Fisheries and Oceans Canada | Pêches et Océans Canada
Ecosystems and Oceans Science | Sciences des écosystèmes et des océans

Pacific Region

Canadian Science Advisory Secretariat
Science Response 2025/013

AN INVESTIGATION OF AGEING REQUIREMENTS TO SUPPORT THE BRITISH COLUMBIA SABLEFISH (*ANOPLOPOMA FIMBRIA*) OPERATING MODEL

CONTEXT

Since the inception of the Management Strategy Evaluation (MSE) process in 2011 (Cox et al. 2011), the Sablefish operating model (OM) has informed the British Columbia (BC) fishery's harvest strategy approach by assessing the stock against reference points and simulating future stock dynamics and fishery responses. Originally, the MSE process was initiated to address the decline of the Sablefish spawning stock after the 1980s. Despite implementing management changes in 2011, the resulting recovery was slow, and it wasn't until high recruitments began in 2016-2018 that the spawning stock recovery was assured. These 1 - 3 record high year classes moved the stock above optimal biomass levels in a short period, and by 2022 the spawning stock was estimated to be 1.32 times the female spawning biomass at maximum sustainable yield, B_{MSY} (DFO 2023a).

Current scientific challenges to the Sablefish MSE process pivot on ensuring the OM can continue to provide reasonable estimates of stock status in the presence of record high recruitment from the 2016-2018 year classes (DFO 2023a). Accurately estimating abnormally large and small recruitments is critical and carries asymmetric risks because overestimation or underestimation can lead to long-term stock status risks or economic setbacks, respectively.

OM estimates of recruitment strength rely on age composition data from research surveys and commercial fisheries, as does the estimation of fleet-specific gear selectivity parameters that characterize the age and size at which fish are caught. However, recent reductions in sample sizes of age composition data, especially from commercial sources, have increased reliance on less direct data via Bayesian priors on selectivity parameters developed from a mark-recapture tagging program (Cox et al. 2023). It is therefore essential to evaluate how sensitive the OM is to the quantity of age data and to strategically allocate ageing resources among survey and fishery sample sources. Proper age data usage is key to managing the effect of recruitment uncertainty in stock assessments, thereby ensuring that operating models are able to (a) provide timely estimates of stock status that are relatively robust to multiple sources of uncertainty, and (b) simulate realistic stock and fishery monitoring data for testing management procedure performance.

This Science Response process was initiated by DFO Science in the Pacific Region to increase understanding of how age composition sample size and the allocation of ageing effort among age data sources affects Sablefish operating model performance with respect to bias and precision of management parameters. This evaluation is undertaken in the context of the full Sablefish management system by using closed-loop feedback projections that apply the current Sablefish management procedure (MP) to simulated data each year to determine annual catch levels. Establishing a scientifically defensible rationale for both the number of ages and allocation of ages to data sources will inform the annual Sablefish ageing request. While these analyses are specific to Sablefish, the framework developed is relevant to stock assessment programs for multiple species groups that face similar challenges due to limited ageing capacity.

Overall, results show that, as expected, increased age sample sizes produce lower parameter uncertainty in selectivity estimates, especially for age compositions derived from otolith sampling during the annual Sablefish research survey. However, under a scenario where the stock collapses due to a persistent recruitment failure, the sample size and source of Sablefish age composition data does not appear to have a strong influence on detecting spawning stock biomass or status with respect to the limit reference point (LRP). We discuss options for improving the response of the Sablefish management system to persistent low recruitment in the Conclusions section.

Inability to detect persistent recruitment failures notwithstanding, results under average recruitment conditions suggest that sample sizes in the range of 1200 - 1600 ages / year are sufficient to support the current Sablefish operating model. Bias and precision for a subset of key model parameters have little improvement above 1600 total samples, and deteriorate at a faster rate when total sample size drops below 1200 aged fish. The identified lower limit of 1200 samples / year aligns with recent samples sizes, while the upper value of 1600 samples / year is similar to sample sizes from the years 2010-2016.

This Science Response Report results from the regional peer review of March 25, 2024 on the Effects of Age Composition Data on Operating Model Performance for Sablefish (*Anoplopoma fimbira*) in British Columbia.

BACKGROUND

Sampling of Sablefish Age Composition

The age of Sablefish in BC is determined using calcium carbonate structures inside the head of the fish called sagittal otoliths (a.k.a. ear bones). Highly trained age readers can estimate fish age by counting the annuli (annual rings) exposed on otoliths when following the ‘break and burn’ method (Chilton and Beamish 1982; Hanselman et al. 2012). Sablefish are difficult to age because of their small and irregularly shaped otoliths, which can cause difficulties in locating the first annulus.

