Title

Addressing complex fleet structure in fishery stock assessment models: Accounting for a rapidly developing pot fishery for Alaska sablefish (*Anoplopoma fimbria*)

Authors

Matthew L.H. Cheng[1], Daniel R. Goethel[2], Curry J. Cunningham[1]

[1] Department of Fisheries at Lena Point, College of Fisheries and Ocean Sciences, University of Alaska Fairbanks, 17101 Point Lena Loop Rd, Juneau, Alaska 99801, USA

[2]National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Alaska Fisheries Science Center, Auke Bay Laboratories, 17109 Point Lena Loop Road Juneau, Alaska 99801, USA

Abstract

Fisheries management operates under uncertainty, often driven by the dynamic nature of marine ecosystems and associated fisheries. Stock assessment models, which form the scientific basis of decision-making in fisheries management, strive for realistic representations of biological and fishery processes. However, data limitations and knowledge gaps necessitate simplifying assumptions for representing these complex bio-socioeconomic systems, which can increase uncertainty in the assessment process. Addressing time-varying fishery dynamics (i.e., due to regulatory changes or alterations in harvester behavior) is a common and particularly challenging problem for stock assessment models. Time-varying fishery selectivity is widely utilized to address changes in fishery dynamics but may not be adequate when regulatory changes substantially alter gear usage and associated assessment fleet structures. We explore the implications of accounting for, or ignoring, complex temporal changes in fleet structure and selectivity within stock assessment models by utilizing a recent and high-value case study, Alaska sablefish (*Anoplopoma fimbria*). Our findings demonstrate that the treatment of fleet structure (i.e., adding fleet complexity to account for gear transitions) did not greatly influence estimates of spawning biomass trajectories. However, associated selectivity assumptions had substantial impacts on sustainable harvest recommendations. We recommend that the treatment of fleet structure and associated selectivity assumptions should incorporate *a priori* considerations and subject-matter expertise of fishery and biological dynamics to ensure pragmatic and appropriate model parameterizations. Moreover, we advocate for multi-fleet models as a useful diagnostic tool for validating model estimates from single-fleet assessments when uncertainty in fleet dynamics exist.

Keywords: fishery dynamics, fishery selectivity, fleet structure, stock assessment, fisheries management, catch-per-unit effort (CPUE)

Introduction

Stock assessment models form the scientific basis for management advice for many species globally by providing estimates of current stock status and trends, which are then used to project sustainable harvest levels given management reference points. Contemporary assessment models commonly integrate a variety of data sources (i.e., fishery-dependent and fishery-independent) and types (i.e., age-and length-compositions, catch, and effort) into a single 'integrated analysis' (Maunder and Punt, 2013), which help inform important biological quantities (e.g., growth and natural mortality), recruitment processes, and the impact of fishery removals. These biological and fishery processes are often influenced by the dynamic nature of marine ecosystems and the fisheries that operate within them, resulting in the need for assessments to adequately incorporate temporal changes in modeled dynamics (Hilborn, 2003). However, the incorporation of temporal dynamics in stock assessment models are often hindered by considerations regarding model parsimony and data availability. Consequently, it may not be feasible to incorporate temporal variation in all modelled processes.

Temporal changes in fishery dynamics are common and are influenced by technological developments, economic conditions, management regulations and ecosystem factors (Beverton and Holt, 1957; Sainsbury, 1984; Eigaard *et al.*, 2014; Martell and Stewart, 2014). For instance, low catch, reduced economic efficiency, and failure to meet market demands for herring (*Clupea harengus*) fisheries in the United Kingdom facilitated the transition to purse seining and trawling as an alternative to drift-nets during the 1960s (Whitmarsh *et al.*, 1995). Similarly, ecosystem considerations, such as wildlife conflicts, have prompted changes in fishery dynamics (e.g., gear modifications) as a result of socioeconomic and species conservation concerns. In particular, entanglement and depredation (i.e., predators damaging fish or fishing gear) events have resulted in gear modifications (e.g., stronger net material, new gear types; Pol and Carr 2000; Tixier *et al*., 2021). These changes in fishery dynamics due to alterations in gear usage can influence how fishery selectivity and fleet structure are represented in stock assessment models and need to be adequately addressed to enable robust estimates of sustainable harvest levels (Sinclair 1993; Goodyear 1996; Maunder 2002; Martell and Stewart 2014).

In stock assessment models, fishery selectivity is one of the key components representing the impact of fishery removal processes and is commonly defined as the relative probability of capturing an individual as a function of its size, length, or age. Selectivity as defined in stock assessments is mediated by the combination of the following two processes: 1) Contact selection – the probability of capturing an individual if it comes into contact with fishing gear and 2) Availability – the probability that individuals occupy the same area and time during fishing activities (Sampson 2014). Parameterizing selectivity within stock assessment models is challenging because the true underlying functional form is often unknown (Punt *et al.*, 2014), and mis-specifying the selectivity process can substantially impact estimates of management reference points and absolute abundance (Goodyear, 1996; Scott and Sampson, 2011). Selectivity can be approximated using a variety of functional (e.g., asymptotic or dome-shaped) or non-parametric (e.g., splines; Martell and Stewart 2014) forms but must be carefully considered given implications for the estimation of other model parameters (e.g., confounding with natural mortality; Thompson 1994). Furthermore, selectivity often varies as a function of time (Sampson and Scott, 2012), which can be represented using time-blocks (i.e., selectivity is constant within a given block), while penalized maximum likelihood or state-space approaches can be used to represent smooth selectivity transitions (Nielsen and Berg, 2014). Ignoring time-varying processes in fishery selectivity can potentially result in consistent directional bias in model estimates (e.g., spawning stock biomass), which has been demonstrated in various simulation studies (Linton and Bence, 2011; Martell and Stewart, 2014; Szuwalski *et al.*, 2018). However, selectivity is assumed to be time-invariant in many integrated assessment models, often due to data limitations, parameter estimability, model complexity, and considerations regarding over-fitting (Maunder and Punt, 2013; Punt *et al.*, 2014; Punt, 2023). Therefore, balancing the potential biases associated with ignoring temporal variation in fishery dynamics (i.e., selectivity) must be carefully considered against the aforementioned factors. Furthermore, careful consideration should be given to the potential for increased uncertainty in estimated parameters as sample sizes decrease with concomitant increases in modelled dimensions.

Similar to the treatment of selectivity, assumptions about fleet structure within stock assessments can also have important implications for the reliability of estimated management quantities. Fleets within stock assessments can be aggregated or disaggregated by spatial units, sectors, or gear types, depending on characteristics of availability and removal processes. The treatment of fleet structure in integrated stock assessment models are well-studied in the context of representing spatial dynamics (Cope and Punt, 2011; Berger *et al.*, 2012; Hurtado-Ferro *et al.*, 2014; Waterhouse *et al.*, 2014), and parsimonious spatial fleet structure can be determined using multivariate regression trees to identify differences in catch-rates and age or length-compositions (Lennert-Cody *et al.*, 2010, 2013). However, less guidance exists on the relative importance of modeling the full diversity of fishery gear types (e.g., how to determine the number of unique fishery fleets to model) or the consequences of combining data across gears to represent a single fleet within assessment models. Single-fleet models can be practical if all fishery gear types share similar patterns in harvest size or age (e.g., Nielsen *et al.*, 2021). However, if the magnitude of catch and patterns in size or age of removals differ largely between gears but are modeled as a single aggregated fleet, removal processes within the assessment can be misrepresented (e.g., the misallocation of mortality to certain age or length classes). Punt *et al*. (2014) demonstrated that ignoring fleet structure (i.e., aggregating data and model dimensions across multiple sectors) resulted in important differences in recommended harvest levels (650 tons) for a pink ling (*Genypterus blacodes*) stock assessment.

Several benefits can be envisioned from disaggregating fleet structure of the fishery by gear type in an assessment, which include enhanced model diagnostics, better representations of age- or size-based fishery removal processes, and an improved reflection of local ecological knowledge. By disaggregating fishery fleets in the assessment, data conflicts across gear types can also be more clearly detected by inspecting residuals with an improved resolution (i.e., disaggregated by fishery fleet), which can facilitate model refinements through an enhanced understanding of explicit drivers of model instability (Punt *et al*., 2014). In a similar vein, fleet-specific models can be useful tools for validating or identifying structural uncertainty in single-fleet approaches if model results are inconsistent (Nielsen *et al*., 2021). By more accurately representing the actual gears used in the fishery, fleet disaggregation by gear type can also facilitate stakeholder acceptance and trust of model results because the assessment model better represents their empirical observations. With respect to catch projections, fleet-specific scenarios can aid in developing advice that can inform allocation decisions regarding fishing effort or recommended harvest levels and can facilitate the development of more robust management procedures (Bastardie *et al.*, 2010a, 2010b; Baudron *et al.*, 2010; Pascoe *et al.*, 2010). Beyond the benefits provided for stock assessment models, investigations into fishery fleet structure can enhance the understanding of harvester behaviors (Andersen *et al.*, 2012), aid in the estimation of fishery discards and subsequent measures for mitigating fishery discards (Fernández *et al.*, 2010; Holmes *et al.*, 2011), and facilitate ecosystem-based fisheries management (Gascuel *et al.*, 2012; Ulrich *et al.*, 2012).

In this study, we examine the treatment of fleet structure in the context of rapidly changing gear usage within the Alaska sablefish (*Anoplopoma fimbria)* fishery by exploring the implications of explicitly modeling, or ignoring, changes in fishery dynamics in the associated stock assessment model. Alaska sablefish are a deep-dwelling species that exhibit high movement rates and ontogenetic migration patterns. Juvenile sablefish typically migrate from nearshore to deeper offshore areas with adults typically inhabiting depths much deeper than 200m (Hanselman *et al.*, 2015; Goethel *et al.*, 2021). The sablefish fishery is one of the most economically valuable groundfish fisheries in Alaska (e.g., providing $94 million in ex-vessel value in 2015; Fissel *et al.*, 2016; Hanselman *et al*., 2019), and transitioned from an open-access fishery to an Individual Fishing Quota (IFQ) system in 1995, which greatly increased fishery catch rates and reduced harvest of immature females by 17% and immature males by 11% (Sigler and Lunsford, 2001). Prior to 2017, the sablefish fishery was prosecuted primarily using hook-and-line gear across the Gulf of Alaska, with a small portion of the fleet using pot-gear (rigid pots) in the Bering Sea and Aleutian Islands region. However, increases in sperm whale (*Physeter macrocephalus*) depredation events across the central and eastern Gulf of Alaska resulted in substantial economic loss for harvesters (Peterson *et al.*, 2014), which prompted interest in using pot-gear to mitigate depredation events. In 2017, pot-gear was legalized in the Gulf of Alaska for use in the directed fishery as an alternative to hook-and-line gear (Hanselman *et al.*, 2018), and removals from the pot fishery have since increased substantially in the region (Fig. 1). Specifically, pot-gear represented about 4% of the average total harvest of sablefish from 2010 to 2017 but rapidly increased to 55% of the total harvest in 2021.These dramatic increases in the use of pot-gear have been facilitated by the introduction of alternative pot-gear configurations in 2019 (e.g., collapsible "slinky" pots; Goethel *et al.*, 2020; Sullivan *et al.*, 2022), providing a more space efficient alternative to traditional rigid pots. In particular, slinky pots are collapsible and lightweight and allow for pot-gear fishing to be more accessible for smaller vessels limited by on-deck storage capabilities (Sullivan *et al.*, 2022). However, the contact selection process of pot-gear within the sablefish fishery is not well established at present. For example, it remains unclear whether the entrance of pot-gear may constrain the entry of larger individuals. Moreover, the absence of regulations and data collected on the use of escape rings for pot gear in the federal sablefish fishery further compounds the difficulty in interpreting selectivity processes for the pot fleet. Nonetheless, limited investigations have demonstrated that the size-distributions of individuals captured using slinky pots are comparable to hook-and-line gear (Sullivan *et al.*, 2022). Furthermore, both gear types appear to be deployed at overlapping depths (200-1100m). In particular, hook-and-line gear are deployed fairly uniformly across the 300-750m range, while pot-gear deployments are more concentrated towards depths of 400-550m (Goethel *et al.*, 2023; Appendix 3E).

Given these rapid changes in fleet structure, in part facilitated by the development of new gear configurations, there is a need to evaluate the implications of alternative treatments of fleet structure in the assessment model to reflect fishery dynamics more realistically. Currently, the Alaska sablefish assessment implicitly accounts for the rapid transition in pot-gear through a selectivity time-block within the single, aggregated 'fixed gear' (i.e., combined hook-and-line and pot) fleet, but potential biases from assuming an aggregated fixed-gear fleet has yet to be explored. Using the Alaska sablefish stock assessment model, we compare different parameterizations of fleet structure by fishery gear to evaluate the impact of alternative methods for representing rapid changes in fishery dynamics on model estimates and resulting management advice. Specifically, we investigate the implications of: 1) the addition of a new pot fleet, 2) associated pot fleet selectivity parameterizations, and 3) the use of either an aggregated index (combining hook-and-line and pot-gear) or a fleet-specific index. We seek to provide pragmatic guidance on the treatment of fleet structure when multiple gear types exist, especially when new fisheries or gear types emerge rapidly with a limited time series of data to inform associated parameter estimation.

Methods

In this study, the 2021 Federal sablefish operational assessment model (with minor modifications) was evaluated against model variants representing alternate assumptions about fishery dynamics to understand the implications of disaggregating modelled fishing fleet(s) based on gear type. Sablefish in Alaska federal waters are assumed to represent a single reproductive population but with sex-specific growth and selectivity processes (Goethel *et al.*, 2021). Furthermore, the assessment uses an integrated statistical catch-at-age framework that treats the hook-and-line and pot-gears as a single fixed-gear fleet (i.e., catch along with age- and size-composition data are aggregated), with a nominal fishery-dependent catch-per-unit effort (CPUE) index (i.e., using only data from the hook-and-line gear) also fit (Fig. 2). To evaluate the treatment of fleet structure on the sablefish stock assessment, we used gear-disaggregated data to enable modeling the hook-and-line and pot-gears as separate fishery fleets within the stock assessment model. A unique trawl fishery fleet is also explicitly modelled in the operational assessment, however the structure of the trawl fleet was not altered in any model runs. The trawl fishery fleet composes about 20% of total removals on average (Goethel *et al.*, 2022a). Following good practices for using fishery-dependent indices in stock assessments, standardized fleet-specific indices based on fishery catch-per-unit effort (CPUE) data were also developed for model runs in this study (Hoyle *et al.*, 2024), given the use of a nominal hook-and-line gear fishery-dependent index in the operational assessment. To explore alternative approaches to fleet specification, three axes of comparison were explored in this study, including:

1) How the stock assessment model was parametrized to address fishery fleet structure (i.e., whether or not pot and hook-and-line fleets were aggregated).
2) How the fishery CPUE index accounted for fishery fleet structure (i.e., CPUE index that aggregates both hook-and-line and pot-gear, or disaggregating by gear and developing separate fishery-dependent indices).
3) How pot-gear selectivity was parametrized.

The sablefish case study is a useful example for fisheries that are experiencing rapid and abrupt shifts in fishery fleet structure and provides practical guidance and considerations for assessment model parameterizations. Here, we present a brief synopsis of the Alaska sablefish operational assessment model. We then identify how data from each gear type are analyzed (including the development of standardized CPUE indices by gear type), highlight key structural updates to the operational model necessary for the incorporation of a pot fishery fleet, and conclude by describing metrics used to compare performance of model parameterizations. Parameters across all model runs were estimated using penalized maximum likelihood estimation by minimizing an objective function, which consisted of likelihood components for catch (lognormal likelihood), abundance/biomass indices (lognormal likelihood), compositional data (multinomial likelihood), and priors and penalties (recruitment, natural mortality, fishing mortality).

Operational Assessment

As described above, the operational assessment model is an age-and sex-structured integrated assessment model assuming a panmictic population and is developed in AD Model Builder (Fournier *et al.*, 2012). The assessment model assumes mean recruitment where recruits enter into the population at age two, with annual recruitment deviations estimated (i.e., assuming a penalized likelihood and a recruitment standard deviation term that is fixed at 1.2). Cohorts by age are tracked over time following an exponential mortality model, where natural mortality is estimated with an informative prior. The general model structure can be found in Goethel *et al*. (2021) and is also provided in Appendix B. The operational assessment model integrates catch data, abundance/biomass indices, and compositional data (age and length) from both fishery-independent and -dependent sources to estimate past demographic trends, quantities of interest (e.g., biomass levels), and biological reference points (e.g., Acceptable Biological Catch, 40% of unfished SSB; *B40%*). Population trends are primarily informed by the Alaska Longline Survey conducted by the National Marine Fisheries Service (NMFS). Age-compositional data are input as sex-aggregated, which is the approach taken in the 2021 federal sablefish operational assessment due to sample size limitations, while length-compositional data are input as sex-specific (Goethel *et al.*, 2021). Additionally, age- and length-composition data for all fisheries and surveys follow multinomially distributed errors, with input sample sizes of 20 (i.e., the variance weighting parameter for the multinomial distribution). Model weights (applied to the aggregate dataset) are then determined using Francis-reweighting (Francis, 2011) and are used in the final model runs. Preliminary explorations indicated that model weights determined by Francis-reweighting were fairly insensitive to the assumed input sample sizes.

Removals from the fishery are currently represented by two unique fishery fleets: 1) a trawl fleet comprised mostly of incidental catch and 2) a directed fixed-gear fleet that reflects aggregated fishery dynamics of both the hook-and-line and pot-gears. Selectivity for the directed fixed-gear fleet assumes a logistic function where time-varying processes are represented by three time-blocks (i.e., selectivity is constant within a given block) to account for regulatory changes (i.e., the shift from open access to an IFQ fishery in 1995 and the allowance of pot-gear regulatory change in the Gulf of Alaska in 2017; Goethel *et al*., 2021). Note that the last time-block (2016+) accounts for both high recruitment events beginning in 2016 and the pot-gear regulatory shift in 2017. In the operational stock assessment, the directed fixed-gear fleet is fit to a CPUE index from 1990 to 2020 specific only to the hook-and-line fleet, which does not include fishery-dependent CPUE data from the pot fleet. However, the present study replaces the nominal index used in the operational assessment with a gear-aggregated standardized biomass index from 1995 to 2020 that combines both hook-and-line and pot-gear data, following the methods of Cheng *et al*., (2023a). This was done to facilitate comparisons among alternative model structures and to better address the rapid expansion of the pot-gear fishery. The operational assessment model updated with the gear-aggregated standardized biomass index served as the basis for comparison (i.e., the null model in this study) across alternative model runs and will be referred to as model *Combined-Logistic* hereafter.

