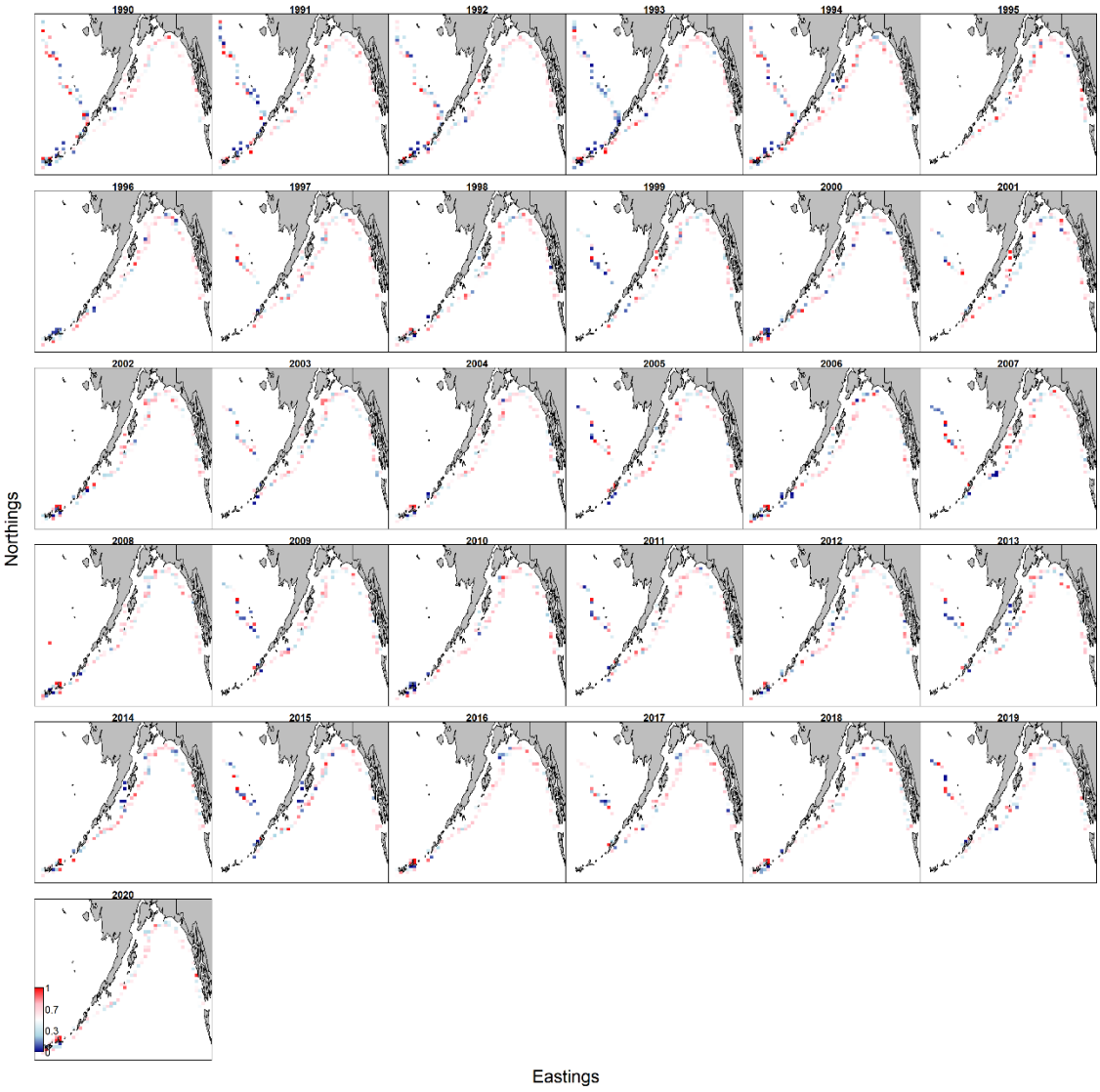


Figure C.11: Pearson residuals for estimated log-relative abundance for years with survey data for Alaskan standardization. The residuals are scaled between 0 and 1 to eliminate the influence of outlier points on visualization.

C.3 Sensitivity Analyses for the Operating Model

A small set of sensitivity analyses was conducted to explore the behavior of the OM, particularly in relation to processes and/or parameters that were pre-specified instead of being estimated. The sections below summarize the experiments, highlight major differences between the sensitivity runs and the conditioned OM (the "base OM"). Unless otherwise specified, these sensitivity analyses started at the parameter values used in the base operating model, and allowed for re-estimation of key scaling and recruitment parameters, including deviations (R_0^k,τ_y^{ik},σ_R, and $\tilde{R}_y^k$).

C.3.1 Sensitivity to movement rates

The OM was run with movement set to the values from the "high" or "low" movement scenarios (*Appendix C.1)*, as well as for a scenario where there was no movement For the low and high movement scenarios, all non-movement parameters are set to those estimated for the base OM to illustrate the impact of changing only the movement paradigm on derived quantities; the time to complete satisfactory conditioning under these alternative paradigms was prohibitive. In the sensitivity analysis with no movement between sub-areas, key scaling and recruitment parameters including recruitment deviations were estimated, to investigate whether this additional flexibility allowed the model to recover the original time series trend in the absence of movement.

No movement-based sensitivities result in collapsed populations. The scenarios with "low" and "high" movement rates did not result in spawning stock biomass trajectories qualitatively very different from those for the base OM for any management region, despite the fact that recruitment downscaling and other parameters were set to the base model values (Figure C.12). For Alaska and the US West Coast, spawning stock biomass differs by an average of 5% from 2000-2019 under the "low" and "high" movement scenarios. This suggests that the range of uncertainty in movement rates described in *Appendix C.1* does not correspond to large differences in the scale of spawning biomass in these regions. The spawning stock biomass trajectories for British Columbia are also qualitatively similar to those for the base OM under the low and high movement scenarios, but lower movement rates resulted in lower biomass estimates (10% on average, 2000-2019) and higher movement rates resulted in biomass estimates elevated to the same degree as the other two management regions (5% on average, 2000-2019). These findings are consistent with the fact that the US West Coast and Alaska have higher self-seeding rates than British Columbia (Figure 3.6),

and that growth is less uncertain in sub-areas C1 and A4 (Figure 3.3; Figure C.2), leading to less variation among movement-at-age scenarios in those regions than in British Columbia. In contrast, the British Columbia management region has both greater uncertainty in length-at-age (Figure 3.3; Figure C.2) and lower rates of self-seeding across all scenarios (Figure 3.6), suggesting dependence on immigration. For reference, the average spawning stock biomass subsidy (in the base OM) from British Columbia to adjacent areas amounts to nearly 30% of British Columbia biomass whereas less than 8% of West Coast or Alaskan biomass is a result of immigration from British Columbia. This dependency is also confirmed by the "low" movement scenario resulting in slightly less spawning stock biomass off British Columbia. Similarly, the estimated proportion of recruitment (not shown) from Stock 3 assigned to sub-area B3 (north British Columbia) averaged 21% after the onset of survey data when movement was assumed to be zero, whereas in the base OM it averaged 16% for the same period.

The spawning stock biomass trajectories from the sensitivity analysis without movement differ from the base OM in all regions. In Alaska and British Columbia, the years before the onset of regular survey data (<1990) are marked by elevated spawning stock biomass values, that decline to be within 10% of the base OM in the terminal 10 years of the model period (Figure C.12). The shape of the spawning stock biomass trajectory without movement for the US West Coast is similar to that for the base OM, although biomass values are on average 55% higher than the base OM from the late 1970s onward (Figure C.12).

These results suggest that the perception of model scale in the base OM is sensitive to the assumption of connectivity between regions, particularly in the absence of multiple data sources. In the case of British Columbia and Alaska, it appears that the simultaneous presence of survey and compositional data (from 2000 onwards) results in similar estimates of model scale for both regions, even if is assumed that there is no movement. In contrast, the onset of survey data (~1980) for the US West Coast instead results in an accumulation of biomass in that region in the absence of movement. This suggests that the dynamics of the US West Coast population in the base OM is contingent upon the assumption of export to adjacent areas.