Otoliths were removed from Sablefish collected for biological sampling during the offshore Sablefish stratified random component of the annual research survey (referred to as the Stratified Random Survey, or StRS), as well as from a subset of mainland inlets (Lacko et al. 2023a). In addition, otoliths are collected from commercial Sablefish catch. For Sablefish longline trap and longline hook fleets (referred to as ‘trap’ and ‘longline’ respectively in the remainder of this document), otoliths were collected as part of a voluntary catch sampling program (Lacko et al. 2023b). For the trawl fleet, otoliths were collected under the at-sea observer program between 1996 and 2020. Since 2020, a gap in data collection from the trawl fleet has emerged due to the move to at-sea electronic monitoring in place of at-sea observers, which began due to social distancing requirements during the Covid-19 pandemic (Walker et al., In prep[1]). At present, over 4000 otoliths are collected from the Sablefish survey and catch sampling programs each year, which is a reduction from the over 6000 that were collected in the early 2000s (Figure 1). Further reductions are expected in the future due to recent changes to the survey bisampling protocol. In 2023 (not shown in Figure 1), 2900 otoliths were collected from the survey. Survey biosampling rates will be revisited again in a few years time to evaluate whether further changes are warranted.

Only a portion of collected otoliths are aged each year due to limited age reader resources. The number of age readings per year was highest between 2009 and 2016, during which time the

[1] Walker, L.C., Rogers, L.A., Anderson, S.A., and Haggarty D.R. In review. A review of biological samples collected from commercial groundfish fisheries, 1996-2022. Can. Tech. Rep. Fish. Aquat. Sci.

Sablefish MSE process was being developed. Annual numbers have dropped to around 1000-1200 in recent years.

Ageing effort has been primarily focused on ageing samples collected from the annual research survey; however, ageing of otoliths from commercial fisheries has also occurred (Figure 1). Among commercial fisheries, ageing effort has historically focused on ageing trap fleet otoliths due to the high proportion of catch attributed to the trap gear type. However, in recent years, the proportion of Sablefish taken by longline hook gear has increased such that trap and longline fleet catches are now similar. While sampling rates from longline fisheries have been low or zero, recent changes in catch sampling protocols and program communication have resulted in more samples, which affords future opportunities to estimate longline fleet selectivity directly from age composition data rather than depending solely on tag-release recovery data (Johnson et al. 2025).

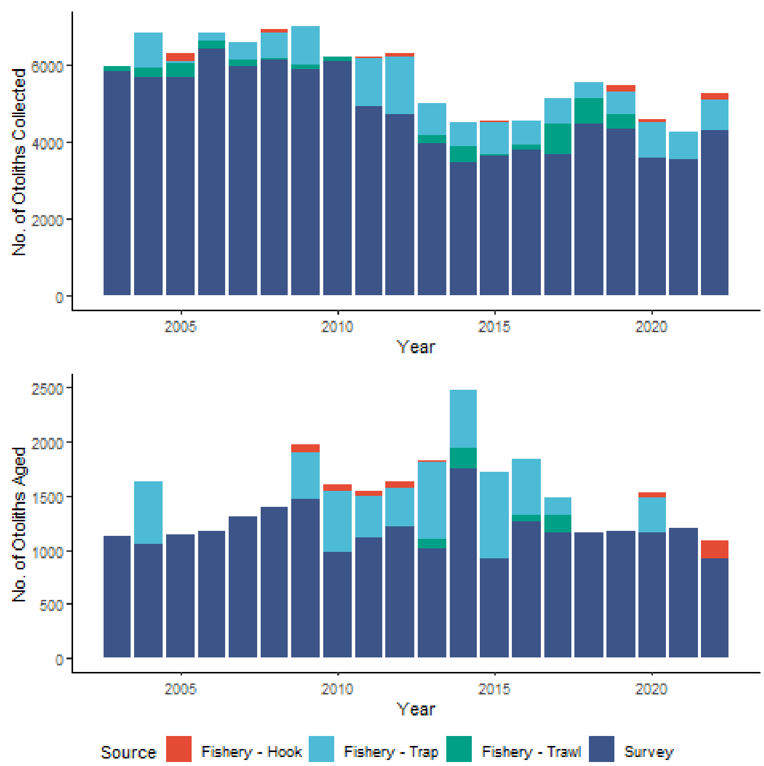

Figure 1. Number of otoliths collected from field sampling by year (top) and number of otoliths for which ages have been estimated each year (bottom), by gear type.

Sablefish Operating Model

The Sablefish operating model is a sex-structured statistical catch-at-age model that incorporates: (i) gear-specific landed catches (trap, longline hook, and trawl), (ii) at-sea releases from each gear type, (iii) three abundance indices (two historical, one current from the StRS), (iv) age compositions from the trap fleet, StRS, and the defunct standardised survey, and (v) length composition data from the trawl fleet (DFO 2023a). A complete description of the data can be found in the most recent operating model update (Johnson et al. 2025).

Selectivity-at-length within the model is estimated by fitting to age and length compositions. Age composition data are available for the StRS and trap fleet, with sex-specific selectivity-at-age determined from a sexually dimorphic model of Sablefish growth. Similarly, the trawl fleet selectivity is estimated directly from length composition data. Finally, age- or length- data have not been collected in sufficient quantities for estimating longline hook selectivity, which instead relies exclusively on Bayesian selectivity priors. Bayesian prior distributions for selectivity functions for all three commercial sectors are derived from the long-running Sablefish tag release-recovery study (Jones and Cox 2018; Cox et al. 2023).