Disaggregating Data from the Fixed-Gear Fleet

To incorporate a unique pot fleet and represent changes in sablefish fishery dynamics, we separated catch data and age- and length-composition data from the fixed-gear and pot fleets. As noted previously, pot fishing was permitted in the Bering Sea and Aleutian Islands region prior to 2017, while it was legalized in the Gulf of Alaska during 2017. Consequently, fishery data for the pot fleet have been collected since 1991, albeit in limited quantities and spatial coverage. Disaggregating data sources from the fixed-gear fleet resulted in a pot-specific catch time-series that ranged from 1991 to 2021, length-composition data that ranged from 1999 to 2021, and age-composition data that ranged from 2004 to 2021 (Fig 2). Both age- and length-composition data from the pot fleet include more breaks in the time-series relative to hook-and-line gear due to comparatively lower fishing effort and proportional sampling, resulting in limited sample sizes (Fig. 2). For age-composition data, years with observations that had less than 20 samples from a given gear type were removed (2014 and 2015), while data from years with length-composition data that had less than 100 samples were removed (2014 and 2015). This was done to ensure that compositional data were generally representative of removal processes in the pot fishery and were not derived from a limited number of sampling events. Sample sizes from the pot fishery only began increasing after the regulatory shift (due to an increased effort in the pot fishery), such that the pot fishery fleet had relatively lower sample sizes for compositional data prior to 2017.

Development of Fishery-Dependent Standardized Indices

Fishery-dependent biomass indices were developed using Generalized Additive Models following the methods described in Cheng *et al*. (2023a). As noted, the nominal index used in the 2021 operational sablefish assessment model was replaced by a gear-aggregated standardized biomass index. The overall interpretation of model results between model *Combined-Logistic* and the 2021 operational sablefish stock assessment remained similar. Relative to the nominal index (1990 to 2020), all standardized biomass indices developed in the current study omitted data prior to 1995 because of the shift towards an IFQ system, which resulted in large increases in catch efficiency (Sigler and Lunsford, 2001). Furthermore, all fishery-dependent standardized biomass indices explicitly incorporated spatial information (i.e., longitude and latitude) using tensor product smoothers. For the gear-aggregated biomass index, catch rate data from hook-and-line (effort = catch-per-hook) and pot (effort = catch-per-pot) fleet were combined between 1995 to 2020. Gear-disaggregated biomass indices were developed between 1995 to 2020 and 2003 to 2020 for the hook-and-line and pot fishery, respectively. The shorter time-series for the pot-specific index is attributed to removing years prior to 2003 that had low sample sizes and observer coverage. For the pot index, trends prior to 2017 are representative of the Bering Sea and Aleutian Islands region. Trends after 2017 are representative of both the Bering Sea and Aleutian Islands, and the Gulf of Alaska (i.e., consistent with the entire spatial extent of sablefish management). Model selection for index standardization model terms was conducted using 5-fold cross validation.

Assessment Fleet Structure and Selectivity

To accommodate the addition of a new pot fishery additional parameters had to be estimated, which include annual instantaneous fishing mortality rates (32 parameters: 1 parameter describing mean log fishing mortality and 31 independent annual deviations from the mean) for the pot fleet and associated sex-specific selectivity parameters (total number of estimated parameters depended on the specific functional form; $\leq$ 4 additional selectivity parameters). Although a variety of selectivity parameterizations for the pot fleet were initially explored (i.e., normal, exponential-logistic, double-logistic, double-normal), only two were retained based on explorations of model performance (including an invertible Hessian matrix, reasonable selectivity forms, and model performance). The first selectivity parametrization was the logistic function (model *Pot-Logistic*), which was time-invariant:

$$s_{a,s,f} = \left[1 + e^{-\delta_{s,f}\left(a - a_{s,f}^{50\%}\right)}\right]^{-1} \quad \text{(Eq. 1)}$$

where subscripts a, s, and f denote ages, sexes, and fleets. δ denotes the shape parameter of the logistic function and $a^{50\%}$ represents the age-at-50% vulnerability to fleet f. Sex-specific $\delta_{s,f}$ for the pot fleet were shared with the hook-and-line fleet during the 2016-2021 time-block. Parameter sharing was necessary because sex-specific shape parameters for the pot fleet were estimated at an upper bound due to model instability (resulted in knife-edged selectivity). The second selectivity parametrization was the re-parameterized gamma function (model *Pot-Gamma*; Punt *et al.*, 1996), which was also time-invariant:

$$s_{a,s,f} = \left(\frac{a}{a_{s,f}^{max}}\right)^{\left(\frac{a_{s,f}^{max}}{p}\right)} e^{\frac{a_{s,f}^{max}-a}{p}} \qquad \text{(Eq. 2.1)}$$

$$p = 0.5 * \left[\sqrt{a_{s,f}^{max^2} + 4\gamma_{s,f}^2} - a_{s,f}^{max}\right] \qquad \text{(Eq. 2.2)}$$

where γ is an estimated shape parameter that describes the steepness of the descending limb, p is a derived quantity representing the power parameter (not estimated), and a^{max} is an estimated parameter that describes the age-at-maximum selection.

Model Scenarios, Comparisons, and Performance

A total of three model variants were explored (Table 1):

1) An aggregated fixed-gear fleet structure assuming logistic selectivity, fit to a gear-aggregated standardized biomass index that combines catch rate and composition data from the hook-and-line and pot-gear (*Combined-Logistic*).
2) A disaggregated fixed-gear fleet structure assuming logistic selectivity for both the hook-and-line and pot fleet, fit to separate standardized biomass indices and composition data for hook-and-line and pot-gear (*Pot-Logistic*).

3) A disaggregated fixed-gear fleet structure assuming logistic selectivity for the hook-and-line fleet and gamma selectivity for the pot fleet, fit to separate standardized biomass indices and composition data for the hook-and-line and pot-gear (*Pot-Gamma*).

For models that were fit to gear-disaggregated biomass indices, a catchability time-block was imposed for the pot index in 2017 to account for the regulatory shift pertaining to pot-gear. Incorporating the catchability time-block is considered best practice for accounting for changes in gear-use and regulations in stock assessment models (Wilberg *et al.*, 2009). Preliminary explorations indicated that allowing for a catchability time-block allowed for improved model fits to the index.

Given that stock assessments often utilize different data sources and data weights (Maunder and Piner, 2017), it is difficult to objectively identify tradeoffs in model parsimony and model fit using commonly employed model selection methods (i.e., information criterion methods). Consequently, stock assessments often use a variety of diagnostic tools and subject-matter expertise to evaluate model fit, parsimony, and realism for determining optimal model structures (Carvalho *et al.*, 2021). Therefore, model performance was assessed by investigating common model diagnostics, and using subject-matter expertise to determine whether model estimates were reasonable given *a priori* knowledge of fishery and biological processes. Comparisons of important model outputs used for the basis of fisheries management decisions (i.e., biological reference points and projected harvest recommendations) were also explored to understand the implications of alternative treatments of fleet structure and selectivity.

Model adequacy and performance was based upon: 1) convergence diagnostics, 2) parameter correlations, 3) model fits to data, 4) retrospective patterns, and 5) likelihood profiles.

Convergence diagnostics included inspection of an invertible Hessian matrix and a maximum gradient component < 0.001 (Carvalho *et al.*, 2021). We also examined the matrix of parameter correlations for the presence of highly correlated parameter pairs > 0.95, which could be indicative of unstable and spurious model solutions (Carvalho *et al.*, 2021). One-step-ahead (OSA) residuals of compositional data for hook-and-line and pot fleets were inspected to evaluate potential misspecification of selectivity forms through the presence of systematic patterns (Thygesen *et al.*, 2017; Trijoulet *et al.*, 2023). Furthermore, to compare the average magnitude of residuals for a given composition type across models, a metric of mean absolute residuals was computed. Failure to account for time-varying processes and misspecification of selectivity forms can also manifest as retrospective patterns and may result in consistent inappropriate management advice (Linton and Bence, 2011; Martell and Stewart, 2014). To assess the direction and magnitude of retrospective inconsistencies across models, we conducted 3-year retrospective "peels" (i.e., data are sequentially removed and models are re-estimated for each truncated dataset) and computed Mohn's ρ for estimated spawning stock biomass (SSB) and fully-selected fishing mortality rates:

$$b_p = \left(\frac{X_{Y-y,p} - X_{Y-y,\mathrm{ref}}}{X_{Y-y,\mathrm{ref}}}\right) \qquad \text{(Eq. 3.1)}$$

$$\rho = \sum_{p=1}^{n} \frac{b_p}{n} \qquad \text{(Eq. 3.2)}$$

where b_p represents the relative retrospective inconsistency for "peel" *p, X* is the metric of interest, *Y* is the final year for a given projection, *y* is the last year of an assessment with fewer years of data used, and *ref* is the reference peel (the most recent assessment year). Mohn's ρ is then computed by taking the average relative inconsistencies across all peels. Positive values of Mohn's ρ represent positive inconsistencies in the estimated quantity, and vice versa. Considering the time-blocking model structures across all model variants, in addition to parameter sharing with the hook-and-line 2016-2021 time-block for model *Pot-Logistic*, larger data peels were not conducted for comparability purposes. Nevertheless, the retrospective performance for model variants across these three peels can still provide insight into model consistency and short-term retrospective behavior. Finally, to investigate the presence of conflicts among data sources and model consistency (Lee *et al.*, 2014), we constructed likelihood profiles for survey catchability (sablefish longline survey) and mean recruitment, both of which are key scaling parameters within the sablefish stock assessment. Likelihood profiles were constructed by incrementally increasing log survey catchability and log mean recruitment values across a fixed range. Large differences in negative log-likelihood values over small changes in parameter values are likely to be indicative of model misspecification, poorly parameterized model structures, or highly correlated parameter pairs (Punt *et al.*, 2014; Carvalho *et al.*, 2021).

To understand the implications of selectivity, fleet structure, and biomass indices on stock status, we compared differences in estimates of fully-selected fishing mortality rates, predicted recruitment, SSB trends and projections, and the ratio of SSB with the *B40%* reference point across models. Population projections were conducted by assuming mean recruitment, used fishery selectivity estimates from the most recent time block, and assumed a fishing mortality rate equal to *F40%.* Here, *F40%* is the fishing mortality rate that reduces the spawning biomass-per-recruit to 40% of the average unfished spawning biomass-per-recruit. Additionally, the ratio of SSB and *B40%* is the basis of the harvest control rule (sloping control rule; Deroba and Bence 2008) used to manage sablefish in Alaska that determines long-term sustainable harvest levels. When the ratio of terminal year SSB and *B40%* is above 1, harvest levels are increased to maintain the stock at the *B40%* target. In contrast, when this ratio is below 1, harvest levels are reduced to allow the stock to rebuild towards the *B40%* target. For further details on the harvest control rule employed for Alaska sablefish, please refer to Appendix B. Finally, we used expert judgment (e.g., considering process research, fishery dynamics, and biological dynamics) to evaluate model performance and to determine the relative plausibility of model results. Although expert judgment may be subjective in nature, it is commonly used to evaluate stock assessments (Carvalho *et al.*, 2021). Nonetheless, we attempt to provide transparent and sensible rationale when using expert judgment to describe relative model performance.

Results

Development of Fishery-Dependent Standardized Indices

Comparisons of the gear-aggregated index (combined hook-and-line and pot-gear) and gear-disaggregated (fleet-specific) biomass indices demonstrated that the year trend derived from the gear-aggregated index was most similar to that of the hook-and-line index. The gear-aggregated index shows a small increase in the year 2020, whereas the hook-and-line index stabilizes (Fig. 3). Year trends from the pot index demonstrated large increases occurring in 2015, coinciding with periods of large recruitment events, which are often first observed in the Bering Sea and Aleutian Islands region, where the pot fishery operated prior to the regulatory change in 2017. Overall, year trends developed from the standardized indices do not seem implausible given *a priori* knowledge of biological processes for sablefish.

Estimation of Selectivity

Estimated logistic selectivity for females and males across models for the hook-and-line fleet (2016-2021 time-block) indicated that model *Pot-Logistic* was most similar to model *Combined-Logistic* with respect to the slope of the ascending limb and the initial age at maximum selection (Fig. 4). Similarities in the estimated hook-and-line selectivity between *Pot-Logistic* and *Combined-Logistic* are likely a consequence of model *Pot-Logistic* sharing the shape parameter by sexes, and between the hook-and-line and pot fleet. In contrast, hook-and-line selectivity for *Pot-Gamma* differed moderately relative to model *Combined-Logistic.* Specifically, younger fish appeared to be less vulnerable to fishing across sexes for model *Pot-Gamma*, and these differences were more pronounced for males (Fig. 4). Unsurprisingly, pot-specific selectivity for *Pot-Logistic* took on similar forms to selectivity estimates from the hook-and-line fleet, likely due to the sharing of the shape parameter by sex, which constrained the ascending limb of the logistic curve. When selectivity for the pot fleet was assumed to be dome-shaped following a gamma function (*Pot-Gamma*), the age at maximum selection was similar for females and males, occurring at ages five and six respectively. Additionally, the initial age at maximum selection describing pot selectivity between models *Pot-Gamma* and *Pot-Logistic* corresponded closely with each other across both sexes (Fig. 4). However, estimated pot-specific selectivity for *Pot-Gamma* across sexes indicated extreme and possibly unrealistic dome-shaped selectivity, where older age classes were less vulnerable to removals, and the rate at which selectivity at age declined was much faster for females.

Model Performance

All model variants presented in Table 1 had invertible Hessian matrices with maximum gradient components that were < 0.001, suggesting that these models achieved convergence. Although, we detected several highly correlated parameter pairs (> 0.95), many of these correlated parameter pairs were also present in model *Combined-Logistic* and largely consisted of fishing mortality and recruitment deviations. Notable highly correlated parameters were those associated with logistic selectivity (age at 50% selection and shape parameters) for males in the Alaska NMFS Longline Survey, which was only present in model *Pot-Gamma*. Retrospective analysis for SSB and fully-selected fishing mortality rates did not appear to suggest substantial retrospective inconsistencies (Fig. A1 and Fig. A2) (i.e., within cutoff values as defined by Hurtado-Ferro *et al*., 2015) for any of the models explored. Additionally, likelihood profiles for longline survey catchability did not exhibit abnormal likelihood surfaces (i.e., not trapped in local minima) and all data sources were generally in agreement across model variants (Fig. A3). Similarly, likelihood profiles for mean recruitment were generally in agreement across *Combined-Logistic*, *Pot-Logistic*, and *Pot-Gamma* models where the recruitment penalty (panel labelled as "Other" in Fig. A4) was the most influential. However, the likelihood response surface of mean recruitment for model *Pot-Gamma* was fairly uneven, which could be indicative of high parameter correlations (e.g., survey selectivity), and a poorly parametrized model.

Evaluation of Model Fits

Model fits to the gear-aggregated standardized biomass index for *Combined-Logistic* were acceptable and were fairly similar relative to models that incorporated a standardized hook-and-line index (gear-disaggregated models; Fig. 3). However, fits to the pot biomass index were mediocre for *Pot-Logistic* and *Pot-Gamma* models, with *Pot-Gamma* exhibiting slightly improved fits to the index (Fig. 3). Nonetheless, these mediocre fits to biomass indices are likely a result of the lower data weights assigned to the fishery-dependent index, compared to the fishery-independent survey abundance indices to which the model was fit.

Patterns in residuals for hook-and-line composition data were similar across model variants when compared to fits to the fixed-gear fleet for model *Combined-Logistic* and generally demonstrated satisfactory model fits (Fig. A5 and Fig. A6). Satisfactory model fits to the hook-and-line composition data suggest that logistic selectivity is a valid assumption for representing the disaggregated hook-and-line fishery fleet. In addition, the magnitude of absolute residuals across models were also similar. Model fits to pot composition data exhibited stronger systematic residual patterns for model *Pot-Logistic* relative to model *Pot_Gamma* (Fig. A5 and Fig. A6). In particular, runs of positive residuals were detected for ages 2-7 (i.e., smaller fish), which were accompanied by slight runs of negative residuals for older (i.e., larger) fish (Fig. 5). The presence of systematic patterns in residuals were generally less severe for model *Pot-Gamma* when compared to those from model *Pot-Logistic* (Fig. 5). Furthermore, mean absolute residuals were generally slightly larger (i.e., worse fit on average) across both age and length-composition data for the pot fishery for model *Pot-Logistic* relative to residuals from model *Pot-Gamma* (Fig. A5 and Fig. A6).

Estimation of Key Parameters and Management Quantities

Trends in SSB estimates were similar across all models, although estimated trends diverged during the start of the time-series likely due to a lack of informative data during that time-period (Fig. 6). Terminal year SSB estimates differed slightly across models, where *Combined-Logistic* and *Pot-Gamma* estimated the largest (106.39) and smallest (99.63) SSB values, respectively (Fig. 7). Similarly, estimates of *B40%* reference points were also slightly different across all models (Fig. 6 and Fig. 7). Despite these differences, the ratio of terminal SSB and the *B40%* reference point were almost identical (range: 0.87-0.90) across model variants, such that the estimated stock status across models were fairly similar (Fig. 7; upper right panel). Projections of SSB into the year 2036 also exhibited similar trajectories across all models, although we note differences in the scale of these estimates; the scale to which SSB increased was the largest for model *Combined-Logistic* and lowest for model *Pot-Gamma* (Fig. 6). In addition, projected declines following the peak SSB were less pronounced for model *Pot-Gamma* (Fig. 6), presumably due to the minimal selection of older ages as assumed by dome-shaped selectivity. Similar to the concordant nature of SSB estimates across models, estimates of predicted recruitment from 2016 to 2021 also exhibited comparable trends (Fig. 7).

Estimates of both fully-selected (sum of fleet-specific fishing mortality rates) and fleet-specific fishing mortality rates also generally followed consistent patterns across all model variants (Fig.7 and Fig. A7) but with differences in scale. Specifically, the scale of the fishing mortality rates for the pot fleet (also reflected in fully-selected fishing mortality rates) were much higher for model *Pot-Gamma* (Fig. A7), which is necessary in the presence of dome-shaped selectivity to adequately fit to catch observations. Acceptable Biological Catch (ABC) estimates were fairly different across all models. In particular, model *Pot-Gamma* estimated ABC values that were demonstrably higher compared to models *Combined-Logistic* and *Pot-Logistic* (Fig. 7). Considering that recruitment estimates were consistent across model variants, the higher ABC estimates resulting from models assuming dome-shaped selectivity (i.e. *Pot-Gamma*) is likely due to the lower modelled vulnerability of older age-classes to the pot fishery. Given that older, mature fish become essentially invulnerable to harvest once they survive the pot fishery process between ages 5 to 15 (i.e., given that hook-and-line harvest rates are comparatively lower; Figure 4), the model assumes a spawning refuge that enables higher removals.