C.3.2 Sensitivity to ignoring age-composition data

A sensitivity analysis was conducted in which the age-composition data did not contribute to the overall joint likelihood, to evaluate whether the conditioned OM could still capture the general trend in spawning stock biomass and survey indices without this information, which slows run times considerably. The time series of spawning stock biomass for this sensitivity analysis is qualitatively similar to the base OM from ~2000-2019 in all regions, but the early model scale is much higher than the base OM (Figure C.13). Terminal (2015-2019) spawning stock biomass in Alaska and British Columbia was 40% and 30% lower than the base OM, respectively, while spawning stock biomass for the US West Coast is 13% higher than the base OM in the same period. The discrepancies between the base OM and this sensitivity are less pronounced following the onset of survey biomass data, and suggest difficulty identifying the model scale for any region before 1990 in the absence of age-composition data.

C.3.3 Sensitivity to density-dependence specification

The base OM models recruitment at the stock level, and estimates a unique deviate for each year-stock combination. This represents the hypothesis that density dependence is localized to each stock and varies temporally, which is ecologically plausible given the extent of the spatial domain (Tolimieri et al., 2018) and existence of known nursery areas for sablefish (Cornthwaite et al., 2020). In these sensitivity analyses, all parameters were set to the values from the base OM except for the recruitment deviations and uncertainty thereof (σ_R). We attempted to estimate the recruitment deviations as random effects, but this proved computationally intractable. A sensitivity analysis was run in which deviations were instead mirrored among stocks, meaning that while density dependence remain localized, the time series of deviations from expected recruitment are assumed to be identical across stocks. This represents the alternative hypothesis that sablefish recruitment is synchronous throughout the Northeast Pacific, and that there are no lags or independent dynamics in recruitment signals across the range.

The examination of the historical OM recruitment deviations revealed stocks 2 and 3 (spanning the US West Coast north of 36˚N to the eastern Gulf of Alaska) to be highly auto-correlated (Figure 3.12), so a sensitivity analysis was run where recruitment was mirrored between these two stocks. This represents the hypothesis that, though density dependence is still localized into two

separate stocks in this area, deviations from expected recruitment are synchronous between them, possibly due to shared ecological conditions.

The sensitivity analysis where recruitment deviations are mirrored across all stocks results in spawning stock biomass trajectories in Alaska ~30% higher than the base OM for the duration of the time series (Figure C.14, top panel). The discrepancy between spawning stock biomass trajectories is trivial (<5%) for British Columbia following the onset of survey data (>1990, Figure C.14, top panel). Stock spawning biomass trajectories off the US West coast are lower than the base OM by ~20% for the duration of the time series (Figure C.14, top panel). This experiment was the only model run for which the Hessian matrix was invertible. The time series of recruitment deviations for the entire population most closely resembles the base OM values for stock 2, where recruitment deviations (in log space) are between 1 and 2 in the 1970s, decline through the mid-1990s, and exhibit a peak in the late 2010s (Figure C.14, bottom panel, blue line). The time series of recruitment deviations in this sensitivity run is less variable than the base OM values for the independent stocks 1 and 4 (Figure C.14). This result might be influenced by the fact that stock 2 encompasses the longest time-series of age-composition data with the largest overall input sample size (Table 3.9), and that young fish (<5) are more strongly selected for in US West Coast surveys and fishery fleets than in Alaska (Figures 3.19 and 3.20). The spawning stock biomass trajectory for British Columbia did not vary greatly across recruitment sensitivities, especially since the onset of survey data in the mid-1980s. This is possibly because the British Columbia management region encompasses shared stocks with both Alaska and the US West Coast, and may indicate that the survey data available in that region can alias spatial dynamics.

The scenario where the recruitment deviations are mirrored within stocks 2 and 3 and otherwise independent did not converge. Spawning stock biomass trajectories for the US West Coast and British Columbia are uniformly lower than the base OM (Figure C.14, top panel, green line); spawning stock biomass was slightly (~12%) higher than the base OM for Alaska until the mid-1990s, after which it was 6%-25% lower. The recruitment deviations estimated for the combined stocks 2 and 3 were less variable than the respective values from the base OM, particularly for stock 3, and more closely resemble the time series from the area north of 36˚N in the base OM (Figure C.14, bottom panel, green line). This is consistent with the fact that sample sizes for age-

composition are larger from the area north of 36°N, and the survey abundances are roughly three times higher in that region.

The investigations into the specification of density dependence suggest that the dynamics within stock 2 are highly influential to the spawning stock biomass and recruitment trajectories for all regions. It is possible that this outcome is predetermined by the abundance of age-composition data in this region.

C.3.4 Figures

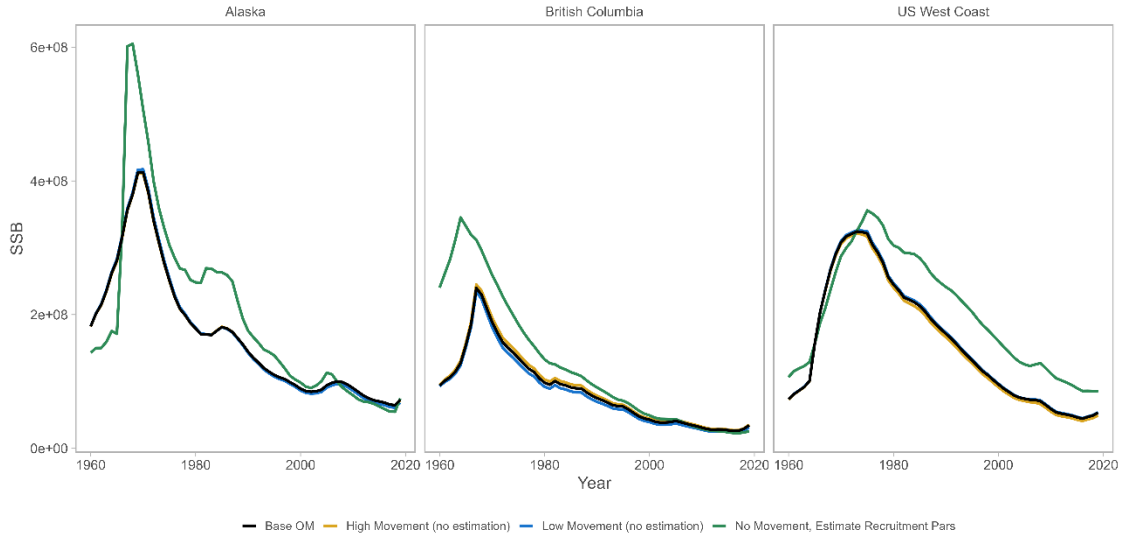


Figure C.12: Sensitivity analyses with various movement regimes (high movement: gold lines; low movement: blue lines; and no movement with recruitment deviations, unfished recruitment, recruitment downscaling, and standard deviation in recruitment estimated: dark green lines). Panels indicate time series of spawning stock biomass by management region. The black line is the base operating model.

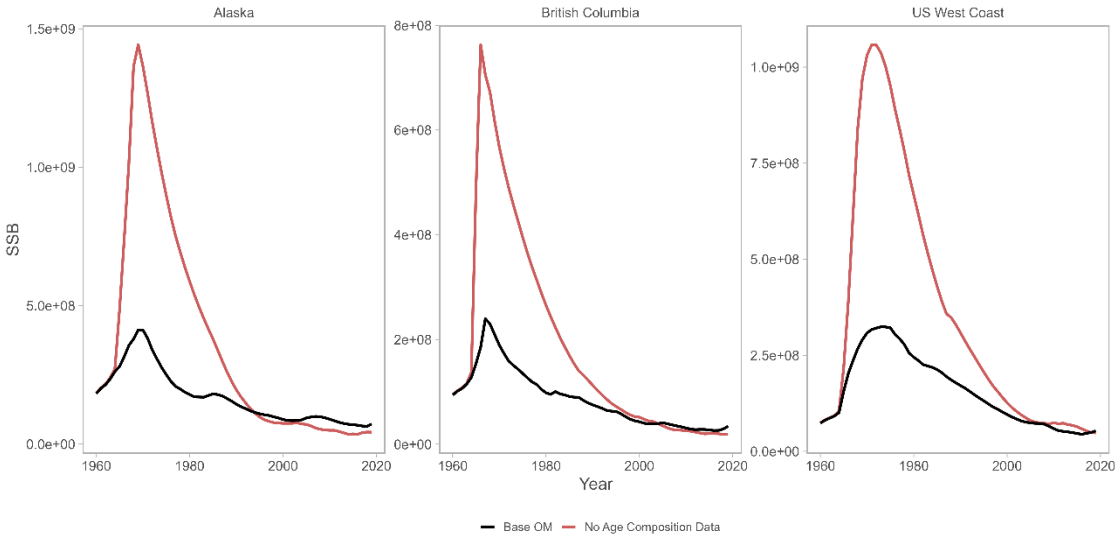


Figure C.13: Sensitivity analysis with age-composition data excluded from the negative log-likelihood (red lines). Panels indicate time series of spawning stock biomass by management region. The black line is the base operating model.