Accurate estimation of abnormally large and small recruitments is a critical requirement for the Sablefish OM. Recent high recruitments for BC Sablefish are associated with higher catches of sub-legal Sablefish (<55cm fork length), which in turn creates increased discarding and discard-induced mortality. Such increases in sub-legal mortality complicate stock reconstructions and projections, given evidence that discard data are diverging from existing model hypotheses (DFO 2023a). Reliable age composition data from multiple fleets are likely required to help distinguish between recruitment events and changing selectivity during model fitting.

A peer-review of the Sablefish model advocated for enhanced age composition sampling from the longline hook and trawl fleets to refine selectivity estimates for both gears (DFO 2023a). Although currently unavailable, the inclusion of these age composition data could potentially enhance stock assessment precision. This Science Response focuses solely on including additional age data from the longline fleet, as there is limited information available for developing the same analyses for trawl data, and a plan for age or length sampling of Sablefish from the trawl fleet under the electronic monitoring program has not been developed. Trawl length compositions are a reasonable age substitute for now as trawl gear intercepts a higher proportion of smaller fish (ages 1-2) than longline fisheries. The smaller difference in size at age for these age classes makes it easier to distinguish recruiting year classes in trawl catches than in trap and longline fleets, which tend to select for older fish with highly variable size-at-age.

ANALYSIS AND RESPONSE

Linking Sample Size to Age Sampling Error Variance

Simulating the effect of alternative Sablefish age sampling designs required a model that relates age composition sample sizes to age sampling error variance. Age sampling error variance is an output of the Sablefish operating model, which was fitted to age data using a logistic-normal age composition likelihood (Schnute and Haigh 2007; Francis 2014), where the age-sampling error variance is the variance of log-residuals between the fitted model and age composition data. As with most modeling approaches, the fitted logistic normal compositional likelihood function could have been used to simulate new data, with assumed data quality varied by increasing or lowering the residual variance used to simulate the data. However, simply assuming alternative values for the variance would not have provided guidance on the sample sizes required to target the desired level of model performance, since there was no explicit link between the simulated variance and the age sample size in the Sablefish OM. Therefore, we developed a model to link age sampling variance to sample size for both the commercial trap fleet and StRS age composition data. For simulating longline hook data, we assumed the same variance and sample size model as estimated for the trap fleet for two reasons. First, the available longline hook age data are limited and therefore excluded from the Sablefish OM. Second, factors contributing to age sampling error variance other than sample size are likely to be similar between trap and hook vessels.

We simulated age sampling error variance based on a given sample size via a Dirichlet process. A Dirichlet process is a stochastic process that generates random discrete probability distributions from an underlying 'base distribution' H and a concentration factor α_0. For the purpose of simulating Sablefish age-composition data for each fleet and sex, the base distribution H is the true vulnerable proportion-at-age $\vec{p} = (p_1, ..., p_K)$, where $K = 35$ is the number of age-classes, $p_i \in [0,1]$, and $\sum_i p_i = 1$. To describe the concentration factor α_0 it helps to first define a Dirichlet distribution. Dirichlet distributions are a multi-variate generalisation of the Beta distribution, and are defined by an order $K \geq 2$ (where K = number of age-classes) and a parameter vector $\vec{\alpha} = (\alpha_1, ..., \alpha_K)$, where $\alpha_i > 0$. Random draws from a Dirichlet distribution are K-dimensional vectors of proportions (i.e., proportions-at-age) $\vec{x} = (x_1, ..., x_K)$, such that $x_i \in [0,1]$ and $\sum_{i=1}^{K} x_i = 1$. For a random variable $\vec{X} = (X_1, ..., X_K) \sim Dir(\vec{\alpha})$, the expected value of each entry in $\vec{X}$ is

$$E[X_i] = \frac{\alpha_i}{\alpha_0}, \tag{1}$$

where $\alpha_0 = \sum_{i=1}^{K} \alpha_i$. The quantity α_0 is the concentration factor for the Dirchlet process $DP(\vec{p}, \alpha_0)$ described above. Distributions drawn from $DP(\vec{p}, \alpha_0)$ may be parameterised as $Dir(\alpha)$, where $\vec{\alpha} = \alpha_0 \cdot \vec{p}$, and for $\vec{X} \sim Dir(\alpha)$, $E(X) = \vec{p}$ is the true population vulnerable proportion-at-age. Random draws $\vec{X} \sim Dir(\vec{\alpha})$ can be considered observed or sampled proportions-at-age with uncertainty, which represents a realisation of the total catch-at-age from the vulnerable population, or alternatively the total of all sampling, across all vessels in a fleet. Of those intermediate sampled proportions-at-age $\vec{X}$, a certain number N of otoliths are sampled, which are modeled as N repeated draws from a Categorical distribution $Cat(K, \vec{x})$, where $\vec{x}$ is a realisation of the random variable $\vec{X} \sim Dir(\alpha_0 \cdot \vec{p})$. The Dirichlet distribution is sometimes applied in fisheries stock assessment models as a prior for the multinomial likelihood used to fit to compositional data (Thorson et al. 2017) because it is the conjugate prior for the multinomial/categorical distribution, and it has a good structure for describing the distribution of catch-at-age samples. We estimated separate concentration factors α_0 for male and female Sablefish in both the trap fishery and StRS survey (referred to as fleets). For each combination of sex and fleet, the Sablefish OM estimates a conditional maximum posterior density estimate (cMPDE) of age sampling error variance over all years of data, subject to a tail-compression process that accumulates observations and model expected values for age classes where age samples make up less than 2% of the total annual sample. The resulting cMPDE of variance was used as the target variance for simulated age composition data. Optimal concentration factors, α_0^*, that produced the target variance were identified using a grid search over α_0 values in the regular grid $G = \{10, 20, 30, ..., 3000\}$. For each candidate $\alpha_0 \in G$ we estimated the associated age sampling error variance, on average, over 100 random draws, developing data for the response of age sampling variance to the concentration factor. The grid search procedure used the following steps for each fleet and sex combination with age observations:

1. For each year t with ageing data, use the operating model numbers-at-age and fleet selectivity-at-age s_a to derive the base distribution $\vec{p}_t = (p_{1,t}, ..., p_{35,t})$ as

$$p_{a,t} = \frac{s_a N_{a,t}}{\sum_a s_a N_{a,t}} \tag{2}$$

   where a is the age-class from 1 to the plus group age of $A = 35$, t is the time step in 1965 - 2021, and $N_{a,t}$ is the estimated number of Sablefish in age-class a at time t.

2. Adjust the base measure $p_{a,t}$ by the ageing error matrix $Q_a a$, i.e., $\vec{p'}_t = Q_a a \cdot \vec{p}_t$.

3. For each candidate value of α_0 in the grid, repeat the following for replicate $i = 1, ..., 100$:

   a. For each year t with age composition data:

   i. take a random draw of age-class probabilities $\vec{\rho}_{i,t} \sim Dir(\alpha_0 \cdot \vec{p'}_t)$,

   ii. draw an age sample from the categorical distribution $Cat(A, \vec{\rho})$ with sample size equal to the observed sample size for year t

b. Compute σ_i^2 over all years with data as the cMPDE of the variance term in the logistic normal compositional likelihood function for age composition data (Johnson et al. 2025), using the base measure $\vec{p}$ as the expected value.

4. Calculate the $\sigma^2 =$ median σ_i^2 for each candidate α_0 in the grid.

The resulting median age sampling error variance estimates for each α_0 in the grid search define a response curve of age sampling error variance over the grid. The final estimate of α_0^* for each fleet/sex combination used for simulations in the remainder of this paper was defined by the point where the median age sampling residual variance associated with α_0 equaled the target variance (i.e., the operating model estimate of σ^2), with a spline used to interpolate the response curve between grid points. See DFO (2020) for description of the ageing error matrix in step (2), which was used to correct for how the true proportions-at-age (i.e., proportions-at-age in the operating model) get mis-classified to adjacent age classes when interpreting otolith patterns.

The Dirichlet processes we estimated for the StRS and trap fishery fleets indicated the expected response of uncertainty to age sample size (i.e., uncertainty decreased with increasing sampling size; Figure 2). Although the annual age sampling variance (calculated from the logistic normal residual errors) was noisier than the conditional MPDE would suggest (Figure 2, top row), this may have been more related to individual sample size in a given year rather than operating model specification or data likelihoods (Figure 2, third row). Estimates of α_0^* range from 222 (male, trap fishery) to 903 (male, StRS), and taken collectively over all sex/fleet combinations, α_0^* was negatively correlated with variance. That is, higher concentration factors were required to achieve lower target variances, and vice versa. For Trap fishery data, male α_0^* values were lower than female, reflecting the higher residual errors for males caught by the trap fishery. In contrast, the order of male/female concentration factor values reversed in the StRS, reflecting poorer fits to female age data during recent years following the large recruitment event(s). Finally, for the given concentration factors under each fleet/sex combination, the modeled relationship showed that age-composition residual standard error decreases with sample size, as expected, and appeared to be approaching a lower limit at higher sample sizes, with most simulated standard errors above 0.2. Overall, the model captures the cMPDE of age sampling standard errors within the range of simulated uncertainty at the average sample size for each fleet/sex combination (Figure 2, third row, red points). These occured at sample sizes of 458 and 526 for StRS male and female sources, respectively. For commercial trap males and females, the average sample sizes were 288 and 331.