Discussion

As management systems continue to confront the dynamic nature of fisheries, it becomes imperative for stock assessment models to adapt accordingly. Our results demonstrate that disaggregating the fixed-gear fleet structure appeared to have minimal impacts on estimates of biomass levels in the case of Alaska sablefish. Given similarities in estimates of biomass levels between multi-fleet and fleet-aggregated models, we believe that disaggregating fleet structure can serve as a useful basis for validating single-fleet models and can provide valuable insight into fleet-specific dynamics. However, our results illustrate that assuming dome-shaped selectivity may lead to overly optimistic harvest recommendations (Cadrin *et al.*, 2016; Northeast Fisheries Science Center (NEFSC), 2019), especially when informed by a limited time-series of age-or length-composition data as was the case for pot fleet in this context. In the following sections, we highlight the importance of considering *a priori* knowledge of fishery and biological dynamics and provide practical guidance for fisheries and associated assessment models experiencing changes in gear usage.

<u>Implications of Disaggregated Fleet Structure</u>

Given the complexity of stock assessment models, which can estimate hundreds of parameters, model parsimony is often an important consideration when selecting among models (Walters and Martell, 2002; Cotter *et al.*, 2004). In comparison to model *Combined-Logistic*, model variants that assumed a disaggregated fishery fleet structure (Table 1) were more complex given the need to estimate new parameters (up to 30 additional parameters) for fleet-specific fishing mortality rates and fleet-and sex-specific selectivity processes. The increased complexity across model variants did not result in substantially improved model performance and provided similar estimates (with exception of reference points for *Pot-Gamma*) relative to the fixed-gear fleet structure as assumed by model *Combined-Logistic,* suggesting that the added complexity may not be necessary, especially given a limited time series available for the pot fleet. However, the process of disaggregating fleet structure can better represent the reality as observed and understood by harvesters and provides additional insight into fleet-specific fishery dynamics. Similar to Nielsen *et al.* (2021), findings from our study also suggest that similarities between fleet-disaggregated models and single-fleet models can be used as a tool to further validate model results, diagnose potential conflicts within a single fleet model, and improve confidence in the stock assessment process.

The Alaska sablefish case study indicated that when extremely rapid changes in fleet composition occur, the most parsimonious approach may be to assume a single fleet for the fixed-gear fleet, while allowing for a change in the selectivity pattern using a time-block, rather than disaggregating the fixed-gear fleet (e.g., *Pot-Logistic* and *Pot-Gamma*) and adding complexity. Compared to previous iterations (2020) of the operational sablefish assessment (without time-block selectivity), the incorporation of time-blocked selectivity demonstrated improved model fits to compositional data for the fixed-gear fleet and improved retrospective patterns (Goethel *et al.*, 2020, 2021). Thus, assuming a single fleet will likely be sensible under rapid shifts in fleet composition, especially if contact selectivity and availability processes do not appear to be drastically different between the existing and emerging fleets. This is likely the case for hook-and-line and pot-gears for Alaska sablefish, where the contact selectivity process of the two gears have been demonstrated to be comparable (Sullivan *et al.*, 2022), although differences in the availability selection process of the two gears remains unclear.

Selectivity and Model Fits to Composition Data

Upon the disaggregation of the fixed-gear fleet structure, fits to the hook-and-line composition data were not substantially degraded relative to the status-quo model. Furthermore, estimates of selectivity for the hook-and-line fleet were most similar between model *Combined-Logistic* and *Pot-Logistic*, likely due to model *Pot-Logistic* sharing shape parameters by sexes, between the hook-and-line and pot fleet. With respect to fits to the pot composition data, model performance varied depending on the assumed selectivity function for the pot fishery. In general, model variants assuming dome-shaped selectivity for the pot fishery resulted in better agreement between predicted and observed composition data compared to logistic selectivity (Fig. 5). Despite improved statistical fit, extreme dome-shaped selectivity as estimated by a re-parametrized gamma function may not be representative of removal processes from the pot fishery in the present study. From our experience, dome-shaped selectivity represented by the gamma function (Punt *et al.*, 1996) is inflexible, relative to other dome-shaped selectivity forms and can result in unrealistically extreme declines in selectivity for older ages, especially with limited data available to inform the descending limb of the function.

Consideration of the information provided by compositional data for informing selectivity is critical in the context of this Alaska sablefish case study, wherein the timeframe for the rapid emergence of the pot fleet in the Gulf of Alaska directly overlaps the observation of several anomalously large recruitment events (2014, 2016, 2017, 2019), resulting in a high abundance of younger individuals within the population. However, due to the limited time-series of composition data available for the pot fishery (Fig. A8 and Fig. A9), other flexible dome-shaped selectivity functions (double-normal, double-logistic, exponential-logistic) to represent the pot fishery were unable to achieve adequate model performance (i.e., non-invertible Hessian). Given the sex-structured nature of the assessment model, and the limited time-series for the pot fishery, additional partitions with respect to gear-types are likely not practical under the current data scenario for Alaska sablefish. In the case of Alaska sablefish where sex-specific dynamics are a key driver of population dynamics, incorporating sexually dimorphic growth is likely more important than accounting for gear-specific differences. However, for fisheries where sexually dimorphic growth is negligible, accounting for an additional gear dimension may prove to be a potentially crucial and estimable partition.

Considering harvester targeting practices and market demands (Goethel *et al.*, 2021), selectivity estimates based on previous tagging studies of sablefish (Maloney and Sigler, 2008; Jones and Cox, 2018), the highly migratory nature of sablefish (Hanselman *et al.*, 2015; O'Boyle *et al.*, 2016), and comparable length-compositions observed between the two gears during gear comparison studies (Sullivan *et al.*, 2022), it is unlikely that the rate of selection for older individuals declines as rapidly as estimated for model *Pot-Gamma* (Fig. 4). Improved model fits as a result of assuming dome-shaped selectivity could potentially be attributed to high recruitment events during 2014, 2016, 2017, and 2019. These high recruitment events coincide with the regulatory shift in pot-gear in 2017, such that the pot composition data reflect a dominance of younger fish, potentially obscuring the signal of older individuals being removed from the population (Goethel *et al*., 2021). Furthermore, high recruitment events tend to first be observed in the Bering Sea and Aleutian Islands, where the pot fishery primarily operated prior to 2017. Thus, it is plausible that pot-gear selects for younger individuals through availability selection (Sampson, 2014), resulting in dome-shaped selectihvity (Sampson and Scott, 2012). However, the steep descending limb as estimated in *Pot-Gamma* is unlikely as discussed above. In particular, individuals move from nearshore to offshore regions (depths > 200m) as they mature, and the depth ranges (>400m) that the pot fishery primarily operates in suggests that selection of older individuals should be higher than is estimated by model *Pot-Gamma*. However, given the extreme demographic state of the population, the removal of these old individuals are likely inundated by the abundance of young individuals. These dynamics are likely further accentuated by hypothesized density dependent effects, where younger individuals have appeared to inhabit deeper depths following these recent high recruitment events (Goethel *et al.*, 2021). Although model *Pot-Logistic* appeared slightly mis-specified when fit to the composition data for pot-gear (Fig. 4), other model diagnostics (i.e., likelihood profiles, parameter correlation) did not suggest a major cause for concern. Thus, given the biological and fishery dynamics associated with Alaska sablefish, model variants assuming logistic selectivity might be more appropriate for the purpose of representing removals from the Alaska sablefish pot fishery, especially with the limited time series of data for the emerging pot fleet currently available. More complex selectivity parameterizations (e.g., double-normal, double-logistic) could potentially reconcile conflicts between model fits and *a priori* knowledge, but often failed to achieve convergence as previously noted. Incorporation of priors to investigate the degree of doming may also reconcile such conflicts, but were not explored as they were beyond the scope of the current study. Consequently, our results suggest that the optimal selectivity form to represent a new emerging fishery should likely depend on *a priori* knowledge of data quality and representativeness of the functional form (Privitera-Johnson *et al.*, 2022; Punt, 2023).

Treatment of Biomass Indices

Overall, the use of aggregated and disaggregated biomass indices did not demonstrate apparent differences in model performance and key model results were also similar (with exception of reference points). Model fits for both the gear-aggregated and standardized hook-and-line index were generally similar and appropriate (Fig. 3), but were mediocre for the standardized pot index (*Pot-Logistic* and *Pot-Gamma*). These lack of differences in model performance and results are likely attributed to the lower relative weights applied to the fishery-dependent indices. Nevertheless, using gear-aggregated standardized biomass indices can leverage additional spatiotemporal information available from different gears, which can potentially provide more informative and robust trends in stock status (Cheng *et al.*, 2023a). While methods incorporating spatiotemporal information other than tensor product smooths are available (e.g., Gaussian Markov Random Fields; Rue and Tjelmeland, 2002; Thorson and Barnett, 2017; Thorson, 2019), they were not further explored, given that it was beyond the scope of the study. Furthermore, some studies have found that different spatiotemporal interpolation methods (i.e., tensor products compared to Gaussian Markov Random Fields) can demonstrate similar model performance (Brodie *et al.*, 2020; Stock *et al.*, 2020). Thus, alternative methods for accounting for spatiotemporal correlations in the index standardization process are unlikely to have greatly impacted the interpretation of results in this study. For assessments assuming a disaggregated fleet structure, the use of fleet-specific indhhhhices can improve transparency in the assessment process and better reflects empirical observations from harvesters, which can help facilitate agreeable management outcomes when changes are necessitated (Goethel *et al.*, 2019; Barbeaux *et al.*, 2020).

Estimation of Key Management Quantities and Population Status

Trends in SSB and the ratio of terminal year SSB and *B40%* were fairly similar across all models explored in this study, irrespective of the treatment of fleet structure. However, differences in selectivity assumptions for models represented with a disaggregated fixed-gear fleet resulted in substantially different recommended harvest levels. In particular, model *Pot-Gamma* estimated ABC values that were much higher, despite similar estimates of population status across models. Such differences are likely ascribed to the reduced vulnerability of older mature age classes to the pot fishery given the strong dome-shape estimated for selectivity. SSB projections into the year 2036 exhibited less pronounced declines for model *Pot-Gamma* (Fig. 6), which are also presumably attributed to the older cohorts recruiting to ages unavailable to the pot fleet, resulting in higher levels of SSB maintained in the long-term. Despite improved statistical fit to the pot composition data when assuming dome-shaped selectivity, harvest levels were sensitive to the assumed choice of selectivity forms and may suggest the need to rely on the knowledge of biological and fishery processes, especially during these initial periods of change in fleet structure. Similar to findings from Bohaboy *et al*. (2022), the implementation of dome-shaped selectivity when multiple fisheries exists can result in obscure interactions between selectivity and harvest recommendations. Findings from our study further underscore the sensitivity of management references points to selectivity assumptions (Scott and Sampson, 2011; Butterworth *et al.*, 2014), and the value of subject matter expertise in stock assessment (Rosenberg and Restrepo, 1994). Furthermore, we recommend that fleet structure and selectivity are carefully explored in tandem, especially when there are rapid shifts in fleet structure.

Caveats and Future Work

The need to directly account for multi-dimensional processes (e.g., gear, space, time, sex) within stock assessments is well recognized (Wang *et al*., 2005; Goethel *et al*., 2011). Given that sexually dimorphic growth is a key driver in sablefish population dynamics (Goethel *et al.*, 2021), incorporating both sex and gear partitions limited the estimation of sex- and fleet-specific selectivity parameters for model *Pot-Logistic*, and sharing of sex-specific selectivity parameters for the pot fleet was necessary to achieve adequate model performance. Although such parameterizations are imperfect, we believe that parameter sharing with the hook-and-line 2016-2021 time-block is reasonable considering that a majority of the removals from the pot fishery began in 2017. Sharing of parameters is not uncommon, and is similar to the "Robin Hood" approach described by Punt *et al*. (2011), but parameter values in the current study are assumed to be the same among fleets instead of estimated with penalties or priors. Furthermore, the re-parametrized gamma function used in model *Pot-Gamma* can often be inflexible (restricted to 2 parameters) when compared to other domed-shaped selectivity forms. The limited time-series of compositional data available for the pot gear further impeded the ability to estimate more flexible domed-shape parameterizations due to the increased number of parameters to be estimated from extremely limited data sample sizes. Moreover, the limited compositional data combined with the rapidly changing population demographics (i.e., an extremely small and young population in recent years) resulted in unrealistically extreme doming of the selectivity when using the gamma function, as discussed above.

In addition to the limited time-series of available compositional data, other components incorporated within the assessment model could have impacted the estimation of selectivity. For instance, age-composition data were input as sex-aggregated, while length-compositions are input as sex-specific, but it remains unclear how the treatment of compositional data might adversely impact the estimation of selectivity processes. Ageing error and selectivity are also known to interact with each other, which can impact estimates of cohort size (both under and overestimation; Bradford, 1991; Punt *et al.*, 2008), inaccurate estimates of population status, and biases in management reference points (Henríquez *et al.*, 2016). However, given that an ageing-error matrix is directly incorporated in the assessment model to account for uncertainty in the ageing process, ageing error is unlikely to have substantially impacted the estimation of selectivity in the context of this study.

Model configurations in the present study represent the pot fishery as a single fleet despite the use of multiple pot types (rigid pots and "slinky pots"). However, considering the recent introduction of "slinky pots" in 2019, there are likely insufficient data available to further partition out an additional gear dimension. Furthermore, historical fishing effort from the pot fishery was primarily concentrated in the Bering Sea and Aleutian Islands , which later expanded into the Gulf of Alaska in 2017, facilitated by the pot regulatory change. Such changes in the spatial distribution of fishing effort can potentially impact spatial harvest patterns and availability of cohorts, which can be further accentuated by the ontogenetic movement patterns Alaska sablefish exhibit. For instance, O'Boyle *et al.* (2016) showed that age-specific movements, along with spatially heterogenous fishing mortality rates can result in dome-shaped selectivity, despite contact selection following patterns of asymptotic selectivity. Similarly, Sampson and Scott (2011, 2012) demonstrated that when stocks are not well-mixed and experience spatially uneven fishing mortality patterns, dome-shaped selectivity can also manifest. Thus, the aforementioned factors further complicates the estimation of fishery selectivity processes when assuming a single area assessment model, as is the case in the current study.

Allowing for additional flexibility in fishery selectivity processes (i.e., continuous time-variation rather than discrete changes) and the use of spatial stock assessment models (spatially-explicit or -implicit) (Cope and Punt, 2011; Stewart and Martell, 2014; Waterhouse *et al.*, 2014; O'Boyle *et al.*, 2016; Lee *et al.*, 2017) may help better characterize these removal processes.

However, continuous time-varying selectivity approaches were not further explored given difficulties in achieving model convergence. Lastly, a moderate proportion of individuals in the plus-group were detected in pot age-composition data relative to younger age-bins (Fig. 5), which may suggest the need to expand the number of individuals modelled within the assessment model, but were not explored in this study.

As data from the pot-gear fishery increases over time, future work should explore alternative models that allow for more flexible selectivity functional forms and/or accounts for time-varying selectivity processes in the Alaska sablefish stock assessment. In particular, multi-dimensional autoregressive processes in selectivity (i.e., age, year, or cohort effects) could be fruitful to explore (Cheng *et al*., 2023b; Xu *et al*., 2020, 2019). Future studies could also conduct simulation analyses to evaluate the implications of ignoring fleet structure, assuming a single fleet with continuous or time-blocked time-varying selectivity, or disaggregating fleet structure when a new fleet emerges.

General Recommendations on Fleet Disaggregation

Data availability are a key determinant in constraining the dimensions that an assessment model can represent (Chen *et al.*, 2003; Hodgdon *et al.*, 2022). Although modelling selectivity as a time-varying process has been identified as best practice (Martell and Stewart, 2014), the dimensions represented within an assessment model should also be based upon considerations regarding data quantity and quality (Privitera-Johnson *et al*., 2022; Punt 2023), model parsimony, and *a priori* understanding of fishery and stock dynamics (Rosenberg and Restrepo, 1994; Francis, 2011; Hulson and Hanselman, 2014; Carvalho *et al.*, 2021). Thus, decisions with respect to model structure and assumptions should not be based purely on statistical fit. The involvement of both stakeholders and harvesters can also be fruitful in the assessment process, which can help fill in knowledge gaps through the inclusion of local knowledge, facilitate information sharing and provide insight for identifying pragmatic stock assessment parameterizations (Duplisea, 2018; Goethel *et al*., 2022; Johannes *et al*., 2008; Neis *et al*., 1999; Peterson *et al*., 2014). In addition, alternative sensible parameterizations of selectivity through parameter sharing, penalties, or aggregating selectivities among modelled partitions (e.g., sex-invariant selectivity) to achieve adequate model performance would be fruitful to explore in scenarios where limited time-series exist (Punt *et al.*, 2011). When multiple fishery fleets are present, we recommend disaggregating fleet structure to compare against single fleet parameterizations if these model structures are supported by the data available. Doing so facilitates comparisons between single- and multi-fleet assessment models, enables analysts to better understand model behavior, aids in model validation, and improves tactical and strategic decision-making. Furthermore, analyzing fleet structure can enable improved fishery monitoring procedures, understanding of spatial and fleet-specific harvest patterns (Eigaard *et al.*, 2011), and the development of fleet-based catch, effort, and discard management procedures (Ulrich *et al.*, 2002; Bastardie *et al.*, 2010b, 2010a; Holmes *et al.*, 2011; Nielsen *et al.*, 2021). Finally, we recommend using simulation analyses and management strategy evaluations to identify pragmatic model parameterizations that are paired with management procedures robust to differential fishery process and dynamic changes to fleet structures. Although the incorporation of an additional gear dimension does not appear to be an immediate concern for Alaska sablefish, adequately emulating fleet-specific dynamics might be more impactful for assessment models with fewer modelled dimensions (i.e., negligible sex-specific dynamics), and will likely be of more merit in cases where fleet structure changes slowly.

Acknowledgments

We thank the Alaska sablefish industry for their efforts in providing vessel logbooks, the International Pacific Halibut Commission for collecting logbooks dockside, and the NOAA North Pacific Observer Program for their long-term data collection on fishery-dependent data. M.LH.C was supported by the National Science Foundation Graduate Research Fellowship Program. Megan Feddern, Genoa Sullaway, and Lillian Hart provided helpful reviews of initial drafts of the manuscript. We thank Pete Hulson, Cara Rodgveller, Chris Lunsford, and Franz Mueter for providing critical feedback and discussions on findings from the study. We also acknowledge Dana Hanselman for originally developing the ADMB sablefish assessment model. We thank the two anonymous reviewers for their helpful comments. Findings and conclusions of this study are those of the authors, and do not represent the views of the National Marine Fisheries Service, NOAA.