Figure C.14: Sensitivity analyses when the recruitment deviations are mirrored among all stocks (blue lines) and mirrored for stocks 2 and 3 (green lines). Top panels indicate the time series of spawning stock biomass by management region; bottom panels are time series of recruitment deviations by stock. The black line indicates values from the base operating model.

Appendix D. Supplementary Information for Chapter 4

D.1 Estimation Method Structure and Dynamics

For a stock consisting of multiple spatial areas, the general dynamics of population biomass B (in t) and abundance N (in numbers) in spatial area i during year y are given by Equations D.1 and D.2. For ease of presentation, the notation i refers to refer to spatial areas, though many demographic parameters and processes are specific to stocks k. It is possible for a stock k to be comprised of more than one spatial area i, in which case demographic parameters would be shared.

$$B_y^i = \left(1 - \sum_{i \neq j} \boldsymbol{X}^{i,j}\right)\left[\, s_{y-1}^i \left(\alpha^i N_{y-1}^i + \rho^i B_{y-1}^i\right)\right] \quad \text{(D.1)}$$

$$+ \sum\nolimits_{i \neq j} \boldsymbol{X}^{j,i}\left[\, s_{y-1}^j \left(\alpha^j N_{y-1}^j + \rho^j B_{y-1}^j\right)\right] + \omega^i (R_y^i)$$

$$\mathrm{N}_y^i = (1 - \sum_{i \neq j} \mathbf{X}^{i,j})\left[\, \mathrm{s}_{y-1}^i \mathrm{N}_{y-1}^i\right] + \sum_{i \neq j} \mathbf{X}^{j,i}\left[\, \mathrm{s}_{y-1}^j \mathrm{N}_{y-1}^j\right] + \mathrm{R}_y^i \qquad (D.2)$$

where $\mathbf{X}^{i,j}$ is the proportion of biomass (numbers) that moves from area i to area j each year, s_y^i is the survival rate in spatial area i during year y, a function of instantaneous rate of natural mortality M, and the annual harvest rate in spatial area i during year y, u_y^i.

$$s_y^i = \left(1 - u_y^i\right) e^{-M^i} \qquad (D.3)$$

α^i is the Ford-Brody growth model intercept[7] for stock i, computed from the asymptotic weight-at-age (in kg) for spatial area i, W_∞^i:

$$\alpha^i = W_\infty^i \left(1 - \rho^i\right) \quad (D.4)$$

ρ^j is the growth coefficient from a Ford–Walford plot for stock i, calculated from the von Bertalanffy growth rate κ:

$$\rho^i \;=\; exp\left(-\kappa^i\right) \qquad (D.5)$$

[7] This is obtained by assuming fish weight-at-age a (in kg) can be described by $w_a = W_\infty^i \left(1 - exp[\kappa^i a - a_0^i\right)^{b^i}$, and taking the limit to solve for α^i as $a \to \infty$. W_∞^i, κ^i and a_0^i are stock-specific growth parameters estimated externally to the EM using data simulated from the OM. See Figure D.6 and Tables D.3 and D.4.

ω^i is the weight (in kg) of an individual recruiting into the fishery in spatial area i.

Catches are assumed to be taken during the middle of the year. The annual harvest rate for fishery f, u_y^f, is calculated as the ratio between the observed catch by fishery f during year *y* and the biomass available to fishery f, which is the total biomass at the start of year y in the spatial area(s) i where fishery f operates:

$$u_y^f = \frac{C_{obs,y}^f}{\sum_i \phi^{if} B_y^i} \qquad (D.6)$$

where $C_{obs,y}^f$ are the observed catch by fishery f during year y, and ϕ^{if} is a matrix defining whether fishery f operates in spatial area i. The harvest rate applied to spatial area i (exploited by one or more fisheries) is:

$$u_y^i = \sum_f \phi^{if} u_y^f \qquad (D.7)$$

When the assessed stock is comprised of more than one spatial area, the Beverton-Holt recruitment for the stock is allocated to each area based on parameters τ_y^{ik}, allowing for spatial auto-correlation of recruitment within a stock:

$$R_{y+1}^i = \tau_y^{ik} \frac{4h^k R_0^k B_y^k}{B_0^k(1-h^k) + B_y^k(5h^k - 1)} e^{-0.5\sigma_R^2 + \tilde{R}_y^k} \qquad (D.8)$$

where h^k is the steepness of the stock recruitment curve for stock *k* (expected proportion of R_0^k at $0.2B_0^k$); σ_R is the standard deviation of recruitment, R_0^k is the unfished ("virgin") recruitment for stock *k*, $\tilde{R}_y^k$ are random annual recruitment deviations specific to stock k and assumed to be normally distributed with mean zero and a parameter for the standard deviation σ_R, and B_0^k is the unfished biomass of stock k, given by the product of unfished recruitment and unfished biomass per recruit (ϕ_0^k).

$$B_0^k = \phi_0^k R_0^k \qquad (D.9)$$

$$\phi_0^k = \frac{\frac{\alpha^k exp(-M^k)}{1 - exp(-M^k)} + \omega^k\left(1 - exp(-M^k)\right)}{1 - \rho^k exp(-M^k)} \qquad (D.10)$$

B_y^k is the biomass of stock *k* at the start of year *y*; spatial area *i*) is always part of only one stock (*k*):

$$B_y^k = \sum_i \mathbf{\Lambda}^{ik} B_y^i \qquad (D.11)$$

$\mathbf{\Lambda}^{ik}$ is a matrix indicating whether spatial area *i* is nested within stock *k*, and τ_y^{ik} is the proportion of recruits to stock *k* that recruit to spatial area *i* during year *y*. If the assessed stock is comprised of only one spatial area, τ_y^{ik} is set to 1. (In the EM, τ^{ik} is estimated as a time-invariant parameter). The population is initialized at the start of the modelled period (1960) under the assumption that it was in fished equilibrium at that time. Given equilibrium biomass-per-recruit ϕ_{eq}^k, the equilibrium recruitment R_{eq}^k, and an initial recruitment deviation $\tilde{R}_{init}^k$. These equilibrium quantities depend on estimates of unfished biomass, initial survivorship, Ford-Brody growth parameters and the parameters of the Beverton-Holt stock-recruitment relationship for each stock *k*, and do not consider movement between spatial areas:

$$B_{1960}^k = \phi_{eq}^k R_{eq}^k e^{-0.5\sigma_R^2 + \tilde{R}_{init}^k} \qquad (D.12)$$

$$N_{1960}^k = \frac{R_{eq}^k}{1 - exp(-M^k)(1 - u_{init}^f)} e^{-0.5\sigma_R^2 + \tilde{R}_{init}^k} \qquad (D.13)$$

The equilibrium biomass-per-recruit, ϕ_{eq}^k considers exploitation in the initial year (u_{init}^f), which is fixed to 0.01 to avoid confounding with $\tilde{R}_{init}^k$ (see footnote for derivation of this syntax).[8]:

$$\phi_{eq}^k = \frac{\frac{\alpha^k exp(-M^k)\left(1 - u_{init}^f\right)}{1 - exp(-M^k)\left(1 - u_{init}^f\right)} + \omega^k \left(1 - exp(-M^k)\left(1 - u_{init}^f\right)\right)}{1 - \rho^k exp(-M^k)\left(1 - u_{init}^f\right)} \qquad (D.14)$$

Equilibrium recruitment R_{eq}^k is given by:

[8] This is obtained by solving $B_0^k = exp(-M^k)(\alpha^k N_0^k + \rho^k B_0^k) + \omega^k R_0^k$ for B_0^k / R_0^k, where $N_0^k = \frac{R_0^k}{1 - exp(-M^k)}$

$$R_{eq}^{k} = \frac{\phi_{eq}^{k} - \phi_{0}^{k}\frac{1-h^{k}}{4h^{k}}}{\frac{5h^{k}-1}{1-h^{k}B_{0}^{k}}\phi_{eq}^{k}} \qquad (D.15)$$