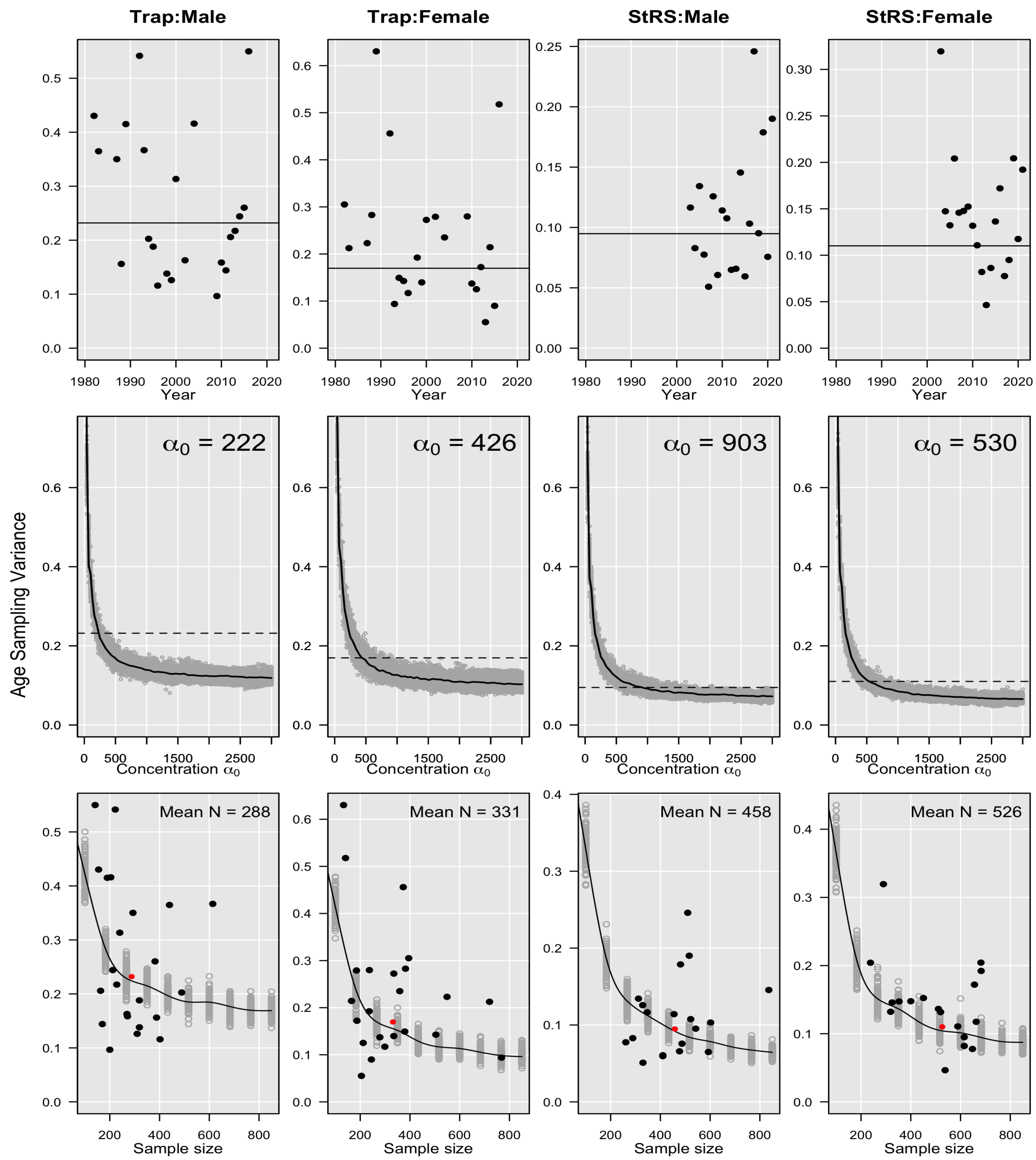

Figure 2. The stages of estimating the Dirichlet process for each fleet (StRS or trap fishery) and Sablefish sex (male and female). The top row shows yearly estimates of age-sampling error variance (black points) as well as the conditional maximum posterior density estimate (cMPDE) calculated over all years (black line). The second row shows the simulated relationship between the Dirichlet process concentration parameter α_0 and age sampling variance found via repeated random draws (grey points show individual simulation replicates and black line is the median for each concentration factor in the grid), with the optimised value of α_0^* shown in the upper right corner. The horizontal dashed line shows the cMPDE of age sampling variance from the operating model. The third row shows the relationship between sample size and age sampling variance (grey points show individual simulation replicates and black line is a spline through median values over the sample size grid) based on the estimated α_0^* values, as well as realised values from the operating model fits (black points), and the red point shows the average sample size compared to the OM cMPDE of variance.

Based on visual inspection, the Dirichlet process approach to simulating age composition data appeared to produce realistic observation uncertainty (Figure 3 and Appendix). There was a similar level of irregularity or roughness in the historical and simulated age compositions (i.e., the proportions did not create a nice smooth distribution). Draws from the categorical distributions defined by the Dirichlet process were less prone to very large outliers where a single age class formed the majority of the composition, which was more common under a logistic normal distribution where errors are unbounded. Additionally, empty age classes occured when the Dirichlet was used, because some low probability ages in the categorical distribution were not drawn. While there was some smoothness in the simulated StRS age frequency distributions (Appendix Figures 9, 10), samples for the StRS had a higher sample size than the trap fishery and the underlying categorical distributions for the StRS were drawn from a Dirichlet process with a higher concentration factor than the trap fishery, both of which implied lower variance, i.e., more smoothness. The simulated trap fishery age compositions had a similar visual pattern to real fishery data and did not appear smooth (Appendix Figures 11, 12), reflecting the smaller sample sizes in both historical and simulated data.