References

Andersen, B. S., Ulrich, C., Eigaard, O. R., and Christensen, A.-S. 2012. Short-term choice behaviour in a mixed fishery: investigating métier selection in the Danish gillnet fishery. ICES Journal of Marine Science, 69: 131–143.

Barbeaux, S. J., Holsman, K., and Zador, S. 2020. Marine Heatwave Stress Test of Ecosystem-Based Fisheries Management in the Gulf of Alaska Pacific Cod Fishery. Frontiers in Marine Science, 7: 703.

Bastardie, F., Vinther, M., Nielsen, J. R., Ulrich, C., and Paulsen, M. S. 2010a. Stock-based vs. fleet-based evaluation of the multi-annual management plan for the cod stocks in the Baltic Sea. Fisheries Research, 101: 188–202.

Bastardie, F., Nielsen, J. R., and Kraus, G. 2010b. The eastern Baltic cod fishery: a fleet-based management strategy evaluation framework to assess the cod recovery plan of 2008. ICES Journal of Marine Science, 67: 71–86.

Baudron, A., Ulrich, C., Nielsen, J. R., and Boje, J. 2010. Comparative evaluation of a mixed-fisheries effort-management system based on the Faroe Islands example. ICES Journal of Marine Science, 67: 1036–1050.

Berger, A. M., Jones, M. L., Zhao, Y., and 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.

Beverton, R., and Holt, S. 1957. On the Dynamics of Exploited Fish Populations. Min. Agric. Fish and Food U.K. Fish Invest. Ser., London. 533 pp.

Bohaboy, E. C., Goethel, D. R., Cass-Calay, S. L., and Patterson, W. F. 2022. A simulation framework to assess management trade-offs associated with recreational harvest slots, discard mortality reduction, and bycatch accountability in a multi-sector fishery. Fisheries Research, 250: 106268.

Bradford, M. J. 1991. Effects of Ageing Errors on Recruitment Time Series Estimated from Sequential Population Analysis. Canadian Journal of Fisheries and Aquatic Sciences, 48: 555–558.

Brodie, S. J., Thorson, J. T., Carroll, G., Hazen, E. L., Bograd, S., Haltuch, M. A., Holsman, K. K., *et al.* 2020. Trade-offs in covariate selection for species distribution models: a methodological comparison. Ecography, 43: 11–24.

Butterworth, D. S., Rademeyer, R. A., Brandão, A., Geromont, H. F., and Johnston, S. J. 2014. Does selectivity matter? A fisheries management perspective. Fisheries Research, 158: 194–204.

Cadrin, S. X., DeCelles, G. R., and Reid, D. 2016. Informing fishery assessment and management with field observations of selectivity and efficiency. Fisheries Research, 184: 9–17.

Carvalho, F., Winker, H., Courtney, D., Kapur, M., Kell, L., Cardinale, M., Schirripa, M., *et al.* 2021. A cookbook for using model diagnostics in integrated stock assessments. Fisheries Research, 240: 105959.

Chen, Y., Chen, L., and Stergiou, K. I. 2003. Impacts of data quantity on fisheries stock assessment. Aquatic Sciences - Research Across Boundaries, 65: 92–98.

Cheng, M. L. H., Rodgveller, C. J., Langan, J. A., and Cunningham, C. J. 2023a. Standardizing

fishery-dependent catch-rate information across gears and data collection programs for Alaska sablefish (*Anoplopoma fimbria*). ICES Journal of Marine Science: fsad037.
Cheng, M. L. H., Thorson, J. T., Ianelli, J. N., and Cunningham, C. J. 2023b. Unlocking the triad of age, year, and cohort effects for stock assessment: Demonstration of a computationally efficient and reproducible framework using weight-at-age. Fisheries Research, 266.
Cope, J. M., and Punt, A. E. 2011. Reconciling stock assessment and management scales under conditions of spatially varying catch histories. Fisheries Research, 107: 22–38.
Cotter, A. J. R., Burt, L., Paxton, C. G. M., Fernandez, C., Buckland, S. T., and Pan, J.-X. 2004. Are stock assessment methods too complicated? Fish and Fisheries, 5: 235–254.
Deroba, J. J., and Bence, J. R. 2008. A review of harvest policies: Understanding relative performance of control rules. Fisheries Research, 94: 210–223.
Duplisea, D. E. 2018. Eliminating implausible fisheries assessment models using fishers' knowledge. Canadian Journal of Fisheries and Aquatic Sciences, 75: 1280–1290.
Eigaard, O. R., Thomsen, B., Hovgaard, H., Nielsen, A., and Rijnsdorp, A. D. 2011. Fishing power increases from technological development in the Faroe Islands longline fishery. Canadian Journal of Fisheries and Aquatic Sciences, 68: 1970–1982.
Eigaard, O. R., Marchal, P., Gislason, H., and Rijnsdorp, A. D. 2014. Technological Development and Fisheries Management. Reviews in Fisheries Science & Aquaculture, 22: 156–174.
Fernández, C., Cerviño, S., Pérez, N., and Jardim, E. 2010. Stock assessment and projections incorporating discard estimates in some years: an application to the hake stock in ICES Divisions VIIIc and IXa. ICES Journal of Marine Science, 67: 1185–1197.
Fissel, B., Dalton, M., Felthoven, R., Garber-Yonts, B., Haynie, A., and Kasperski, S. 2016. Economic Status of the Groundfish Fisheries Off Alaska 2015: 488.
Fournier, D. A., Skaug, H. J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M. N., Nielsen, A., *et al.* 2012. AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. Optimization Methods and Software, 27: 233–249.
Francis, R. I. C. C. 2011. Data weighting in statistical fisheries stock assessment models. Canadian Journal of Fisheries and Aquatic Sciences, 68: 1124–1138.
Francis, R. I. C. C. 2017. Revisiting data weighting in fisheries stock assessment models. Fisheries Research, 192: 5–15.
Gascuel, D., Merino, G., Döring, R., Druon, J. N., Goti, L., Guénette, S., Macher, C., *et al.* 2012. Towards the implementation of an integrated ecosystem fleet-based management of European fisheries. Marine Policy, 36: 1022–1032.
Goethel, D., Hanselman, D. H., Rodgveller, C. J., Fenske, K. H., Shotwell, S. K., Echave, K. B., Malecha, P. W., *et al.* 2020. 3. Assessment of the Sablefish Stock in Alaska: 257.
Goethel, D., Hanselman, D., Rodgveller, C., Echave, K. B., Williams, B., Shotwell, S. K., Sullivan, J., *et al.* 2021. 3. Assessment of the Sablefish Stock in Alaska: 347.
Goethel, D. R., Quinn, T. J., and Cadrin, S. X. 2011. Incorporating Spatial Structure in Stock Assessment: Movement Modeling in Marine Fish Population Dynamics. Reviews in Fisheries Science, 19: 119–136.
Goethel, D. R., Lucey, S. M., Berger, A. M., Gaichas, S. K., Karp, M. A., Lynch, P. D., Walter, J. F., *et al.* 2019. Closing the feedback loop: on stakeholder participation in management strategy evaluation. Canadian Journal of Fisheries and Aquatic Sciences, 76: 1895–1913.

Goethel, D. R., Rodgveller, C. J., Echave, K. B., Shotwell, S. K., Siwicke, K. A., Malecha, P. W., Cheng, M., *et al.* 2022a. Assessment of the Sablefish Stock in Alaska: 182.

Goethel, D. R., Omori, K. L., Punt, A. E., Lynch, P. D., Berger, A. M., de Moor, C. L., Plagányi, É. E., *et al.* 2022b. Oceans of plenty? Challenges, advancements, and future directions for the provision of evidence-based fisheries management advice. Reviews in Fish Biology and Fisheries. https://link.springer.com/10.1007/s11160-022-09726-7 (Accessed 18 November 2022).

Goethel, D. R., Cheng, M. L. H., Echave, K. B., Marsh, C., Rodgveller, C. J., and Siwicke, K. 2023. 3. Assessment of the Sablefish Stock in Alaska. North Pacific Fishery Management Council, Anchorage, AK.

Goodyear, C. P. 1996. Variability of Fishing Mortality by Age: Consequences for Maximum Sustainable Yield. North American Journal of Fisheries Management, 16: 6.

Hanselman, D., Rodgveller, C., Fenske, K., Shotwell, S. K., Echave, K., Malecha, P., and Lunsford, C. 2018. Assessment of the sablefish stock in Alaska: 216.

Hanselman, D., Rodgveller, C., Fenske, K., Shotwell, S. K., Echave, K., Malecha, P., and Lunsford, C. 2019. Assessment of the sablefish stock in Alaska: 263.

Hanselman, D. H., Heifetz, J., Echave, K. B., and Dressel, S. C. 2015. Move it or lose it: movement and mortality of sablefish tagged in Alaska. Canadian Journal of Fisheries and Aquatic Sciences, 72: 238–251.

Henríquez, V., Licandeo, R., Cubillos, L. A., and Cox, S. P. 2016. Interactions between ageing error and selectivity in statistical catch-at-age models: simulations and implications for assessment of the Chilean Patagonian toothfish fishery. ICES Journal of Marine Science, 73: 1074–1090.

Hilborn, R. 2003. The state of the art in stock assessment: where we are and where we are going. Scientia Marina, 67: 15–20.

Hodgdon, C. T., Khalsa, N. S., Li, Y., Sun, M., Boenish, R., and Chen, Y. 2022. Global crustacean stock assessment modelling: Reconciling available data and complexity. Fish and Fisheries, 23: 697–707.

Holmes, S. J., Bailey, N., Campbell, N., Catarino, R., Barratt, K., Gibb, A., and Fernandes, P. G. 2011. Using fishery-dependent data to inform the development and operation of a co-management initiative to reduce cod mortality and cut discards. ICES Journal of Marine Science, 68: 1679–1688.

Hoyle, S. D., Campbell, R. A., Ducharme-Barth, N. D., Grüss, A., Moore, B. R., Thorson, J. T., Tremblay-Boyer, L., *et al.* 2024. Catch per unit effort modelling for stock assessment: A summary of good practices. Fisheries Research, 269: 106860.

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

Hurtado-Ferro, F., Punt, A. E., and Hill, K. T. 2014. Use of multiple selectivity patterns as a proxy for spatial structure. Fisheries Research, 158: 102–115.

Hurtado-Ferro, F., Szuwalski, C. S., Valero, J. L., Anderson, S. C., Cunningham, C. J., Johnson, K. F., Licandeo, R., *et al.* 2015. Looking in the rear-view mirror: bias and retrospective patterns in integrated, age-structured stock assessment models. ICES Journal of Marine Science, 72: 99–110.

Johannes, R. E., Freeman, M. M. R., and Hamilton, R. J. 2008. Ignore fishers' knowledge and miss the boat. Fish and Fisheries, 1: 257–271.

Jones, M. K., and Cox, S. 2018. Size-selectivity for British Columbia Sablefish (Anoplopoma fimbria) estimated from a long-term tagging study. Fisheries Research, 199: 94–106.

Lee, H.-H., Piner, K. R., Methot, R. D., and Maunder, M. N. 2014. Use of likelihood profiling over a global scaling parameter to structure the population dynamics model: An example using blue marlin in the Pacific Ocean. Fisheries Research, 158: 138–146.

Lee, H.-H., Piner, K. R., Maunder, M. N., Taylor, I. G., and Methot, R. D. 2017. Evaluation of alternative modelling approaches to account for spatial effects due to age-based movement. Canadian Journal of Fisheries and Aquatic Sciences, 74: 1832–1844.

Lennert-Cody, C. E., Minami, M., Tomlinson, P. K., and Maunder, M. N. 2010. Exploratory analysis of spatial–temporal patterns in length–frequency data: An example of distributional regression trees. Fisheries Research, 102: 323–326.

Lennert-Cody, C. E., Maunder, M. N., Aires-da-Silva, A., and Minami, M. 2013. Defining population spatial units: Simultaneous analysis of frequency distributions and time series. Fisheries Research, 139: 85–92.

Linton, B. C., and Bence, J. R. 2011. Catch-at-age assessment in the face of time-varying selectivity. ICES Journal of Marine Science, 68: 611–625.

Maloney, N. E., and Sigler, M. F. 2008. Age-specific movement patterns of sablefish (Anoplopoma fimbria) in Alaska. Fishery Bulletin, 106(3): 305–316.

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

Maunder, M. N. 2002. The relationship between fishing methods, fisheries management and the estimation of maximum sustainable yield. Fish and Fisheries, 3: 251–260.

Maunder, M. N., and Punt, A. E. 2013. A review of integrated analysis in fisheries stock assessment. Fisheries Research, 142: 61–74.

Maunder, M. N., and Piner, K. R. 2017. Dealing with data conflicts in statistical inference of population assessment models that integrate information from multiple diverse data sets. Fisheries Research, 192: 16–27.

Neis, B., Schneider, D., Felt, L., Haedrich, R., Fischer, J., and Hutchings, J. 1999. Fisheries assessment: what can be learned from interviewing resource users? Canadian Journal of Fisheries and Aquatic Sciences, 56: 1949–1963.

Nielsen, A., and Berg, C. W. 2014. Estimation of time-varying selectivity in stock assessments using state-space models. Fisheries Research, 158: 96–101.

Nielsen, A., Hintzen, N. T., Mosegaard, H., Trijoulet, V., and Berg, C. W. 2021. Multi-fleet state-space assessment model strengthens confidence in single-fleet SAM and provides fleet-specific forecast options. ICES Journal of Marine Science, 78: 2043–2052.

Northeast Fisheries Science Center (NEFSC). 2019. 66th Northeast Regional Stock Assessment Workshop (66th SAW) Assessment Report. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 19-08; 1170 p. Available from: http://www.nefsc.noaa.gov/ publications/.

O'Boyle, R., Dean, M., and Legault, C. M. 2016. The influence of seasonal migrations on fishery selectivity. ICES Journal of Marine Science, 73: 1774–1787.

Pascoe, S., Punt, A. E., and Dichmont, C. M. 2010. Targeting ability and output controls in Australia's multi-species Northern Prawn Fishery. European Review of Agricultural Economics, 37: 313–334.

Peterson, M. J., Mueter, F., Criddle, K., and Haynie, A. C. 2014. Killer Whale Depredation and Associated Costs to Alaskan Sablefish, Pacific Halibut and Greenland Turbot Longliners. PLoS ONE, 9: e88906.

Pol, M., and Carr, H. A. 2000. OVERVIEW OF GEAR DEVELOPMENTS AND TRENDS IN THE NEW ENGLAND COMMERCIAL FISHING INDUSTRY. Northeastern Naturalist, 7.

Privitera-Johnson, K. M., Methot, R. D., and Punt, A. E. 2022. Towards best practice for specifying selectivity in age-structured integrated stock assessments. Fisheries Research, 249: 106247.

Punt, A. E., Pulfrich, A., Butterworth, D. S., and Penney, A. J. 1996. The effect of hook size on the size-specific selectivity of hottentot *Pachymetopon blochii* (Val.) and on yield per recruit. South African Journal of Marine Science, 17: 155–172.

Punt, A. E., Smith, D. C., KrusicGolub, K., and 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. Canadian Journal of Fisheries and Aquatic Sciences, 65: 1991–2005.

Punt, A. E., Smith, D. C., and Smith, A. D. M. 2011. Among-stock comparisons for improving stock assessments of data-poor stocks: the "Robin Hood" approach. ICES Journal of Marine Science, 68: 972–981.

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

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.

Rosenberg, A. A., and Restrepo, V. R. 1994. Uncertainty and Risk Evaluation in Stock Assessment Advice for U.S. Marine Fisheries. Canadian Journal of Fisheries and Aquatic Sciences, 51: 2715–2720.

Rue, H., and Tjelmeland, H. 2002. Fitting Gaussian Markov Random Fields to Gaussian Fields. Scandinavian Journal of Statistics, 29: 31–49.

Sainsbury, K. J. 1984. Optimal mesh size for tropical multispecies trawl fisheries. ICES Journal of Marine Science, 41: 129–139.

Sampson, D. B., and Scott, R. D. 2012. An exploration of the shapes and stability of population-selection curves: Shapes and stability of population-selection curves. Fish and Fisheries, 13: 89–104.

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

Scott, R. D., and Sampson, D. B. 2011. The sensitivity of long-term yield targets to changes in fishery age-selectivity. Marine Policy, 35: 79–84.

Sigler, M. F., and Lunsford, C. R. 2001. Effects of individual quotas on catching efficiency and spawning potential in the Alaska sablefish fishery. Canadian Journal of Fisheries and Aquatic Sciences, 58: 1300–1312.

Sinclair, A. F. 1993. Partial Recruitment Considerations in Setting Catch Quotas. Canadian Journal of Fisheries and Aquatic Sciences, 50: 734–742.

Stewart, I. J., and Martell, S. J. D. 2014. A historical review of selectivity approaches and retrospective patterns in the Pacific halibut stock assessment. Fisheries Research, 158: 40–49.

Stock, B. C., Ward, E. J., Eguchi, T., Jannot, J. E., Thorson, J. T., Feist, B. E., and Semmens, B. X. 2020. Comparing predictions of fisheries bycatch using multiple spatiotemporal species distribution model frameworks. Canadian Journal of Fisheries and Aquatic Sciences, 77: 146–163.

Sullivan, J., Dimond, J. A., and Malecha, P. 2022. Slinky pot and hook-and-line comparison project during the experimental leg of the 2021 AFSC sablefish longline survey. AFSC Processed Rep. 2022-02, 18 p. Alaska Fish. Sci. Cent., NOAA, Natl. Mar. Fish. Serv., Auke Bay Laboratories,17109 Pt. Lena Loop Road, Juneau, AK, 99801: 27.

Szuwalski, C. S., Ianelli, J. N., and Punt, A. E. 2018. Reducing retrospective patterns in stock assessment and impacts on management performance. ICES Journal of Marine Science, 75: 596–609.

Thompson, G. G. 1994. Confounding of Gear Selectivity and the Natural Mortality Rate in Cases where the Former is a Nonmonotone Function of Age. Canadian Journal of Fisheries and Aquatic Sciences, 51: 2654–2664.

Thorson, J. T., and Barnett, L. A. K. 2017. Comparing estimates of abundance trends and distribution shifts using single- and multispecies models of fishes and biogenic habitat. ICES Journal of Marine Science, 74: 1311–1321.

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.