D.2 Objective Function

To aid in estimation of the model scale, a penalty was placed on the annual, fishery-specific harvest rate u_y^f. The penalizing function assumes that u_y^f can take on any value in the range $(0, \infty)$ but conditionally adjusts values of u_y^f greater than 0.4, by replacing u_y^f with $\frac{0.4}{2-u_y^f/0.4}$. When this occurs, the following penalty term is added to the objective function:

$$Penalty_u = \begin{cases} 0 & if\ u_y^f <= 0.4 \\ \sum_{y,f} 0.01\left(u_y^f - 0.4\right)^2 & if\ u_y^f > 0.4 \end{cases} \quad (D.16)$$

Estimates of instantaneous fishing mortality F from recent regional stock assessments have rarely been greater than 0.15yr^{-1} over the last 20 years, so this penalty is not expected to be unreasonably restrictive (note that the definition of F in the regional stock assessments is not directly equivalent to U in the estimation method). The biomass corresponding to survey f is the sum of estimated biomass over the spatial areas covered by survey f. Estimates of relative abundance (biomass) from each survey are fit under the assumption that the indices are lognormally distributed. Catchability, q, is time-blocked for the West Coast VAST surveys from 1960-1995 (due to a change in survey timing because the survey started earlier in the year after 1995) and from 2003 onwards (due to a switch from the Triennial to the Bottom Trawl surveys).

$$L_{surv} = \sum_y \sum_f \left(0.5\ell n\left[\left(\sigma_y^f\right)^2\right] + \frac{1}{2\left[\left(\sigma_y^f\right)^2\right]}\left(\ell n(B_{Obs,y}^f / B_{Pred,y}^f) - \ell n\left(q_{y*}^f\right)\right)^2 \right) \quad (D.17)$$

q is calculated analytically and given time-blocking in the historical period as in the OM:

$$\ell n\left(q_{y*}^f\right) = \frac{\sum_{y \in y*} \ell n(\frac{B_{Obs,y}^f}{B_{Pred,y}^f}) / \left[\left(\sigma_y^f\right)^2\right]}{\sum_{y \in y*} 1 / \left[\left(\sigma_y^f\right)^2\right]} \quad (D.18)$$

where $B^f_{Obs,y}$ is the observed index of relative biomass for survey *f* and year *y*; $B^f_{Pred,y}$ is the EM-projected exploitable biomass corresponding to the index for survey *f* during year *y*; q^f_{y*} is the catchability coefficient for fishery *f* during the block of years defined by the set *y**; and σ^f_y is the observation error standard deviation for survey *f* and year *y*. Finally, a penalty is added to the objective function for q^f_{y*} to prevent unreasonably small exploitable biomass in spatial areas, particularly those that are isolated by movement and demography (e.g., C1 and A4):

$$Penalty_q = \begin{cases} 0 & if\ q^f_{y*} <= 3 \\ \sum_{y \in y*,f} 0.01\left(q^f_{y*} - 3\right)^2 & if\ q^f_{y*} > 3 \end{cases} \quad (D.19)$$

The recruitment deviations are estimated as random effects, and penalized to be normally distributed and to sum to zero:

$$L_{Rec} = \frac{1}{2}\left(\sum_y \frac{\widetilde{R^2_y}}{\sigma^2_R} + ln(\sigma^2_R)\right) - \frac{\sum_y \widetilde{R^2_y}}{0.01} \quad (D.20)$$

The penalty on the partitioning of recruits within stocks comprised of two spatial areas, τ^{ik}, is based on a beta-distribution parameterized using survey observations (from the operating model):

$$L_\tau = \sum_{ik} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \tau^{ik}\left(1-\tau^{ik}\right)^{\beta-1} (D.21)$$

where $\alpha = \left(\frac{(1-\mu)}{0.01} - \frac{1}{\mu}\right)\mu^2$, $\beta = \alpha\left(\frac{1}{\mu} - 1\right)$ and μ is the mean ratio of observed survey biomass in constituent areas *i* for the final 10 years in the historical period: $\left(\sum_{2010}^{2019}\sum_{i \in k} \frac{B^i_{Obs,y}}{B^k_{Obs,y}}\right)/10$. Ratios from the survey data suggest that most recruits move out of British Columbia, which is consistent with the movement estimates from the analysis of tag-recapture data. Specifically, for the six-area, four-stock EMs, the prior mean for recruitment apportioned northward from the North BC-Gulf of Alaska stock into the Gulf of Alaska was typically 80%; the prior mean for recruits apportioned southward from the South BC-North US West Coast stock into the Northern US West Coast was typically 90% (Figure D.7). For the six-area, three-stock models (where stocks are coincident with management regions), the prior means were between 75%-85% for northbound recruitment within the U.S. West Coast (to the area north of 36°N) and within Alaska (from the Gulf of Alaska into

the Bering Sea) and roughly 50% for the two British Columbia areas. Note that estimation of τ^{ik} is never conducted simultaneously with the estimation of movement rates, as these processes are confounded.

D.3 Tables

Table D.1: Percentage of replicates for which the Hessian matrix was positive definite by projection year and estimation method. The mean column indicates the average across all projection years.

Management Strategy	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	mean
EM 1, Status Quo HCR	100.00	99.06	100.00	99.06	99.06	95.28	95.28	95.28	94.34	94.34	94.34	94.34	94.34	94.34	92.45	94.34	96.23	95.28	95.28	96.23	97.17	96.00
EM 2, Status Quo HCR	71.70	69.81	70.75	72.64	77.36	79.25	75.47	78.30	78.30	82.08	78.30	79.25	79.25	76.42	70.75	80.19	78.30	77.36	76.42	76.42	75.47	76.37
EM 3, Status Quo HCR	11.32	15.09	3.77	2.83	1.89	6.60	7.55	14.15	10.38	9.43	16.04	8.49	6.60	7.55	11.32	10.38	10.38	10.38	6.60	13.21	15.09	9.48
EM 4, Status Quo HCR	99.06	100.00	100.00	100.00	100.00	100.00	100.00	100.00	98.11	97.17	95.28	99.06	100.00	99.06	100.00	98.11	100.00	100.00	100.00	100.00	99.06	99.28
EM 5, Status Quo HCR	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Table D.2: Scaled median (across replicates) scores for the biological performance metrics (columns) for the seven management strategies (rows) and four stocks (panels). Scores have been rounded to two significant digits. The colors indicate the relative performance within each management region-performance metric combination, with lighter shades indicating low-performing scores and darker shades indicating high-performing scores.