The simulated data acceptably captured the distribution of StRS age sampling with quantiles matching, on average, over the historical data set (Figure 3, top row). There are cases where the simulated data were more or less dispersed than the observed data (i.e., QQ plot trend is flatter or steeper, respectively, than the 1-1 line), but they appeared well balanced with respect to the 1-1 line. Additionally, that balance persisted when the same plot was regenerated over several sets of random seeds (not shown). There was evidence of skew, where the Q-Q plot line had a sigmoidal shape, with the skew in simulated age data related to an over- or under-representation of the age 35+ accumulator group. As with the dispersal patterns, the QQ plot's indication of skew in simulated data relative to the observed data appeared to be balanced over the years of data, and not systematic.

The Dirichlet process did not do as well when simulating trap fleet age composition data. Within the trap fleet, the performance for males is worse than for females. Indeed, there are several years where the simulated data were more dispersed than the observed data, and were sometimes very skewed given large plus groups in the observed data that were not produced by the Dirichlet process. Unlike the StRS, the dispersal and skew patterns were not evenly distributed across the 1-1 line, indicating that the problem was systematic. The poorer ability to reproduce the distribution of the data was driven by poor fits of the OM to some years of data, where the underlying operating model population did not have as many age-35+ fish as the data suggested. This was most prevalent in years where trap fleet age composition data were more skewed, with samples concentrated in the age 5-15 range with a large plus group (e.g., 1995, 1998, 1999, 2010, Appendix Figure 11). In those years, the plus group was often underestimated by the operating model (Johnson et al. 2025), which distributed the residual density among other age classes leading to a more dispersed population age-structure than the data suggested.

Finally, longline fishery age compositions looked realistically noisy. However, it was impossible to comment on the quality of those simulated data because there were no real historical age data for comparison (Appendix Figures 13, 14).

Simulations Testing the Effects of Sample Size and Sampling Design

We evaluated the statistical performance of the Sablefish operating model under alternative sample sizes and age sampling designs via closed loop simulation, where sampling design refers to the allocation of ageing effort among otoliths collected from different survey and fishing fleets. Simulations projected the Sablefish population forward in time for 21 years (2022 - 2042) under the current Sablefish management procedure (MP). The Sablefish MP is a specific, repeatable algorithm for computing a recommended annual catch limit that is composed of three components: data collection, an assessment method (a tuned Schaefer state-space production model (SSPM) to extract a signal of stock trend), and a harvest control rule that sets a recommended catch limit based on the SSPM trend (DFO 2023b). In our 21 year projections, all management parameters used within the MP were held constant at the current settings, i.e., the SSPM priors, operational control points and harvest rate were set to the values applied to the fishery in 2024-25 for the entire projection (DFO 2023b). In practice, regular evaluation and updating of the OM and MP occurs at 3-5 year intervals. This difference means that the operating model refits done in the analysis are for measuring statistical performance only. The influence of OM bias on management outcomes is not measured because we are not simulating realistic management responses over the 21 year time period.