Thygesen, U. H., Albertsen, C. M., Berg, C. W., Kristensen, K., and Nielsen, A. 2017. Validation of ecological state space models using the Laplace approximation. Environmental and Ecological Statistics, 24: 317–339.

Tixier, P., Lea, M., Hindell, M. A., Welsford, D., Mazé, C., Gourguet, S., and Arnould, J. P. Y. 2021. When large marine predators feed on fisheries catches: Global patterns of the depredation conflict and directions for coexistence. Fish and Fisheries, 22: 31–53.

Trijoulet, V., Albertsen, C. M., Kristensen, K., Legault, C. M., Miller, T. J., and Nielsen, A. 2023. Model validation for compositional data in stock assessment models: Calculating residuals with correct properties. Fisheries Research, 257: 106487.

Ulrich, C., Pascoe, S., Sparre, P. J., Wilde, J.-W. D., and Marchal, P. 2002. Influence of trends in fishing power on bioeconomics in the North Sea flatfish fishery regulated by catches or by effort quotas. Canadian Journal of Fisheries and Aquatic Sciences, 59: 829–843.

Ulrich, C., Wilson, D. C. K., Nielsen, J. R., Bastardie, F., Reeves, S. A., Andersen, B. S., and Eigaard, O. R. 2012. Challenges and opportunities for fleet- and métier-based approaches for fisheries management under the European Common Fishery Policy. Ocean & Coastal Management, 70: 38–47.

Walters, C., and Martell, S. J. D. 2002. Stock Assessment Needs for Sustainable Fisheries Management. Bulletin of Marine Science, 70: 10.

Wang, S.-P., Sun, C.-L., Punt, A. E., and Yeh, S.-Z. 2005. Evaluation of a sex-specific age-structured assessment method for the swordfish, Xiphias gladius, in the North Pacific Ocean. Fisheries Research, 73: 79–97.

Waterhouse, L., Sampson, D. B., Maunder, M., and Semmens, B. X. 2014. Using areas-as-fleets selectivity to model spatial fishing: Asymptotic curves are unlikely under equilibrium conditions. Fisheries Research, 158: 15–25.

Whitmarsh, D. J., Reid, C. A., Gulvin, C., and Dunn, M. R. 1995. Natural Resource Exploitation and the Role of New Technology: a Case-history of the UK Herring Industry. Environmental Conservation, 22: 103–110.

Wilberg, M. J., Thorson, J. T., Linton, B. C., and Berkson, J. 2009. Incorporating Time-Varying Catchability into Population Dynamic Stock Assessment Models. Reviews in Fisheries Science, 18: 7–24.

Xu, H., Thorson, J. T., Methot, R. D., and Taylor, I. G. 2019. A new semi-parametric method for autocorrelated age- and time-varying selectivity in age-structured assessment models. Canadian Journal of Fisheries and Aquatic Sciences, 76: 268–285.

Xu, H., Thorson, J. T., and Methot, R. D. 2020. Comparing the performance of three data-weighting methods when allowing for time-varying selectivity. Canadian Journal of Fisheries and Aquatic Sciences, 77: 247–263.

Tables

Table 1. Description of model configurations employed. Model *Combined-Logistic* represents the null model in the current study and closely emulates the structure of the operational 2021 federal Alaska sablefish stock assessment (Goethel *et al.*, 2021). Model *Pot-Logistic* assumes a disaggregated fishery fleet structure and estimates logistic selectivity for the pot fishery fleet. Model *Pot-Gamma* also assumes a disaggregated fishery fleet structure but estimates gamma selectivity for the pot fishery fleet.

Model	Fleet structure	Selectivity functional form	Selectivity blocks	Biomass indices	Biomass index blocks	Parameters estimated
Combined-Logistic	Single fixed-gear fleet	Logistic selectivity	3 time-blocks from 1960-1994, 1995-2015, and 2016-2021	Aggregated biomass index (combines hook-and-line and pot-gear data)	2 time-blocks from 1995-2015, 2016-2021	251
Pot-Logistic	Disaggregated fleet structure	Hook-and-line Fleet: Logistic selectivity Pot Fleet: Logistic selectivity	Hook-and-line Fleet: 3 time-blocks from 1960-1994, 1995-2015, and 2016-2021 ($\delta_{s,f}$ shared with the pot fleet) Pot Fleet: Time-invariant (δ_s shared with the 2016-2021 hook-and-line time-block)	Hook-and-line index and pot index are fit independently	Hook-and-line Fleet: 2 time-blocks from 1995-2015, 2016-2021 Pot Fleet: 2 time-blocks from 2003-2016, 2017-2021	287
Pot-Gamma	Disaggregated fleet structure	Hook-and-line Fleet: Logistic selectivity Pot Fleet: Gamma selectivity	Hook-and-line Fleet: 3 time-blocks from 1960-1994, 1995-2015, and 2016-2021 Pot Fleet: Time-invariant	Hook-and-line index and pot index are fit independently	Hook-and-line Fleet: 2 time-blocks from 1995-2015, 2016-2021 Pot Fleet: 2 time-blocks from 2003-2016, 2017-2021	289

Figures

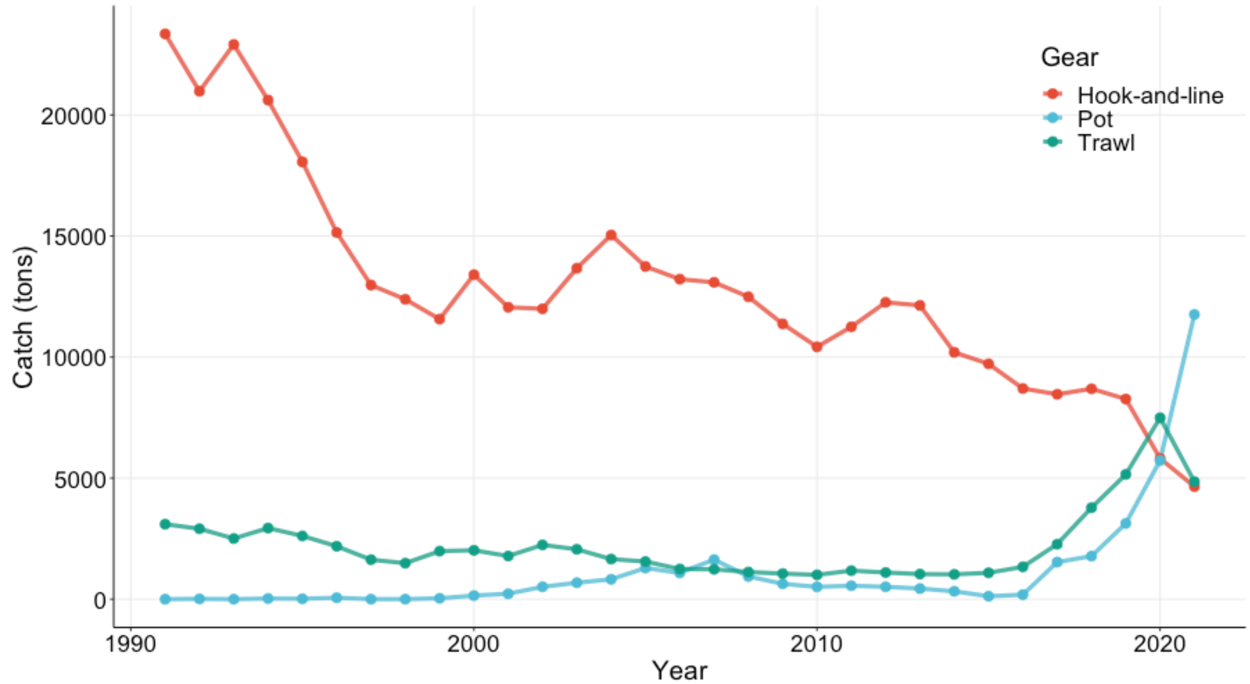

Figure 1. Total catch (tons) from 1990 to 2021 aggregated across sablefish management regions resulting from the hook-and-line, pot, and trawl fleets. Note that the fishery shifted from an open-access fishery to an Individual Fishing Quota (IFQ) program in 1995, and allowed pot-gear fishing in the Gulf of Alaska starting in 2017.

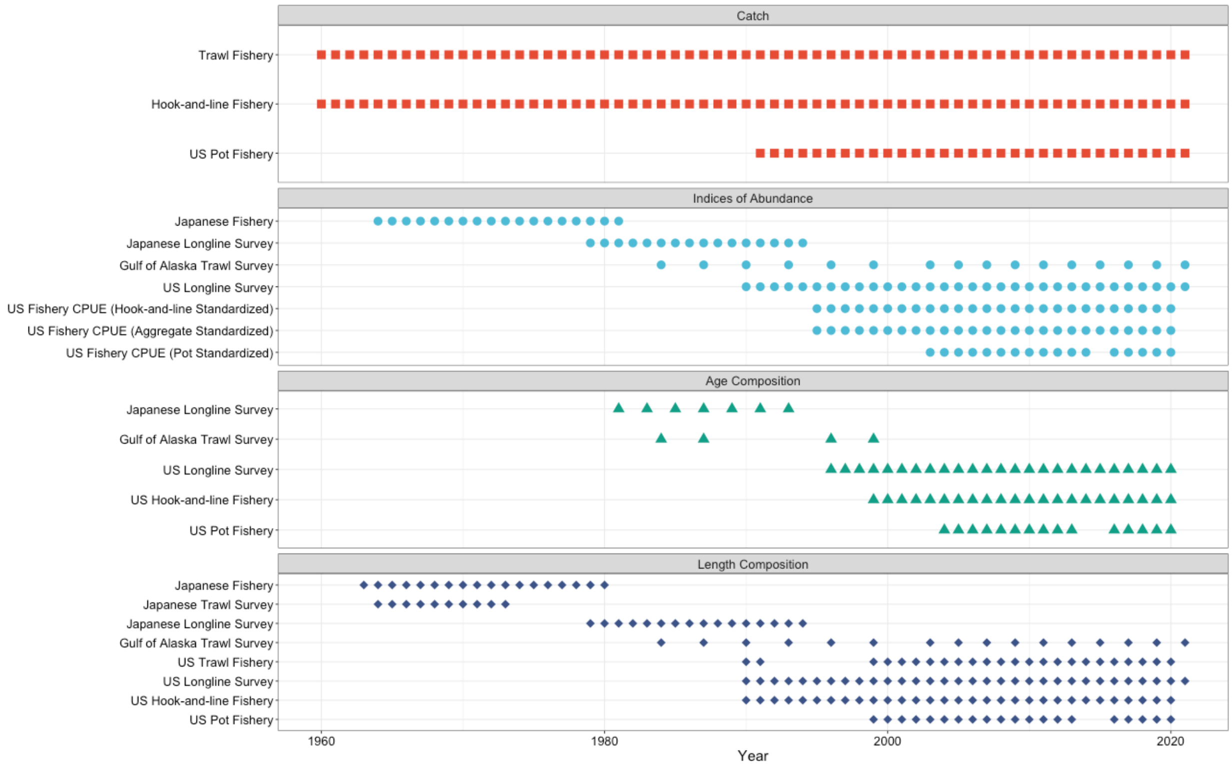

Figure 2. Presence and absence of all data types and sources that models variants are fit to in this study. Note that model *Combined-Logistic*, which closely emulates the 2021 operational assessment model is fit to a single aggregate standardized CPUE index that combines hook-and-line and pot-gear data. In contrast, models *Pot-Gamma* and *Pot-Logistic* are fit to two separate CPUE indices that are gear-specific. Presence of particular data types and sources are indicated by points; absences here are not assigned any points.

Figure 3. Time series of fishery-dependent indices incorporated (grey points and lines) for each model variant. Grey shading represents 95% confidence intervals and blue lines represent the time series to which a given model variant is fit to. Solid colored lines represent predicted values for a given index and assessment model variant.

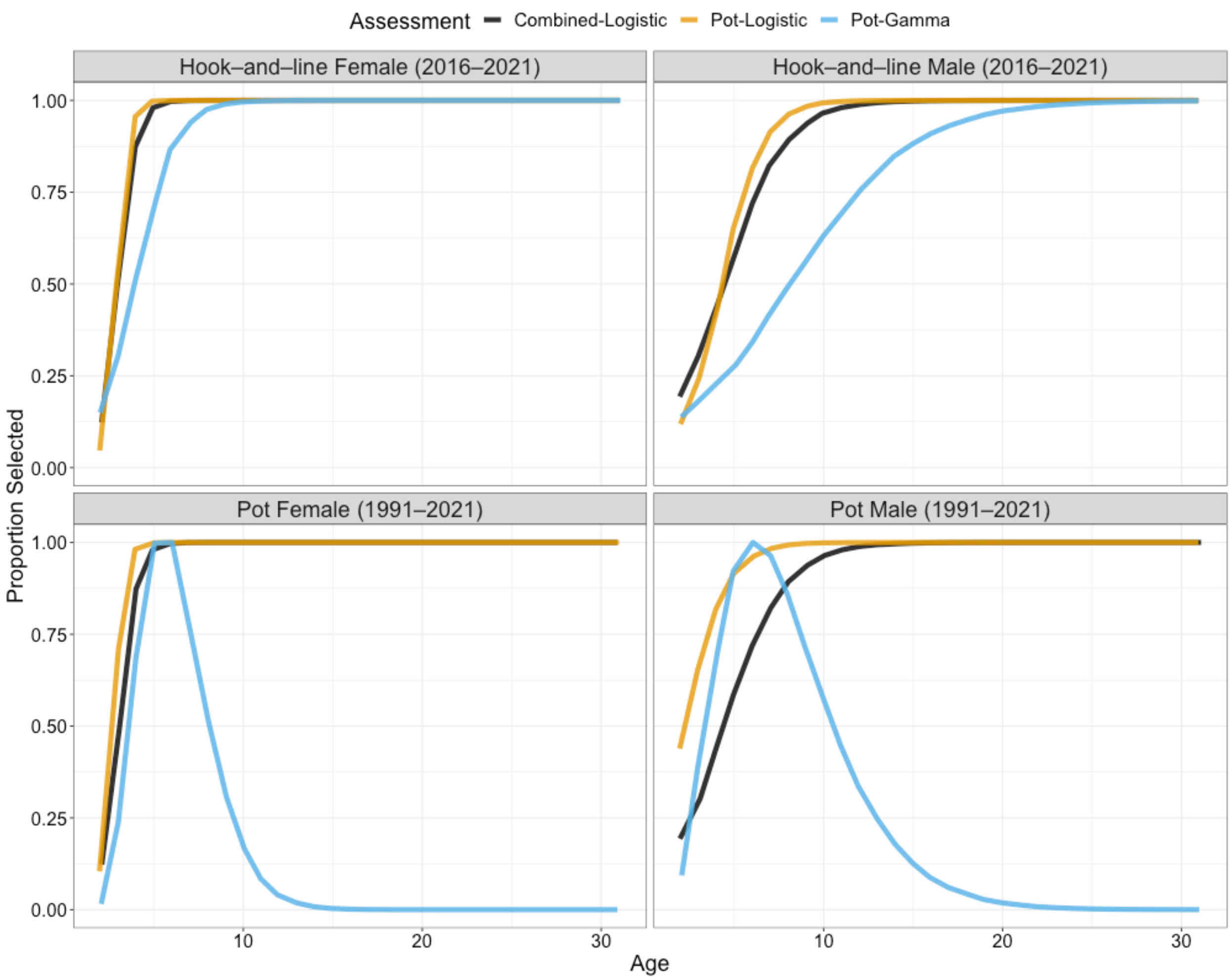

Figure 4. Estimated sex-specific selectivity curves for the hook-and-line and pot fisheries across explored model variants. Selectivities are scaled to have a maximum of 1.0. Selectivity for the hook-and-line fishery is estimated in three separate time-blocks (1960-1994, 1995-2015, 2016-2021), and pot selectivity is assumed to be time-invariant. The estimated selectivities for the fixed-gear fleet from model *Combined-Logistic* is plotted in all panels, given that it is informed by both hook-and-line and pot composition data.

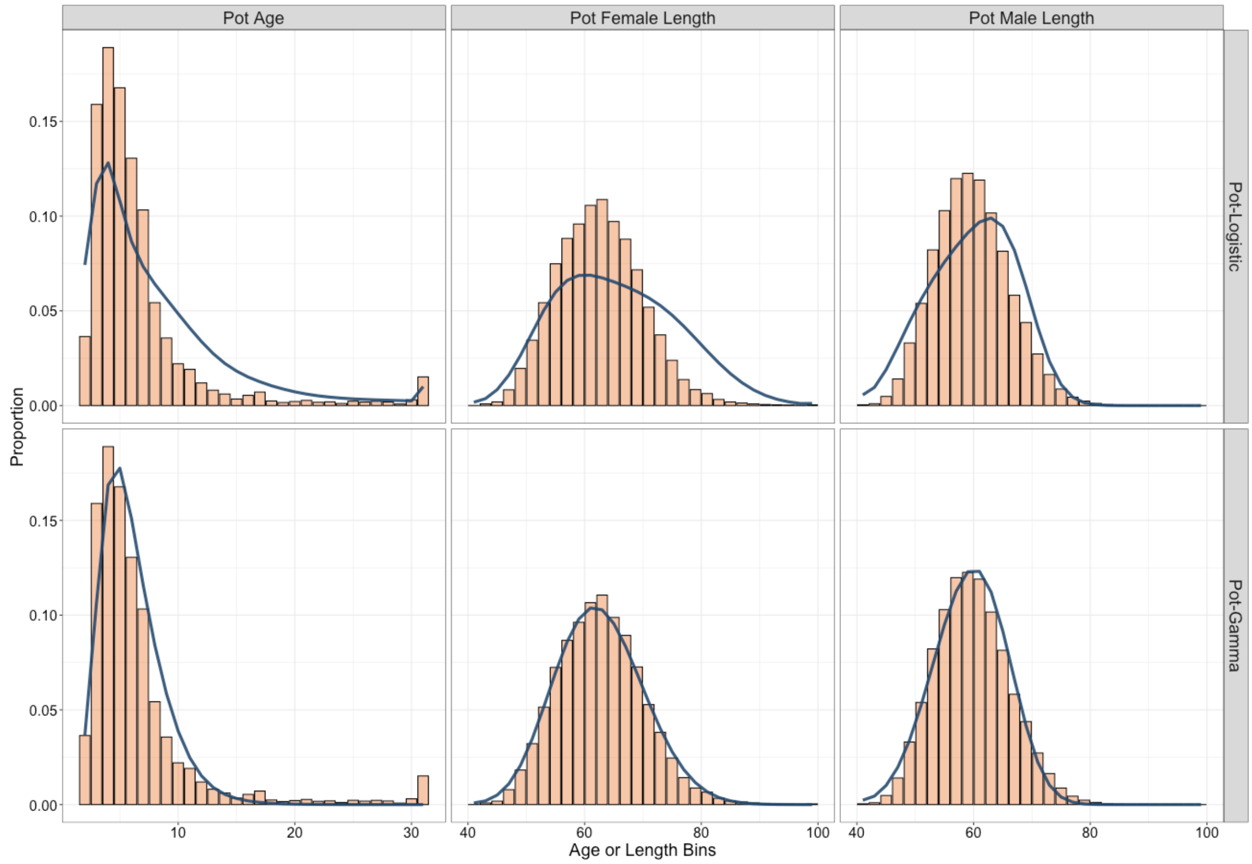

Figure 5. Average fits to compositional data resulting for the two pot fishery model variants explored in the present study. Orange bars represent the average observed proportion and blue lines represent the average model predicted proportion of a given age or length bin across time. Columns represent age-compositions, length-compositions for females, and length-compositions for males (left to right) from the pot fishery and rows indicate model variants (see Table 1 for descriptions of model variants). Note that age-composition data are input as sex-aggregated, while length-compositions are input as sex-specific.