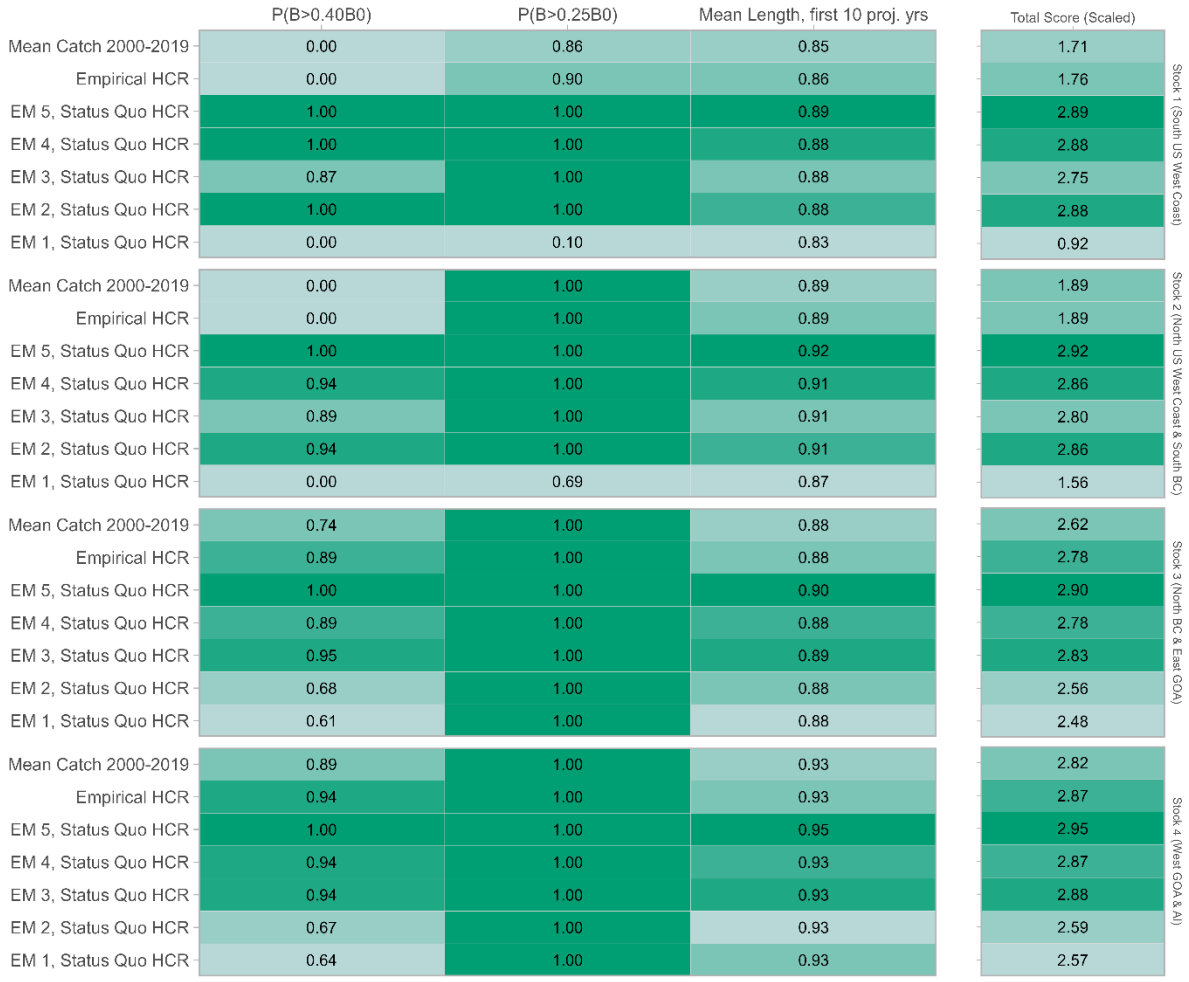


Table D.3: Scaled median scores for the economic performance metrics (columns) for the seven management strategies (rows) and three management regions (panels). The colors indicate the relative performance within each management region-performance metric combination, with lighter shades indicating low-performing scores and darker shades indicating high-performing scores.

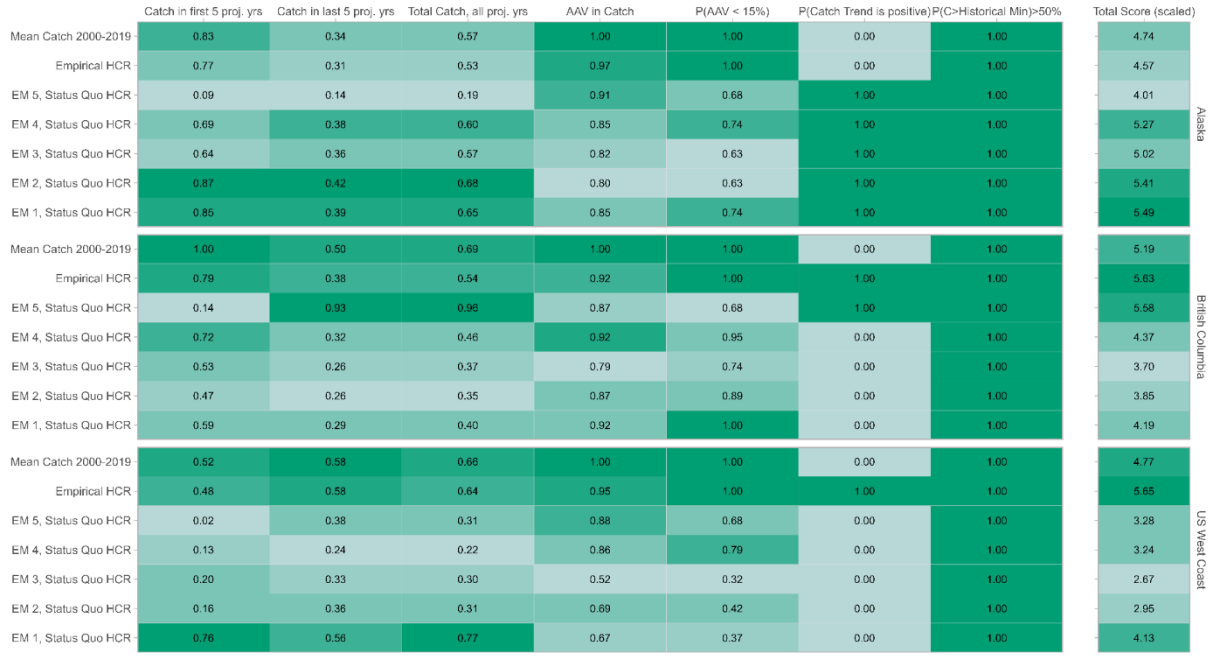


Table D.4: Median reference points (all years, all replicates) by management strategy (rows) and management region (columns). Values have been rounded to two significant digits. Bolded values are those used in the MSE. *British Columbia applies a fixed target harvest rate of 0.055. For the panmictic strategy (rightmost column) the applicable reference points are used to calculate TACs for each region.

		Management Strategy		
Reference Point	Management Region	1	2,3,4	5
U_{MSY} (Equation 4.1)	Alaska	0.2	0.21	
U_{MSY}	British Columbia	0.17	0.18	
U_{MSY}	US West Coast	**0.14**	**0.12**	
U_{MSY}	Panmictic			0.17
$U_{40\%}$ (Equation 4.2)	Alaska	**0.15**	**0.16**	
$U_{40\%}$	British Columbia	0.13	0.13	
$U_{40\%}$	US West Coast	0.11	0.1	
$U_{40\%}$	Panmictic			0.13
B_{MSY} (t)	Alaska	73,600	76,000	
B_{MSY}	British Columbia	**15,200**	**15,500**	
B_{MSY}	US West Coast	**47,400**	**50,800**	
B_{MSY}	Panmictic			146,000
$B_{40\%}$ (t)	Alaska	**96,500**	**99,900**	
$B_{40\%}$	British Columbia	19,700	20,000	
$B_{40\%}$	US West Coast	58,400	63,000	
$B_{40\%}$	Panmictic			189,000

Table D.5: Parameter symbols, definitions, and treatment in the estimation method.

Parameter Symbol	Definition	Treatment
Model Structure		
M	Modeled management region; largest spatial unit at which HCRs are applied	Three management regions correspond to Alaska (AK), British Columbia (BC), and the US West Coast (WC)
K	Modeled stocks; geographic areas of shared demography	Varies by EM: 1, 3 or 4
I	Modeled spatial area; smallest unit at which abundance and biomass are represented	Varies by EM: 1, 3 or 6
F	Fishery or survey	9 fisheries and 6 surveys during the historical period; 9 fisheries and up to 5 surveys during the projection period (aggregated to 3 or 1 surveys for EMs 3-5).
$\mathbf{\Lambda}$	Matrix indicating whether spatial area *i* is nested within stock *k*	
Parameterization		
B_y^i	Biomass (mt) in spatial area *i* at start of year *y*	
N_y^i	Abundance (numbers) in spatial area *i* at the start of year *y*	
M^k	Instantaneous natural mortality rate for stock *k*	Set to values from the OM (ranging from 0.07 to 0.11yr^{-1})
$\mathbf{X}$	Age-based movement matrix	When movement specified, set to values from Rogers et. al. (in prep)
ω^k	Weight of recruits in stock *k*	Estimated using data from OM (see Table D.6)
α^k	Coefficient governing asymptotic maximum weight in stock *k*	see Table D.6
ρ^k	Ford–Walford growth coefficient in stock *k*	see Table D.6
Recruitment		
R_y^k	Recruitment to stock *k* at the start of year *y*	Calculated from Equation D.8
h^k	Steepness of the stock recruitment curve (expected proportion of R_0 at $0.2B_0$) for stock *k*	Set to 0.7 for all stocks
$\widetilde{R}_y^k$	Annual recruitment deviations for stock *k*	Estimated as random effects; normally distributed with mean zero and standard deviation σ_R
$\widetilde{R}_{init}^k$	Initial recruitment deviation for stock *k*	Estimated
σ_R	Standard deviation of recruitment deviations	Estimated for EM1 and fixed otherwise (1.26)
R_0^k	Unfished recruitment for stock *k*	Estimated
τ_y^{ik}	Proportion of recruits in year *y* spawned in stock *k* allocated to spatial area *i; i* $\in k$	Estimated for EMs where stocks are comprised of more than one spatial area; time-invariant with beta prior based on relative survey biomass from 2010-2019 (Equation D.21)
Fisheries and surveys		

u_y^f	Harvest rate by fishery *f* during year *y*	Calculated as the ratio of the catch to the biomass during year *y* available to fishery *f*; penalized when above 0.4
σ_y^f	Standard error of log-survey biomass for survey *f* during year *y*	External estimate of observation error CV, as used in the OM
q_{y*}^f	Survey catchability coefficient for survey *f* during year block *y**	Calculated analytically; penalized when above 3 (Equation D.19)

Table D.6: Pre-specified growth parameters by spatial scenario. These values were estimated using length-at-age and weight-at-length data and parameters from the operating model.