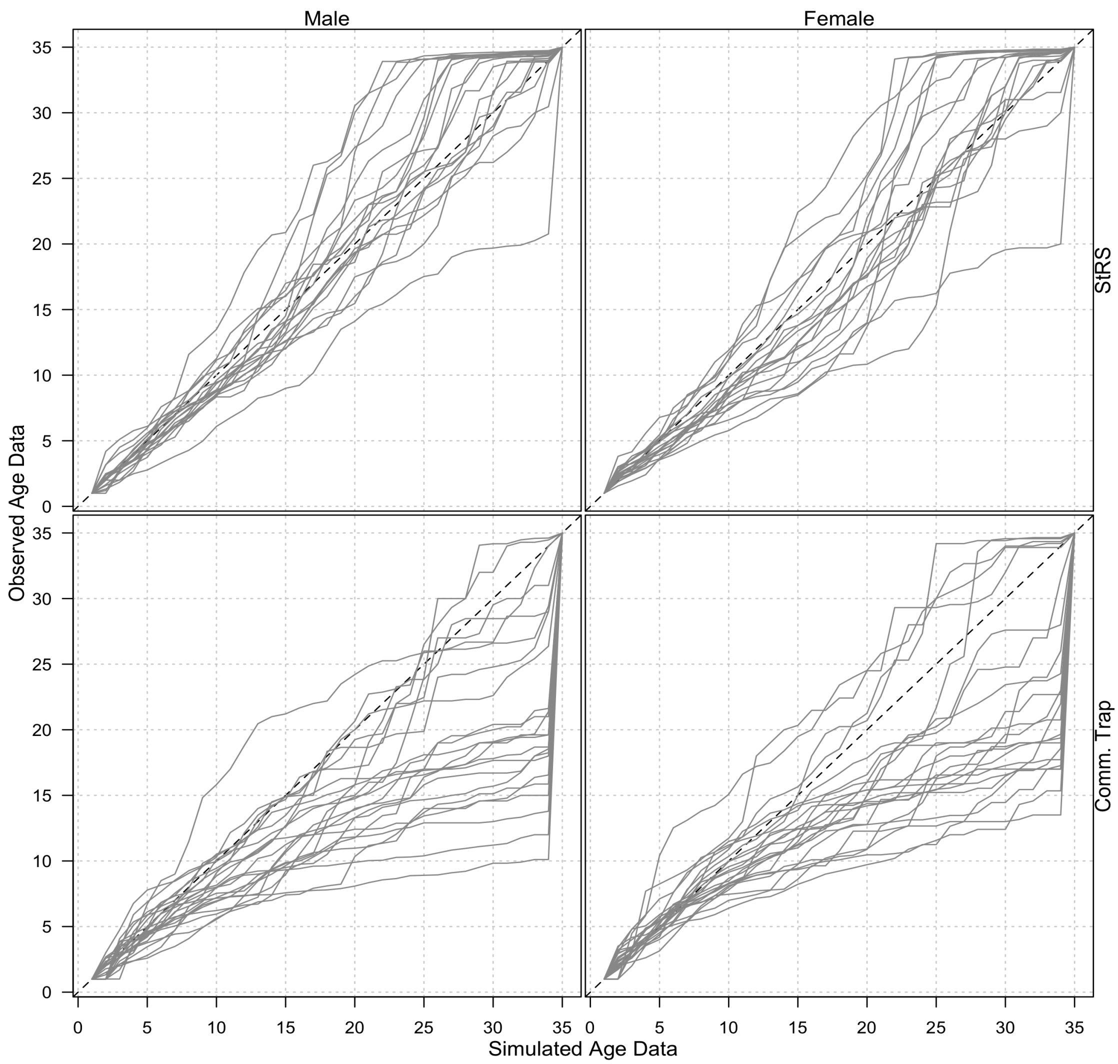

Figure 3. Q-Q plots showing the distributional difference between simulated and observed age compositions for the stratified random survey (top) and commercial trap (bottom) for male and female Sablefish age composition data. Each line represents a single historical year of data, and the closeness of the line to the diagonal reference line indicates the closeness of the cumulative distribution functions for simulated and observed data.

Simulated data in the forward projections included catch from the three commercial sectors, biomass indices from the StRS, and ageing data from selected fleets according to the simulated sampling design. The operating model was refitted to simulated data at 10 year intervals (i.e., years 11 and 21 of the projection). Each time the operating model was refit, estimates of operating model parameters and states were recorded to evaluate model performance over time, i.e., three sets of estimates including the initial OM fit in year 1.

All simulations were repeated for 100 replicates, which differed by the random seed used to draw random errors for simulated observation data. All scenarios had the same random seed for recruitment process errors, leading to the same annual relative year class strength across replicates for each year in the projection period. The same random seed for process error was used for all replicates to ensure comparability among simulation replicates. In addition, the common seed ensured that spawning biomass was below the LRP at the end of the projected time period in all replicates, which was required to test the effect of sampling designs on the ability of the OM to detect a conservation concern. Variability among replicates came from observation error alone. For scenarios where some replicates had non-convergence for refit OMs (characterised as a non-positive definite Hessian matrix), extra replicates were added until the 100 replicate threshold was reached. There were only a few cases where this occurred, described below.

Operating model statistical performance was measured as bias and variance in estimates of key quantities produced by the model refit in the final year of the projections (i.e., year 21, or 2042). Key quantities included selectivity parameter estimates, mortality and recruitment estimates averaged over the projection period, and stock status relative to the limit reference point. Mortality and recruitment averages are defined as

$$\bar{M} = \frac{1}{21} \sum_{t=2022}^{2042} M_t, \tag{3}$$

$$\bar{R} = \frac{1}{21} \sum_{t=2022}^{2042} R_t, \tag{4}$$

where M_t, R_t are the annual values of mortality and recruitment, respectively. The mean values $\bar{M}, \bar{R}$ are estimated for each of the 100 replicates, yielding simulated values $\bar{M}_i$ and $\bar{R}_i$, as well as the refit operating model, yielding estimated values $\widehat{\bar{M}}_i$ and $\hat{\bar{R}}_i$. Pre-2022, the operating model estimates a fixed natural mortality rate for each sex, with simulated operating model values $M_m = 0.052$ for males and $M_f = 0.093$ for females.

Bias and variance of key quantities were derived from the distributions of relative errors, defined as

$$RE_i(\theta) = \frac{\hat{\theta}_i - \theta_i}{\theta_i}, \tag{5}$$

where θ_i is a quantity of interest derived from the 'true' operating model state represented by the OM fit to historical data from 1965-2021, and $\hat{\theta}_i$ is the estimate of θ found by refitting the operating model to simulated data for replicate i. Some true OM values θ_i did not vary with i, such as model parameters B_0 or h, while others did vary with i, such as $\bar{M}_i$ and $\bar{R}_i$. Relative error calculation were repeated over all replicates i, and the resulting relative error distributions were summarised as bias (i.e., median relative errors, or $MREs$) and variance. The initial (first year) and intermediate (eleventh year) model refits were examined using retrospective plots of biomass trajectories to ensure that fits to simulated data were reasonable.