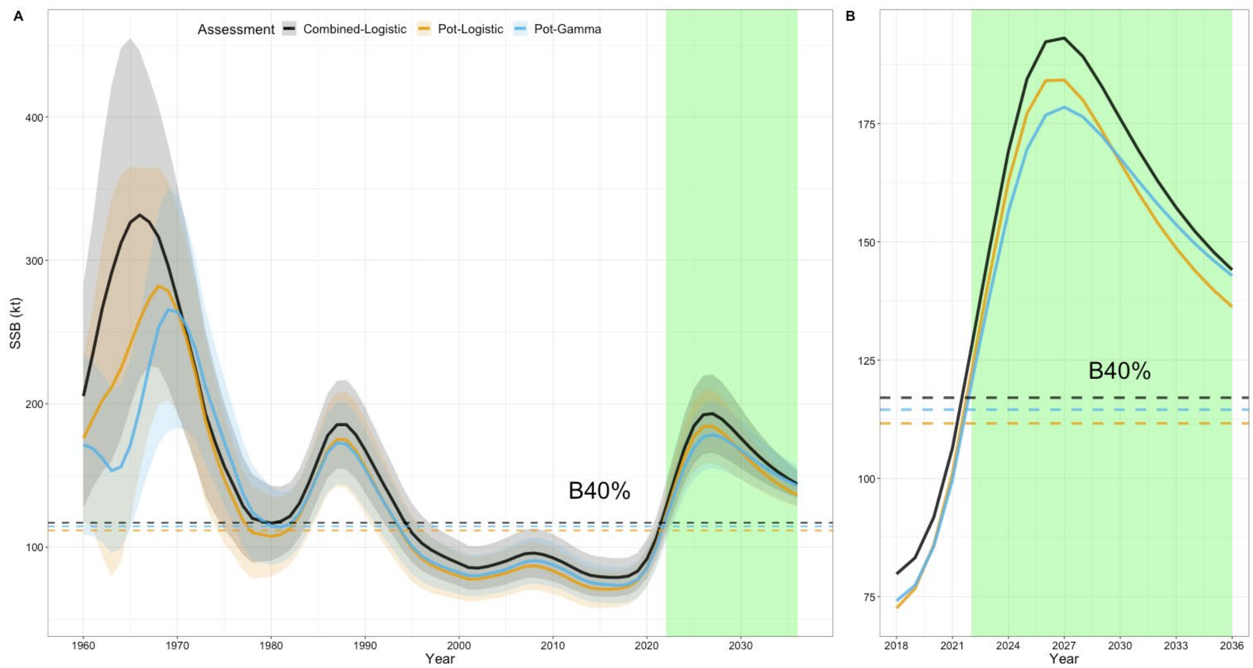

Figure 6. Estimated spawning stock biomass (SSB; solid lines) with associated asymptotic 95% confidence intervals (shading) and B40% reference points for 2021 (dashed lines) across model variants. Solid lines overlapping with green shading represent SSB projections for years 2022 – 2036 (15-year projections). Panel A shows SSB trends across the entire time-series. Panel B shows SSB trends from 2018 to 2036 to better highlight differences in projected SSB across model variants.

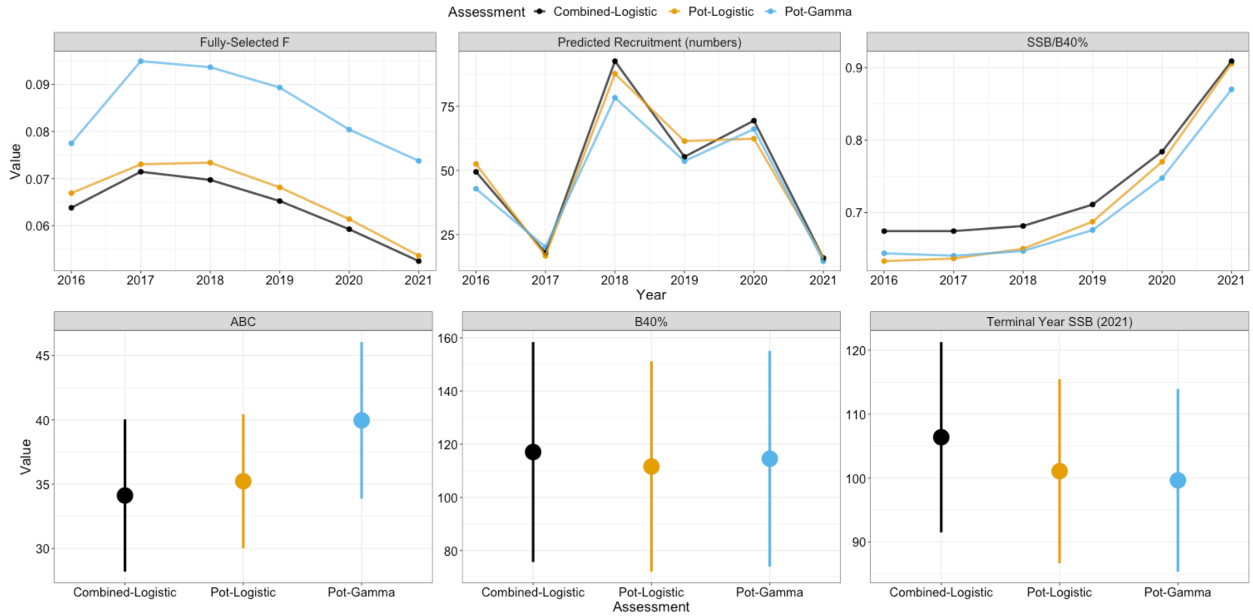

Figure 7. Time series of estimates for fully-selected fishing mortality (sum of fleet-specific fishing mortality rates), predicted recruitment, and stock status (SSB / B40%) in the upper row. Point estimates and associated asymptotic 95% confidence intervals for Acceptable Biological Catch (ABC), B40%, and terminal year (2021) SSB in the bottom row. ABC and B40% are determined internally within the stock assessment and represent the maximum ABC and 40% of unfished biomass, respectively.

Appendix A: Supplementary Figures

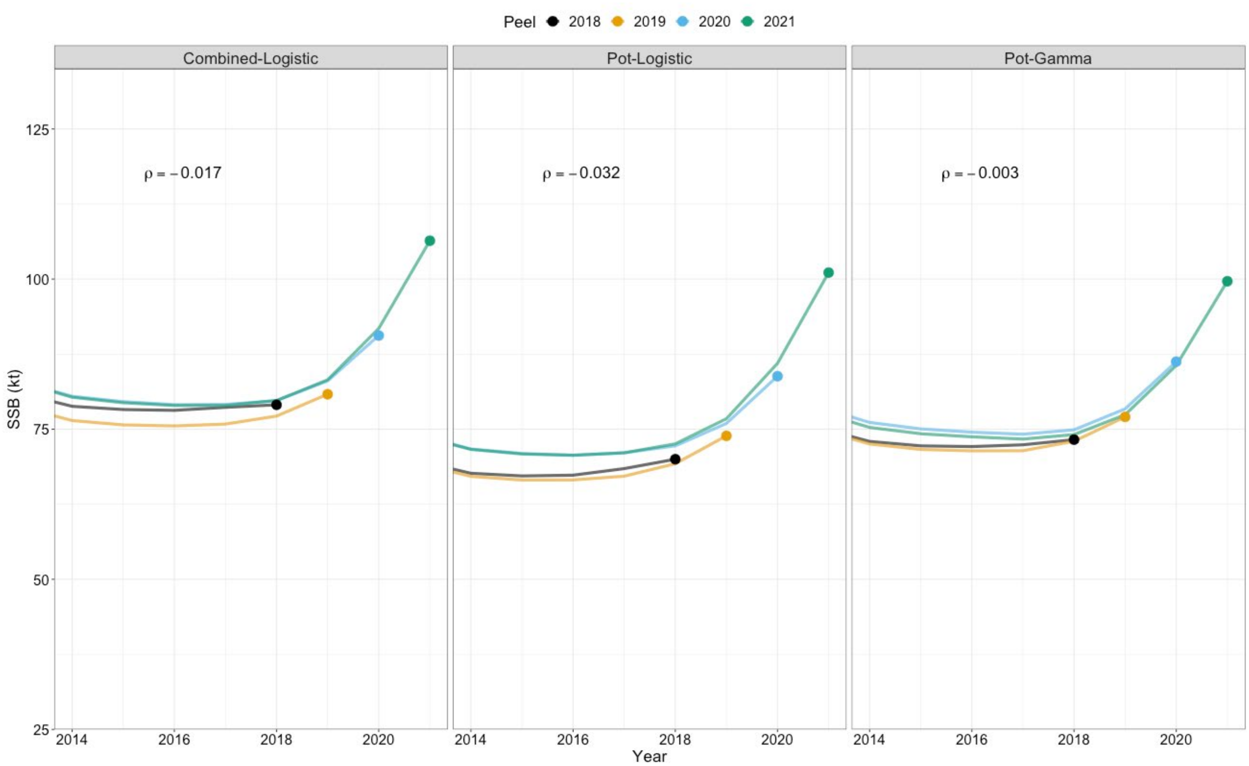

Figure A1. Retrospective patterns from 3-year "peels" of spawning stock biomass (SSB) for sablefish across model variants. Corresponding Mohn's ρ values from retrospective analysis are shown in each panel. Different colors represent estimates for individual "peels" and the estimates from the terminal year assessment (2021) are displayed in green.

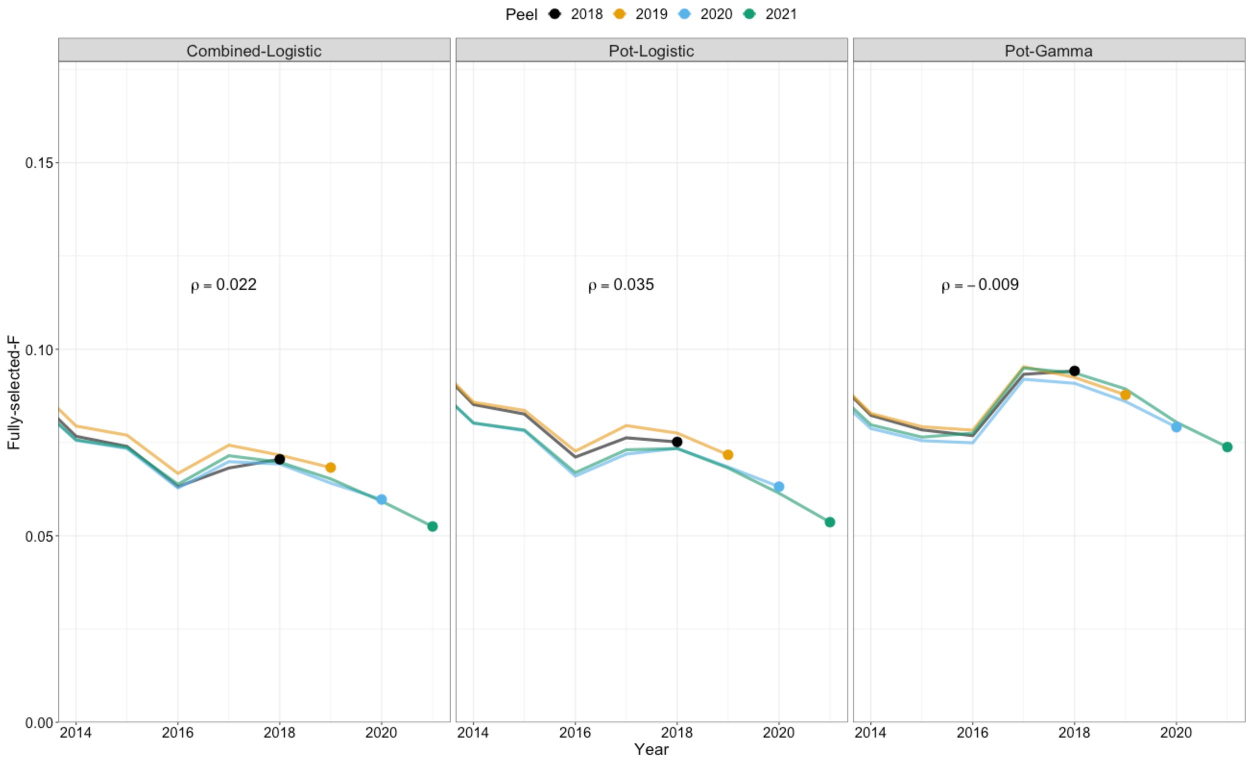

Figure A2. Retrospective patterns from 3-year "peels" of fully-selected fishing mortality rates for sablefish across model variants. Corresponding Mohn's ρ values from retrospective analysis are shown in each panel. Different colors represent estimates for individual "peels" and the estimates from the terminal year assessment (2021) are displayed in green.

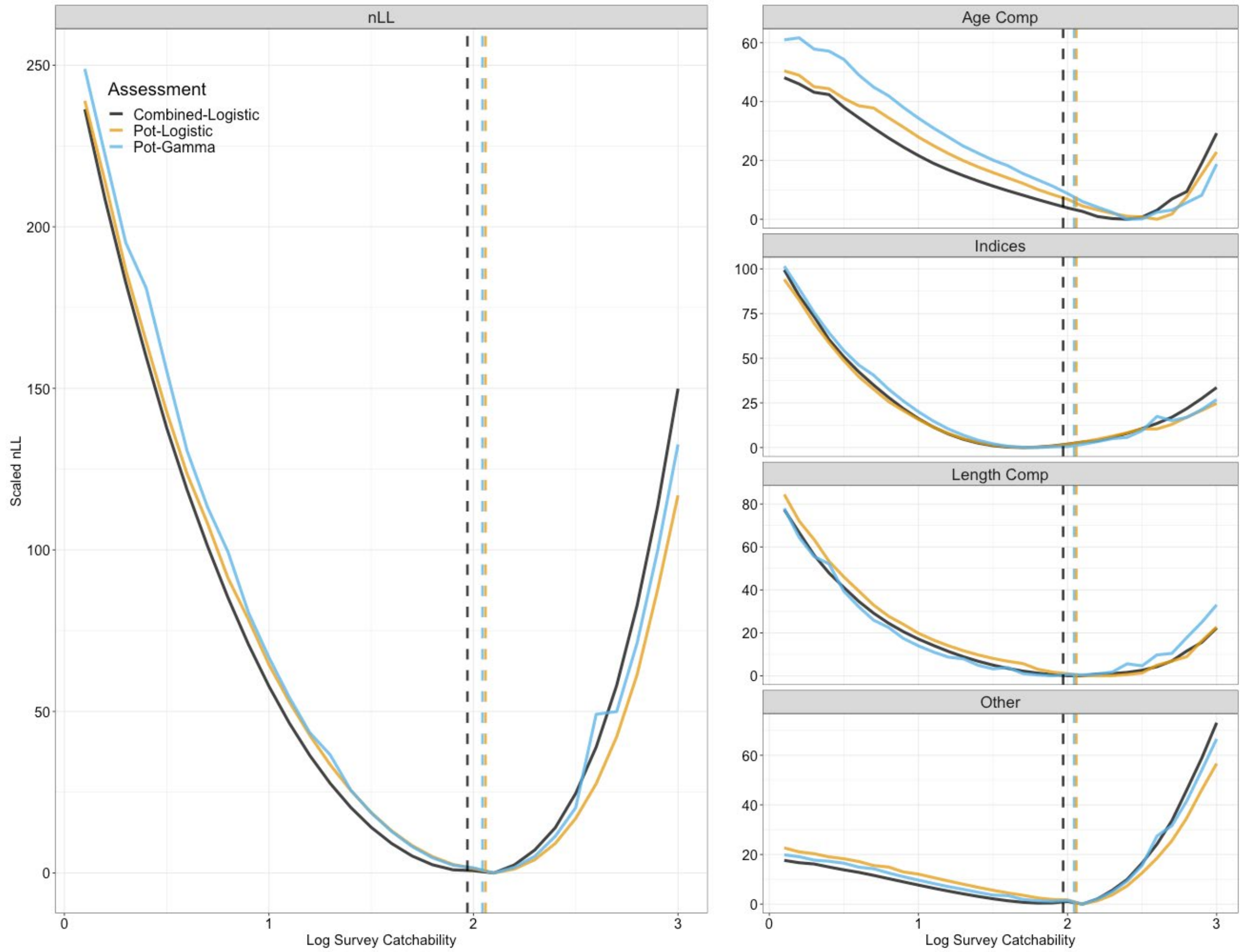

Figure A3. Likelihood profiles for the NMFS longline survey catchability. Catchability values were profiled across values of 0 – 3 in increments of 0.1. Negative log-likelihood (nLL) values for a given data type were scaled by their minimum value to ensure nLL values minimized at 0. Model variants are displayed in different colors, solid lines represent the likelihood profile, and dashed lines represent the maximum likelihood estimate of survey catchability for a given model.