Stock	**α (kg)**	**ρ (kg y^{-1})**	**ω (kg)**
EM 1 (Demographic stocks like the OM)			
1 (US West Coast south of 36°N)	0.507	0.775	1.810
2 (US West Coast north of 36°N and south BC)	0.717	0.817	2.761
3 (north BC and Eastern Gulf of AK)	1.847	0.583	4.378
4 (western GOA)	1.396	0.708	4.364
EM 2-4 (Stocks = management regions)			
US West Coast	0.561	0.803	2.096
British Columbia	0.992	0.742	3.403
Alaska	1.634	0.647	4.457
EM 5 (panmixia)			
Entire domain	1.090	0.726	3.549

D.4 Figures

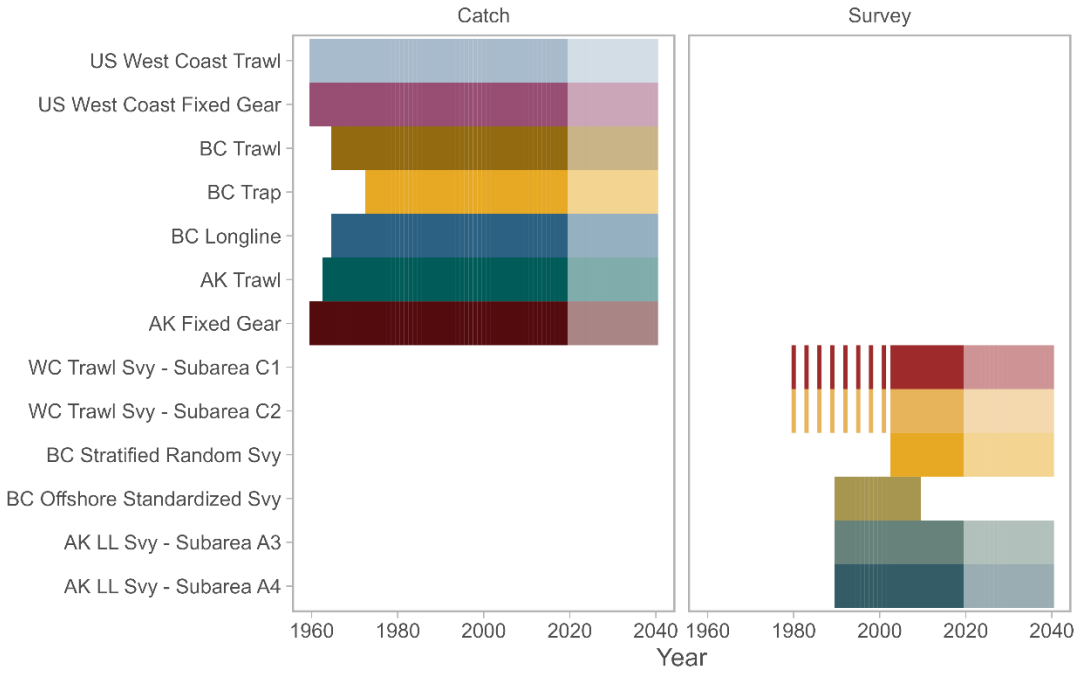


Figure D.1: Years, fisheries, and data types available to the estimation methods. Catches are provided by the management strategy for every year of the projection period (faded colors). For the U.S. West Coast and Alaska survey data are simulated annually, with observation error CVs equal to that for 2019. For British Columbia, survey data are simulated annually for the BC Stratified Random Survey, with observation error CVs equal to that for 2019 (the BC Offshore Standardized Survey was discontinued in 2010).

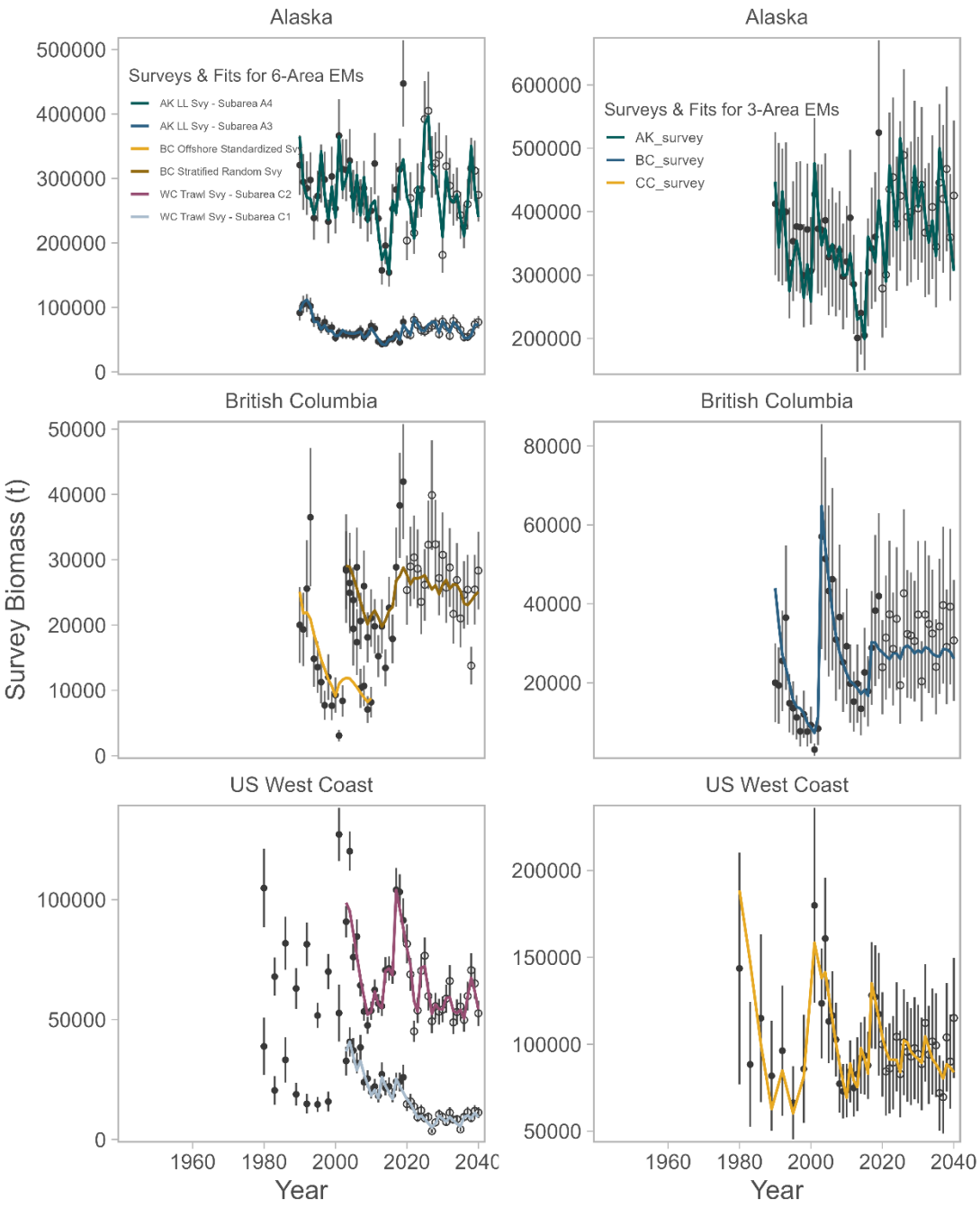


Figure D.2: Typical survey fits for the six-area (left column) and the three-area (right column) estimation methods.