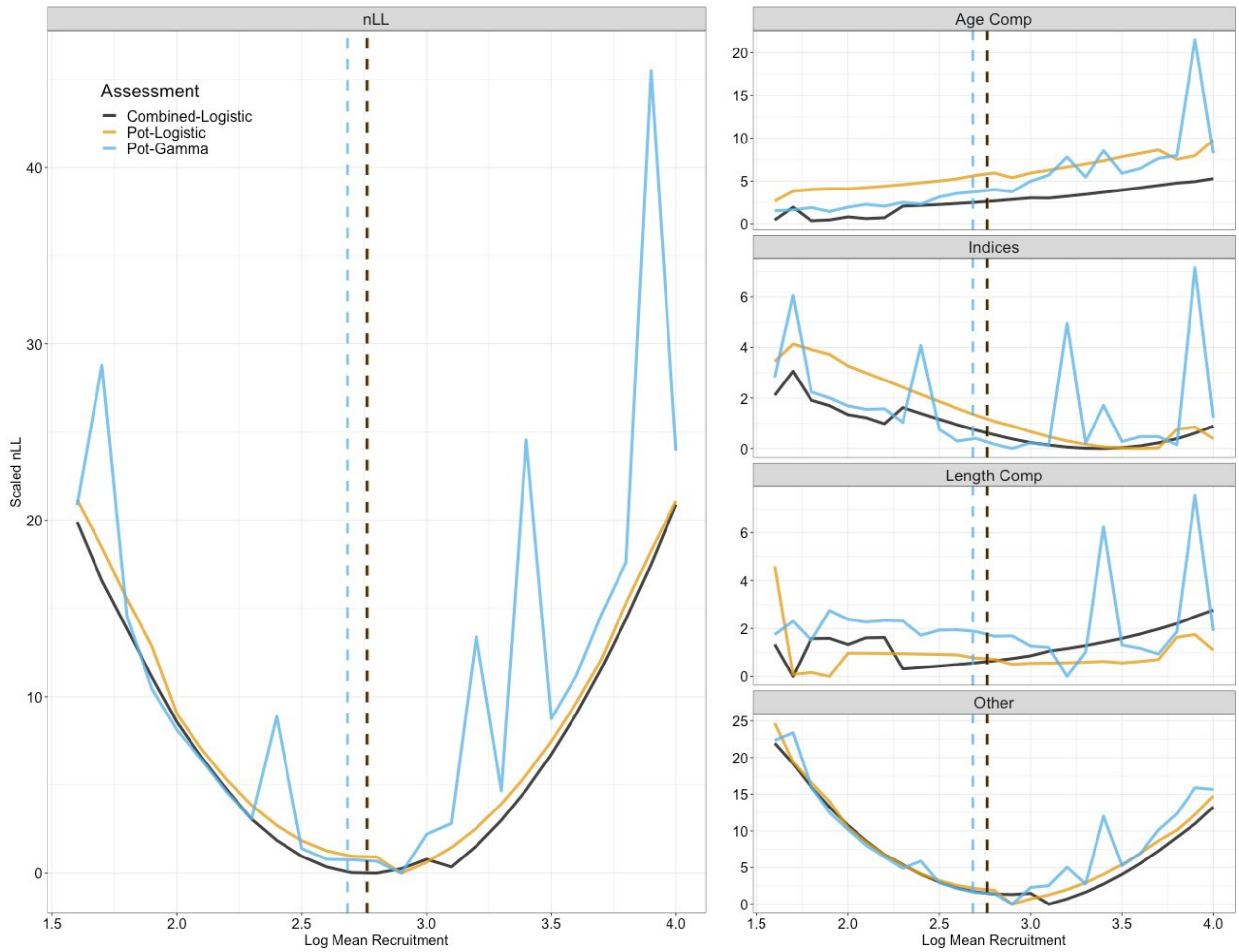

Figure A4. Likelihood profiles for the mean recruitment. Recruitment values were profiled across values of 1.5 – 4 in increments of 0.1. Negative log-likelihood (nLL) values for a given data type were scaled by their minimum value to ensure nLL values minimized at 0. Model variants are displayed in different colors, solid lines represent the likelihood profile, and dashed lines represent the maximum likelihood estimate of mean recruitment for a given model.

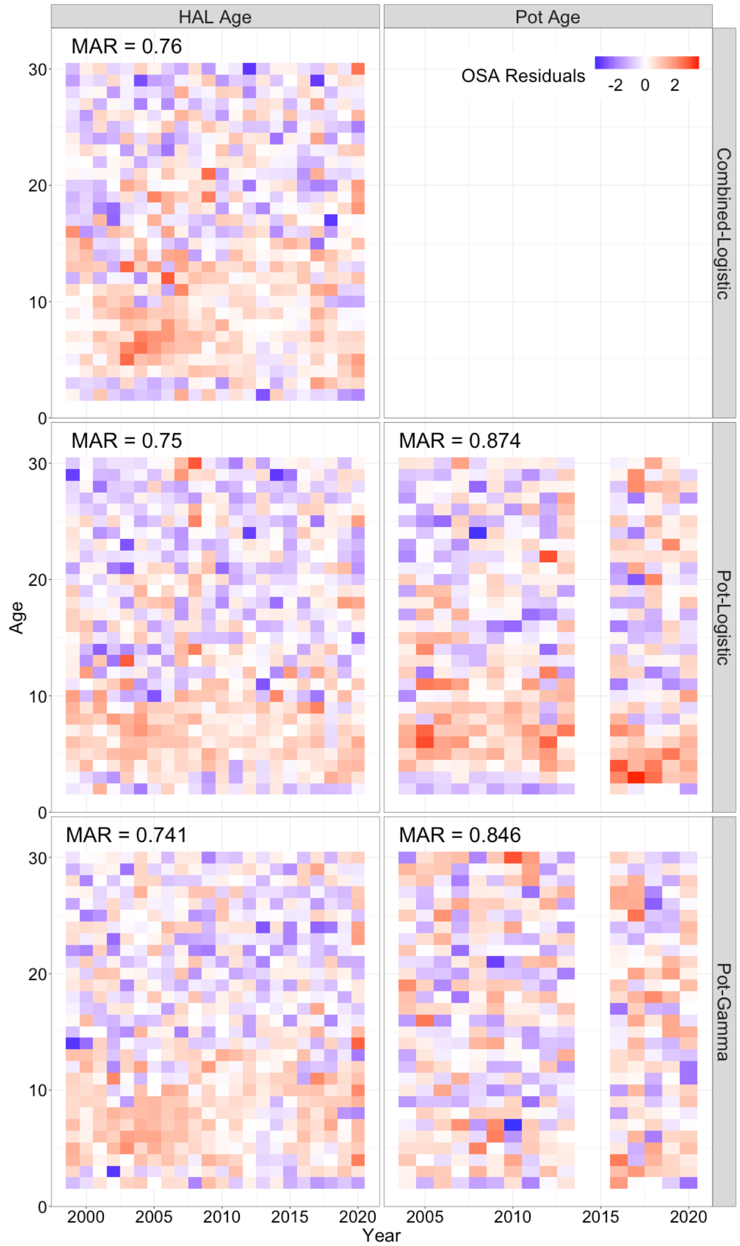

Figure A5. One-step ahead residuals across hook-and-line (HAL) and pot age-composition data (columns) across time (x-axis) and ages (y-axis) for all models evaluated in the study. Red colors are positive residuals and blue colors denote negative residuals. Mean absolute residuals (MAR) presented in the upper left corner of each panel represent the average absolute residuals for a given composition type and assessment model. Larger MAR values are indicative of a worse fit for a given assessment model to a composition type on average.

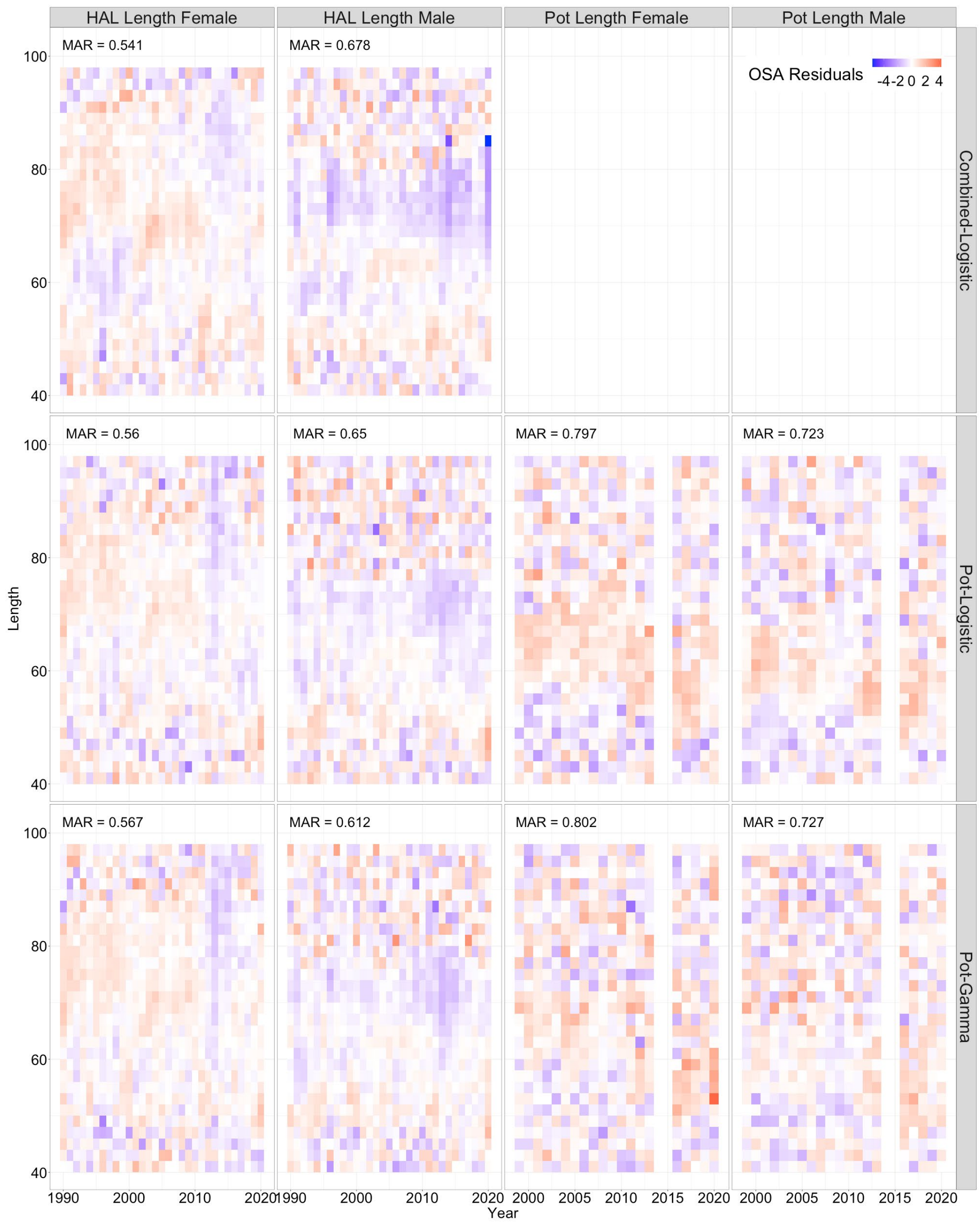

Figure A6. One-step ahead residuals across hook-and-line (HAL) and pot length-composition data (columns) across time (x-axis) and lengths (y-axis) for all models evaluated in the study. Red colors are positive residuals and blue colors denote negative residuals. Mean absolute residual (MAR) values presented in the upper left corner of each panel. Larger MAR values are indicative of a worse fit for a given assessment model to a composition type on average.

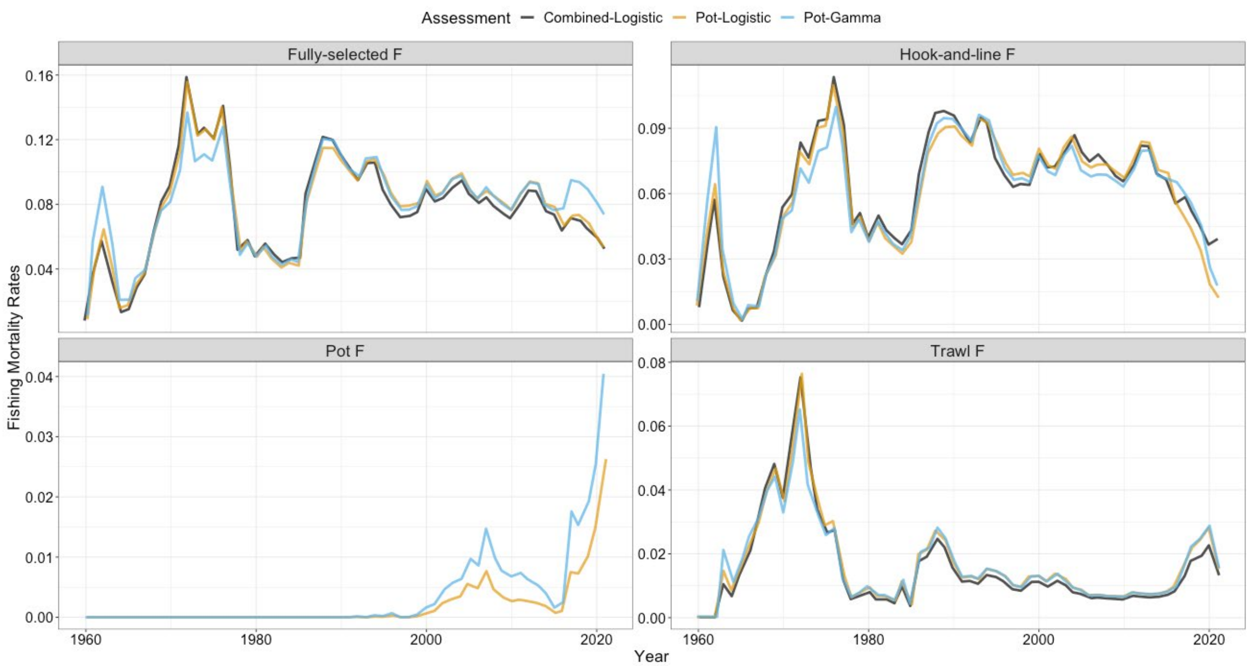

Figure A7. Time-series of fishing mortality rates from 1960-2021 across model variants. The panel denoted as "Fully-selected F" represents the sum of the fishing mortality rates across all fleets. Panels denoted by "Hook-and-line F", "Pot F", and "Trawl F" represent estimated fishing mortality rates for the hook-and-line (or fixed-gear fleet for model *Combined-Logistic*), pot, and trawl fishery, respectively. Note that the scale of the y-axis differs across panels.

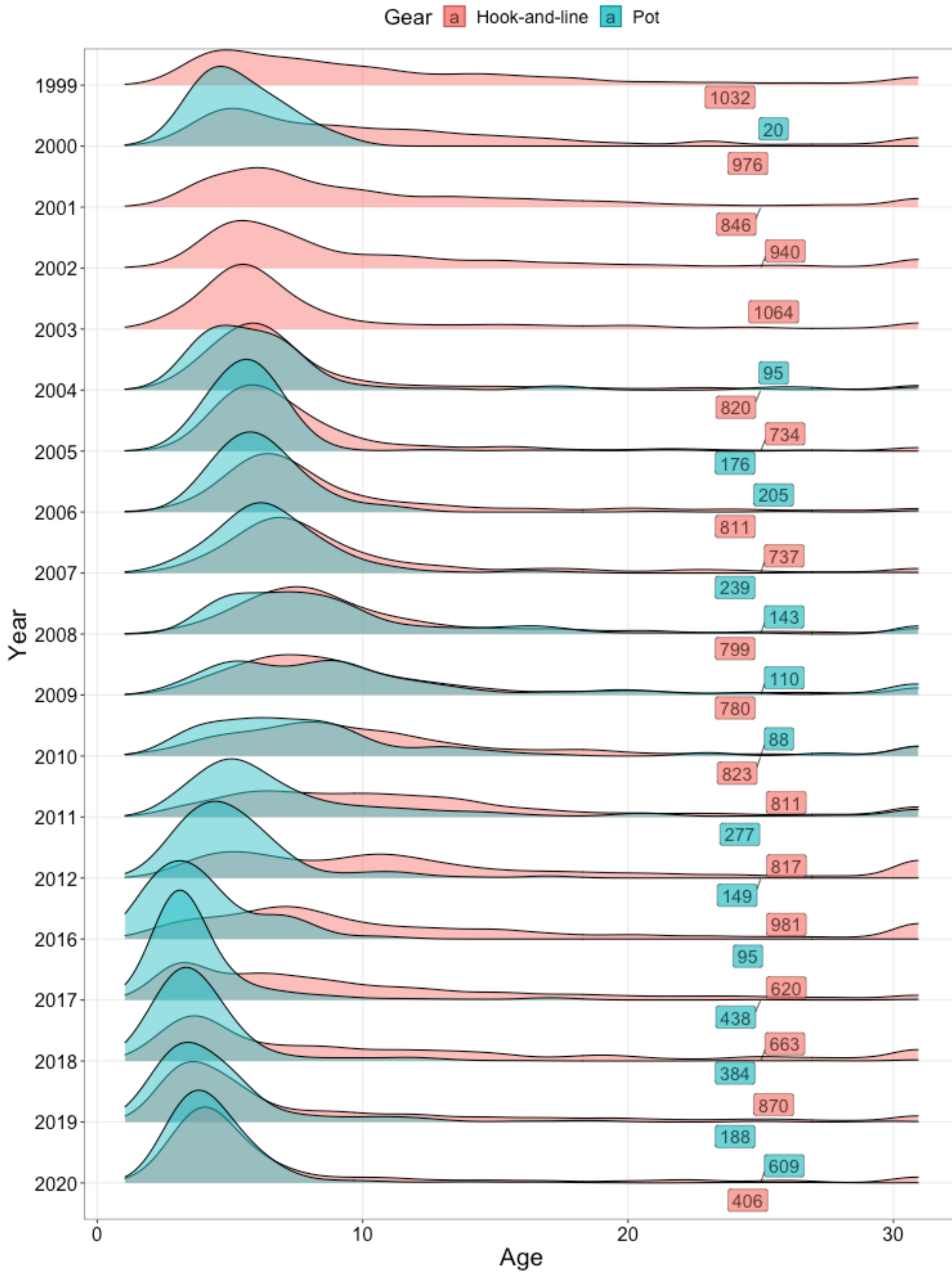

Figure A8. Distribution of ages sampled by hook-and-line gear and pot-gear. Colored labels denote the number of individuals aged for a given gear type.