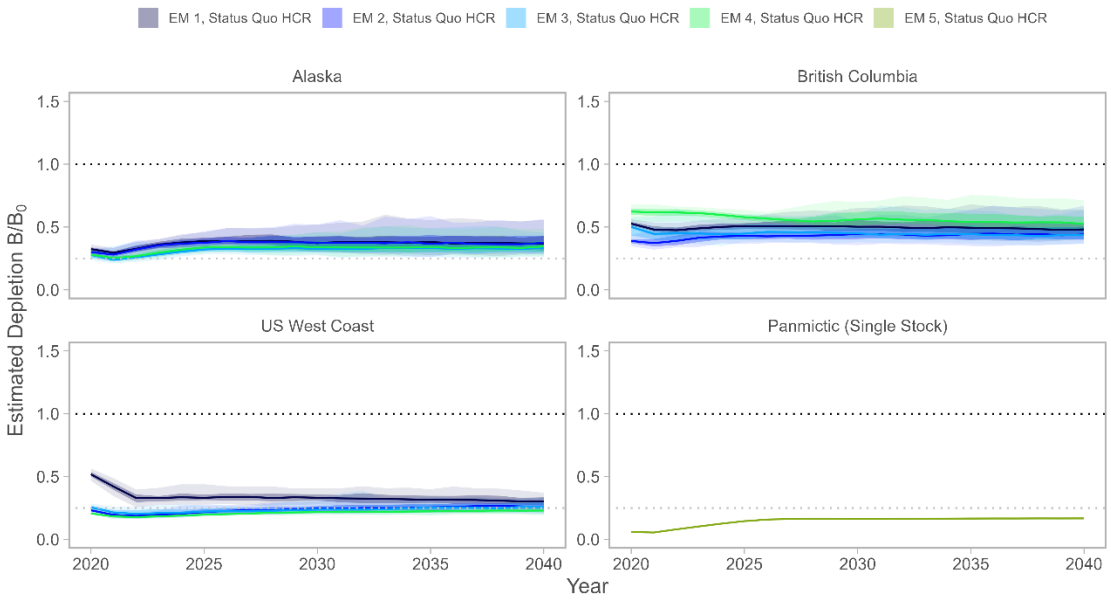


Figure D.3: Estimated depletion for each year based on the assessment conducted in that year for each of the five spatially-structured estimation methods (colors) for management regions (panels) during the projection period. Horizontal dotted lines indicate unfished biomass (dark dotted lines) and 25% thereof (light dotted lines). The solid line is the median, the darker shaded area represents the 25th to 75th quantile, and the lighter shaded area represents the 5th and 95th quantiles for 100 replicates.

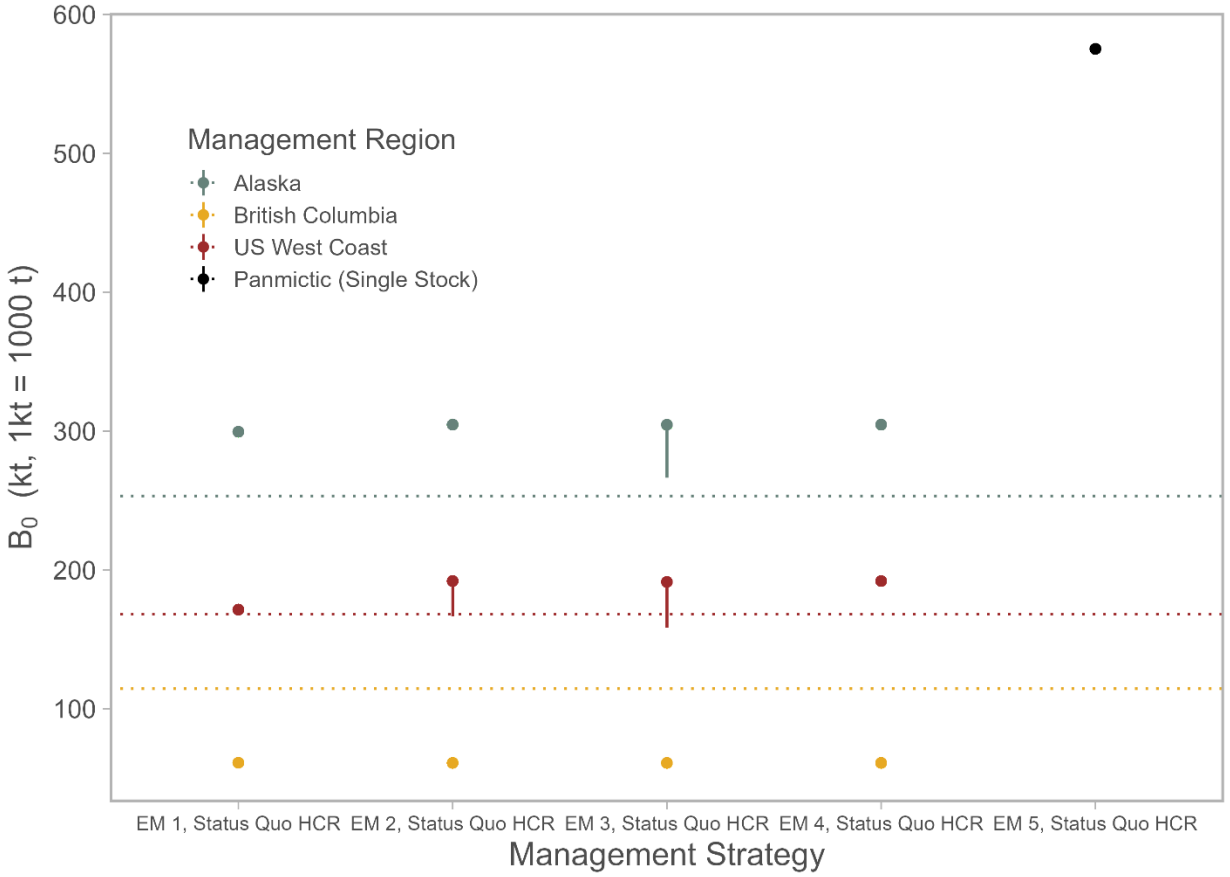


Figure D.4: Estimates of unfished biomass (points) for each management region (colors) by management strategy. The horizontal lines indicate the 5th and 95th quantiles over all replicates, but are not visible for most strategies. Unfished biomass from the OM is shown as a horizontal dotted line.

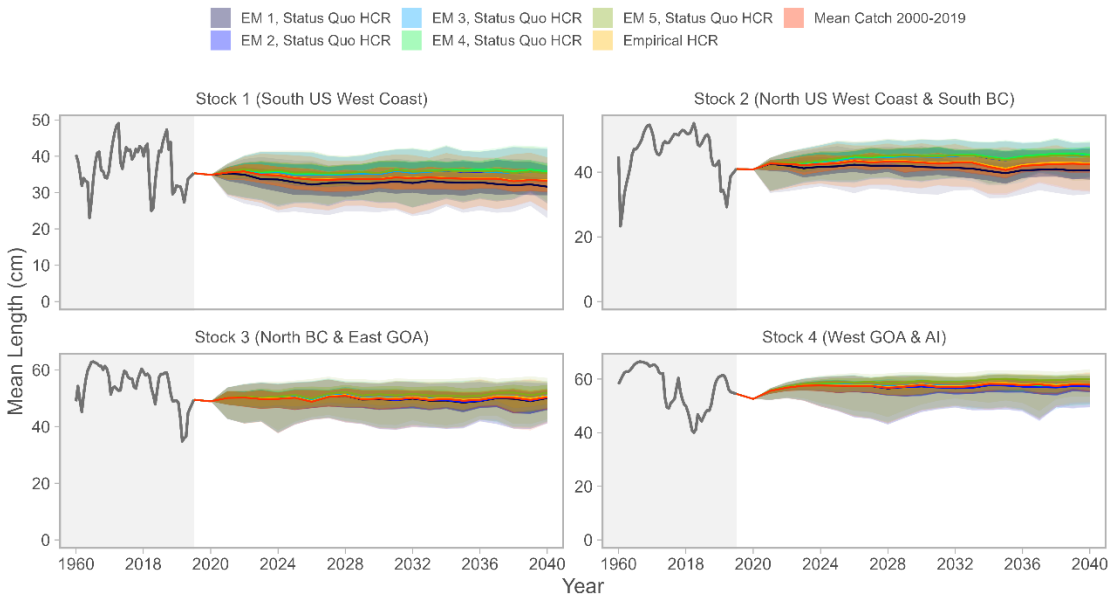


Figure D.5: Operating model trajectories of mean length for the four biological stocks (panels) for the seven management strategies (colors). The x-axis is condensed during the historical period (1960-2019, grey rectangle). The solid line is the median, the darker shaded area is the 50% simulation interval, and the lighter shaded area is the 90% simulation interval.