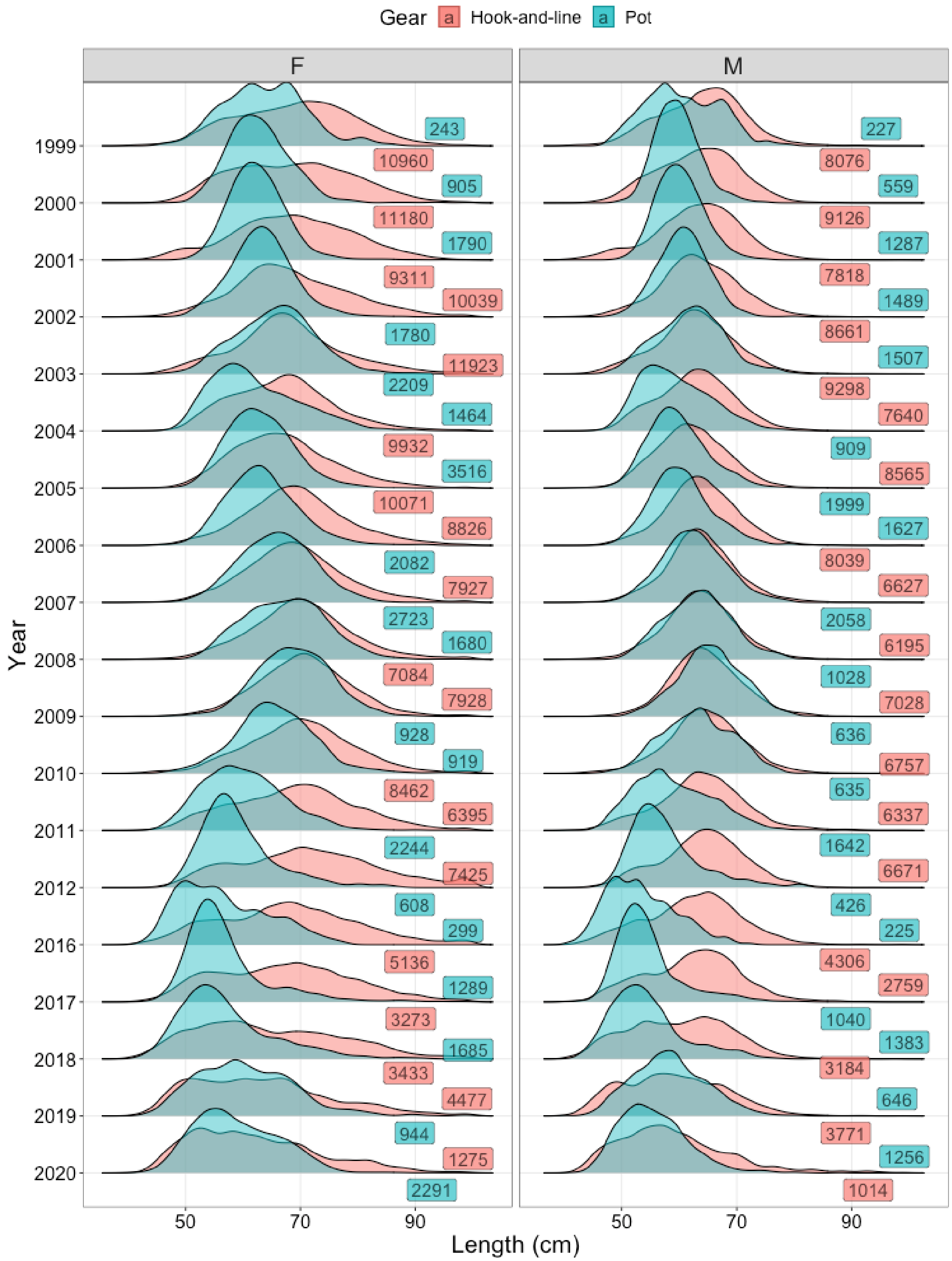

Figure A9. Distribution of lengths sampled by hook-and-line gear and pot-gear across sexes. Colored labels denote the number of individuals aged for a given gear type.

Appendix B: Model Description of the 2021 Federal Sablefish Stock

Assessment Model

General Model Description

The 2021 federal sablefish stock assessment is fit using an age-and sex-structured integrated model assuming a homogenous population in AD Model Builder. Hereafter, several equations will be presented, and definitions of symbols and variables can be found in Table 1 in this appendix. Initial abundance-at-age was determined by the following equation:

$$N_{1,a,s} = \begin{cases} R_1, & a = a_0 \\ e^{(\mu_R+\psi_y)} e^{-(a-a_0)\left(M+F_{hist}^{HAL} * s_{a,s}^{HAL_{1960-1995}}\right)}, & a_0 < a < a_+ \\ e^{(\mu_R)} e^{-(a_+-1)\left(M+F_{hist}^{HAL} * s_{a-1,s}^{HAL_{1960-1995}}\right)} \left[1 - e^{-\left(M+F_{hist}^{HAL} * s_{a-1,s}^{HAL_{1960-1995}}\right)}\right]^{-1} & a = a_+ \end{cases} \quad \text{(Eq. B1)}$$

where recruitment deviations are estimated for each cohort, and is decremented by natural mortality and historical fishing mortality rates resulting from the hook-and-line fishery up until the start of the assessment model (1960) (Goethel *et al*., 2021). The assessment assumes that a stock-recruitment relationship is not estimable (i.e., recruitment is independent of spawning stock biomass):

$$R_y = \begin{cases} e^{(\mu_R+\psi_y)}, & y \neq 2021 \\ e^{(\mu_R)}, & y = 2021 \end{cases}, \psi_y \sim \ln(0, \sigma_R) \quad \text{(Eq. B2)}$$

where recruitment deviates are constrained by a penalized likelihood following a lognormal distribution, with σ_R fixed at 1.2. Numbers-at-age starting in 1960 are determined by:

$$N_{y,a,s} = \begin{cases} R_y & a = 2 \\ N_{y-1,a-1} e^{-Z_{y,a,s}} & 2 < a < 31 \\ N_{y-1,a-1} e^{-Z_{y-1,a-1}} + N_{y-1,a} e^{-Z_{y-1,a}} & a = 31 \end{cases} \quad \text{(Eq. B3.1)}$$

$$Z_{y,a,s} = \sum_f F_{y,a,s,f} + M \quad \text{(Eq. B3.2)}$$

where numbers-at-age in Eq. B3.1 are decremented by total mortality (sum of fishing and natural mortality; Eq. B3.2) and follows a forward projection method. Natural mortality in the assessment is estimated with an informative prior (mean = 0.1, CV = 10%). Catch data in the assessment is predicted using Baranov's catch equation:

$$C_{y,a,s,f} = \frac{F_{y,a,s,f}}{Z_{y,a,s}} N_{y,a,s}(1 - e^{-Z_{y,a,s}}) w_{a,s} \quad \text{(Eq. B4.1)}$$

$$F_{y,a,s,f} = e^{(\mu_f + \rho_{y,f})} * s_{y,a,s,f} \quad \text{(Eq. B4.2)}$$

where Eq. B4.1 is Baranov's catch equation and describes predicted catch as the ratio of fishing mortality and total mortality multiplied by the number of individuals that experienced mortality in year *y*. Eq. B4.2 imposes a separability assumption, where annual fishing mortality rates are multiplied by the selectivity of fleet *f*, to estimate age-specific vulnerabilities. Catch data for a given fleet were assumed to follow a lognormal distribution. Predicted catch-at-age and catch-at-length was given by:

$$P_{y,a,s,f} = N_{y,a,s} s_{y,a,s,f} \left(\sum_{a=2}^{a=31} N_{y,a,s} s_{y,a,s,f} \right)^{-1} \mathbf{A}_{s} \quad \text{(Eq. B5.1)}$$

$$P_{y,a,s,f} = N_{y,a,s} s_{y,a,s,f} \left(\sum_{a=2}^{a=31} N_{y,a,s} s_{y,a,s,f} \right)^{-1} \mathbf{A}_{s}^{l} \quad \text{(Eq. B5.2)}$$

where catch-at-age is multiplied by an ageing error matrix (Fig. B1) to account for uncertainty in the ageing process (Eq. B5.1). For predicted catch-at-length, proportions were determined following Eq. B5.2 and was multiplied by an age-to-length transition matrix, to allow for the age-structured model to fit to sex-structured length-composition data. Age-and length-composition for all fisheries were assumed to follow multinomially distributed errors, with assumed input sample sizes of 20. Given inherent correlations in composition data, input sample sizes were smaller than observed sample sizes to reflect reduced information content resulting from such correlations (Pennington and Volstad, 1994; Francis, 2011). Integrated stock assessments are fit a variety of data sources and are sensitive to input data weights (Maunder and Piner, 2017). Furthermore, multinomial distributions do not allow for correlations that are commonly observed in age-or length-composition data (Francis, 2017). To reconcile these complexities, we applied Francis-reweighting to all explored model variants (Francis, 2011). Data weights for compositional data were determined following a 2-stage approach using method TA1.8 and weighting assumption T3.4 (multiplicative weighting) as described in Francis, 2011. The 2-stage reweighting approach was conducted until data weights and key management quantities appeared converged (Francis, 2017). Preliminary explorations indicated that the relative weights (weights are applied on an aggregate dataset) determined by Francis-reweighting and resulting model estimates were fairly insensitive to the assumed input sample sizes. Abundance/biomass indices were also assumed to follow a lognormal distribution, and the predicted index for a given year was given by:

$$\widehat{I_{y,f}} = q_{y,f} \sum_{a=2}^{a=31} \sum_{1}^{s} N_{y,a,s}\, s_{y,a,s,f} w_{a,s} \qquad \text{(Eq. B6)}$$

For indices of abundance that are represented as numbers, weight-at-age for sex *s* was not included in Eq. B6. Fishery-dependent indices in the current study assumed a coefficient of variation of 10%, as is done in the 2021 federal sablefish stock assessment.

Several data sources are fit within the assessment model. Here, we only describe those that represent an important component of the assessment, but readers should refer to (Goethel *et al.*, 2021) Specifically, the assessment is fit to age-and length-composition data from both the fixed-gear fishery (hook-and-line and pot) and the annual sablefish longline survey, both of which assume logistic selectivity:

$$s_{y,a,s,f} = \left[1 + e^{-\delta_{y,s,f}\left(a - a_{y,s,f}^{50\%}\right)}\right]^{-1} \quad \text{(Eq. B7)}$$

where the fixed-gear fishery assumes three time-blocks in both selectivity and catchability (1960-1994, 1995-2015, 2016-2020) to account for various shifts in management structure and large recruitment events. The assessment is also fit to catch data and length-composition data resulting from the trawl fishery following a re-parameterized gamma function:

$$s_{y,a,s,f} = \left(\frac{a}{a_{y,s,f}^{max}}\right)^{\left(\frac{a_{y,s,f}^{max}}{p}\right)} e^{\frac{a_{y,s,f}^{max} - a}{p}} \quad \text{(Eq. B8.1)}$$

$$p = 0.5 * \left[\sqrt{a_{y,s,f}^{max\,2} + 4\gamma_{y,s,f}^{2}} - a_{y,s,f}^{max}\right] \quad \text{(Eq. B8.2)}$$

where γ (shape parameter) is shared between sexes, to achieve stable model results. Finally, the model is also fit to a biomass index and length-composition from a biennial bottom trawl survey, which assumes a one parameter power function for selectivity:

$$s_{y,a,s,f} = a^{\phi_{f,s}} \quad \text{(Eq. B9)}$$

All selectivities that are included in the model are scaled to have a maximum of 1.

Tier 3 North Pacific Fishery Management Council (NPFMC) Harvest Control Rule

Alaska sablefish are managed under the Tier 3 NPFMC harvest control rule (sloping control rule), which utilizes proxy reference points for maximum sustainable yield (MSY). Specifically, these references points are *B*40%, which represents the long-term average biomass that would be expected under mean recruitment conditions and fishing mortality rates occurring at *F*40%. These reference points are determined from spawning per recruit ratios which represent the ratio between two lifetime egg productions (fished cohort divided by unfished cohort), and ranges between 0 and 1. The resulting catch advice is:

$$F_{\text{ABC}} = \begin{cases} F40\% & \text{if} \dfrac{SSB_{y+1}}{B40\%} > 1 \\ \dfrac{F40\% \left(\dfrac{SSB_{y+1}}{B40\%}\right) - \lambda}{1-\lambda} & \text{if} \dfrac{SSB_{y+1}}{B40\%} < 1 \\ 0 & \text{if} \dfrac{SSB_{y+1}}{B40\%} < \lambda \end{cases} \qquad \text{(Eq. B10)}$$

where the total SSB_{y+1} is the projected spawning stock biomass in the year following the terminal year of the assessment, while assuming mean recruitment and mortality rates from the terminal year of the assessment (fishing and natural mortality). λ is defined as the fraction of $\frac{SSB_{y+1}}{B40\%}$ below which fishing does not occur, and is defined as 0.05 here.

Table 1. Symbols and descriptions of variables for equations used for the sablefish stock assessment model in this study.

Symbol	Description
$N_{y,a,s}$	Abundance for year *y* (1960-2021), age a (2, 3, 4 … 31_{+}) and sex *s* (male or female)
a_0, a_+	Age at recruitment (age 2) and age of plus-group (age 31) respectively
R_y	Recruitment for year *y*
μ_R	Mean log recruitment
ψ_y	Annual recruitment deviation
σ_R	Recruitment variability fixed at 1.2
M	Time-invariant natural mortality
μ_f	Mean log fishing mortality rate for fleet *f* (hook-and-line, trawl, or pot)
$\rho_{y,f}$	Annual fishing mortality deviation for year and fleet *f*
F_{hist}^{HAL}	Historical fishing mortality from the hook-and-line fishery
$F_{y,a,s,f}$	Instantaneous fishing mortality rate for year *y*, age *a*, sex *s,* and fleet *f*
$s_{y,a,s,f}$	Proportion selected for year *y* (estimated as time-blocks), age *a*, sex *s,* and fleet *f*
$a^{50\%}$	Midpoint parameter for a logistic function describing age at 50% selection
δ	Shape parameter describing the rate of increase for a logistic function
a^{max}	Parameter for a re-parameterized gamma function describing age at maximum selection
γ	Shape parameter for a re-parameterized gamma function describing rate of decrease for the descending limb
p	Derived power parameter for a reparametrized gamma function
ϕ	Parameter that determines the slope of the power function
$C_{y,a,s,f}$	Predicted catch (tons) for year *y*, age *a*, sex *s,* and fleet *f*
$Z_{y,a,s}$	Total instantaneous mortality for year *y*, age *a*, sex *s*
$w_{a,s}$	Average weight at age *a* and sex *s*
$P_{y,a,s,f}, P_{y,l,s,f}$	Predicted proportions at age *a* or length *l* (41, 43, 45 … 99) respectively, for year *y*, sex, *s*, and fleet *f*
$\mathbf{A_s}, \mathbf{A_s^l}$	Ageing error matrix and age-to-length transition matrix for sex *s*, respectively

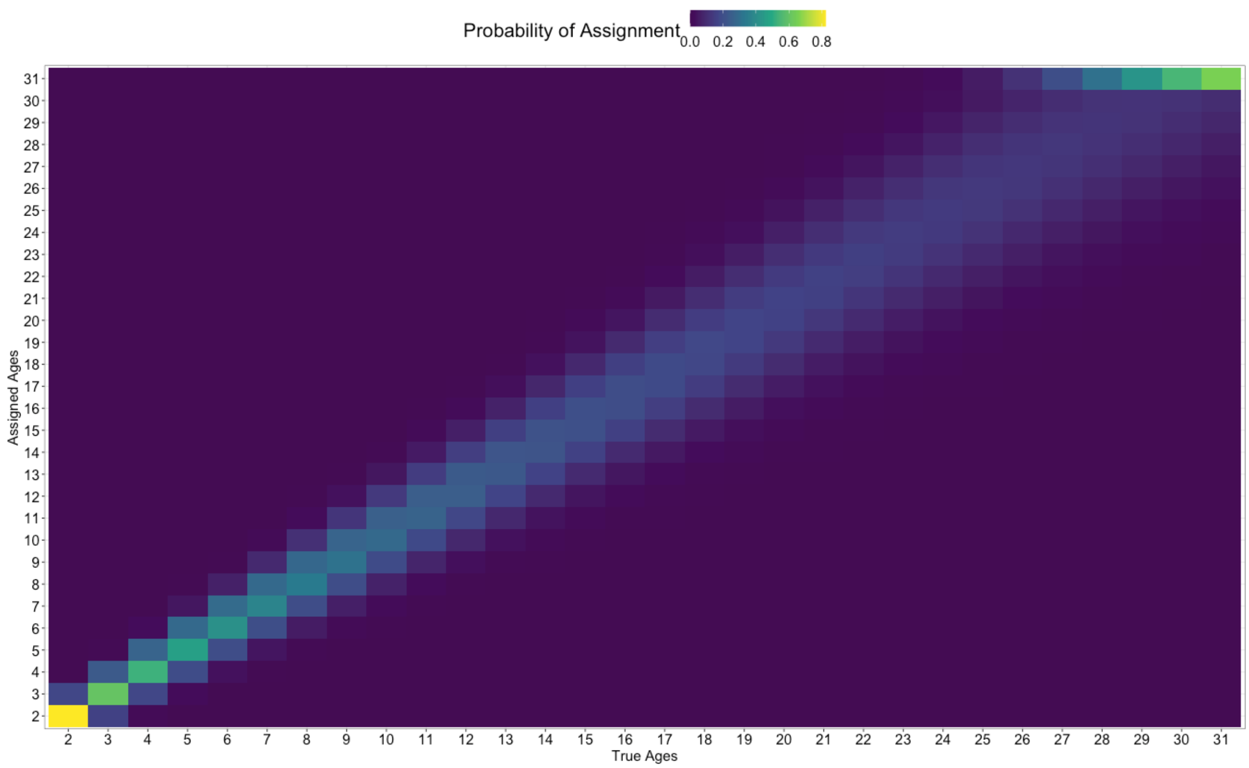

Figure B1. Ageing error matrix used in the 2021 operational sablefish assessment model. True ages are denoted on the x-axis, while reader assigned ages are denoted on the y-axis. Colors represent the probability of assignment to a given age-class.

Appendix B: References

Francis, R. I. C. C. 2011. Data weighting in statistical fisheries stock assessment models. Canadian Journal of Fisheries and Aquatic Sciences, 68: 1124–1138.

Francis, R. I. C. C. 2017. Revisiting data weighting in fisheries stock assessment models. Fisheries Research, 192: 5–15.

Goethel, D., Hanselman, D., Rodgveller, C., Echave, K. B., Williams, B., Shotwell, S. K., Sullivan, J., *et al.* 2021. 3. Assessment of the Sablefish Stock in Alaska: 347.

Maunder, M. N., and Piner, K. R. 2017. Dealing with data conflicts in statistical inference of population assessment models that integrate information from multiple diverse data sets. Fisheries Research, 192: 16–27.

Pennington, M., and Volstad, J. H. 1994. Assessing the Effect of Intra-Haul Correlation and Variable Density on Estimates of Population Characteristics from Marine Surveys. Biometrics, 50: 725.