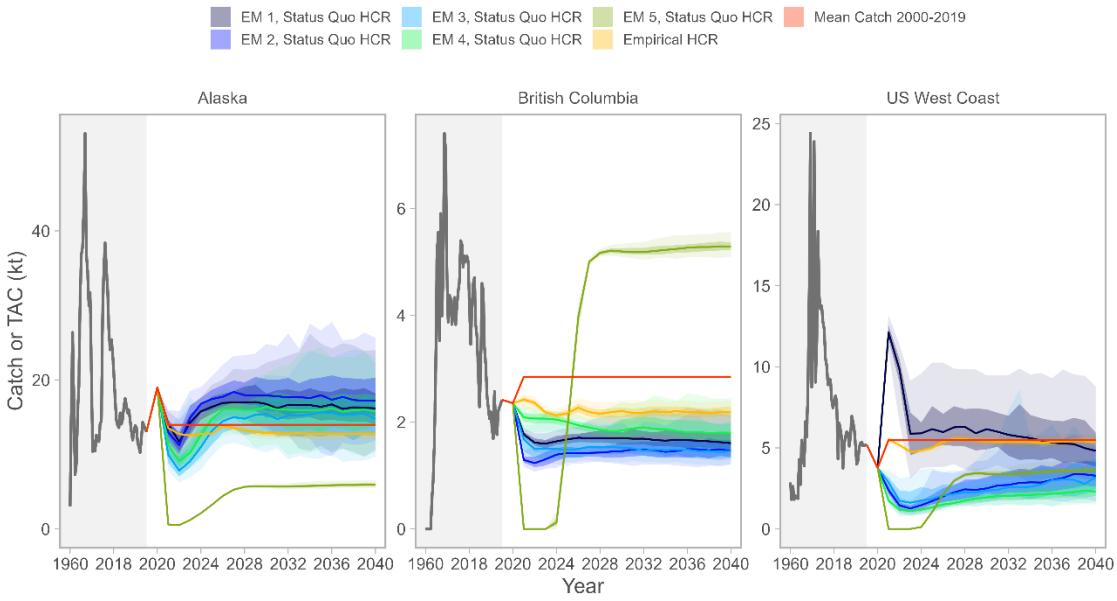


Figure D.6: Operating model time-trajectories of catch for the three management regions (panels) for the seven management strategies (colors). The x-axis is condensed during the historical period (1960-2019, grey rectangle). The solid line is the median, the darker shaded area is the 50% simulation interval, and the lighter shaded area is the 90% simulation interval. Actual observed catches are used for year 2020.

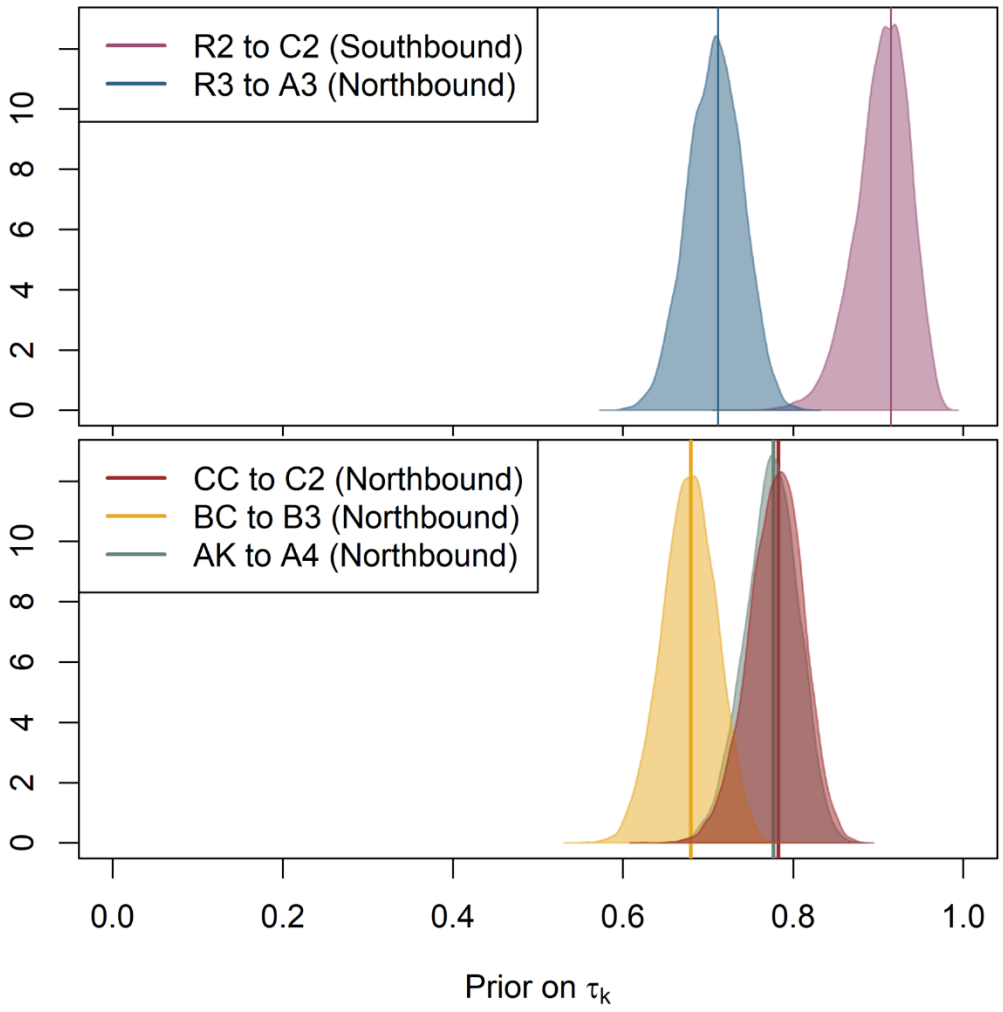


Figure D.7: Characteristic priors (shaded distributions) and posterior means (vertical lines) for the proportion of recruits from multi-area stocks apportioned to member spatial areas. (Top) priors for apportionment of recruits from Stock 3 into the Gulf of Alaska (blue), and from Stock 2 into the US West Coast north of 36^0N (pink). (Bottom) Priors for the apportionment of recruits from the US West Coast to the spatial area north of 36^0N (red), from Alaska to the Bering Sea (green), and from BC to northern BC (yellow).

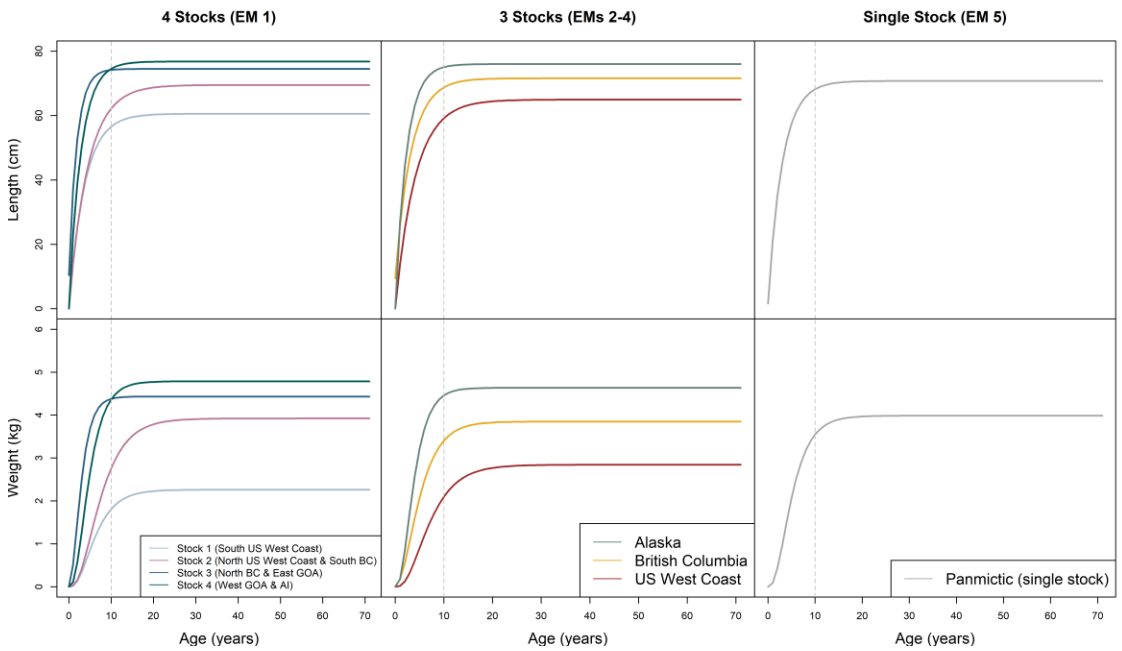


Figure D.8: Von Bertalanffy growth curves (top row) and weight-at-age (bottom row) estimated from OM length-at-age data for various stock structures. The vertical dashed line corresponds to age 10.