Simulation scenarios

The effects of sample size and age sampling design on bias and precision were tested against two recruitment scenarios and three scenarios about the distribution of fishery catch among trap and longline hook fleet (Table 1). Recruitment scenarios were used to examine how bias and precision changed when stock biomass declined quickly towards an LRP, while catch distribution scenarios were used to test how bias and precision changed when the proportion of fish taken by each fleet was adjusted, thereby influencing the removals of each age-class via fleet selectivity curves.

We defined the first recruitment scenario **baseR** by assuming that the stock-recruitment relationship estimated in the model history continued into the projections unmodified. The second recruitment scenario **0.1R0** represents a persistent recruitment failure beginning in the first projection year. This recruitment failure was simulated by scaling R_0 (or equivalently, the numerator of the Beverton-Holt stock-recruitment relationship) by 0.144. The scalar 0.1 was found by tuning the operating model spawning stock biomass to decline to the LRP of $0.4 \cdot B_{MSY}$ in 2042, thereby simulating a situation where a management decision point may be required (i.e., the need to develop a rebuilding plan).

Under the **0.1R0** scenario, three alternative catch distribution scenarios were also used to test the statistical performance of the OM under different sample sizes and sampling designs. These catch scenarios differed in the division of directed Sablefish catch between the trap and longline fisheries. While trawl fisheries are also allocated a portion (8.75%) of the directed Sablefish total allowable catch (TAC), we do not consider varying this allocation here as the trawl fleet is not part of the current analysis. The three catch distribution scenarios include a status-quo catch division (**SQuo**), where trap and longline fisheries each catch roughly half of the non-trawl directed Sablefish TAC, a 'trap-heavy' scenario (**TrShift**) where the trap fleet takes 70% of the non-trawl TAC starting in the first projection year, and a 'longline-heavy' scenario (**LLShift**), where longline takes 70% of the non-trawl TAC starting in the first projection year. The two extreme shifts to 70% trap and longline fisheries, respectively, are based on historical catch distributions observed for the BC Sablefish fishery.

Table 1. Simulation scenario factors and levels. The shift of catch to predominantly trap or longline fisheries is only applied to the 0.1R0 recruitment scenario.

Factor	Levels	Description	Label
Recruitment	R0	Beverton-Holt SR model with true R_0	baseR
	0.1R0	Beverton-Holt SR model with $0.144 \cdot R_0$	0.1R0
Catch distribution	Status Quo	Directed TAC based on recent average	SQuo
	Trap Heavy	70% of directed TAC taken by longline trap	TrShift
	Hook heavy	70% of directed TAC taken by longline hook	LLShift

The method for simulating age composition data with alternative sample sizes matched steps 1 - 3 above in the procedure for finding the concentration factor α_0^*. The abundance index data were simulated exactly the same as for MP evaluations, and catches were simulated by repeated application of the Sablefish MP for the entire time period (DFO 2023b; Johnson et al. 2025). The MP was not updated based on estimates from the simulated refits of the OM at years 11 and 21. Simulated longline fleet age composition data were generated using the trap fleet concentration factor as stated previously, given that no longline specific concentration factor can be derived from the current OM.

The experimental grid design, which controls the level of age sampling effort and allocations of ageing effort among fleets, differed by simulation scenario (Table 2).

For the **baseR** scenario, the range of total age sample sizes varied from 400 to 2000 ages, with 1600 considered a 'baseline' level based on recent sample sizes (Figure 1). The average number of observed Sablefish ages since the inception of the Sablefish MSE in 2011 is 1557, with a range from 1090 to 2481. Only a single sampling design was applied to the **baseR** scenario, which was based on the current target allocation of 60% of samples coming from the StRS and 40% coming from the trap fleet (labelled '60:40:0' in Table 2). While ageing of commercial samples would ideally be distributed between both trap and longline hook fleets, this allocation was precluded due to the paucity of longline hook samples. The target allocation of 60:40:0 has not been met for several recent years due to prioritizing StRS samples over trap fleet samples in response to declining ageing effort; however, we maintain it here as our baseline target.

For the **0.1R0** scenario, an extreme value approach was taken, where large (1600 or 2000) and small (600 or 800) sample sizes were tested for the target design of '60:40:0', as well as a design that includes some longline age composition data (Table 2). The extreme value approach determined if there is any significant effect of sample size on the quality of model estimates when the stock is rapidly declining under the **0.1R0** scenario, while at the same time reducing simulation run-time associated with completing a range of intermediate values.

Under the **0.1R0** scenario, when refitting the OM to simulated data there was an option to allow time-varying M (Table 2). This option was added after preliminary results revealed that the Sablefish OM with constant mortality rates (*conM*) was unable to accurately model the persistent recruitment failure under the **0.1R0** scenario when relying on recruitment process errors alone. Therefore, we added model flexibility via time-varying natural mortality for both sexes, denoted in the results by *tvM*, which improved the OM's ability to model the decline. Time-varying mortality is implemented as the following simple random walk

$$M_{x,t} = \begin{cases} M_x & t \leq 2021 \\ M_{x,t-1} e^{\epsilon_t} & t > 2021, \end{cases} \tag{6}$$

where $M_{x,t}$ are male ($x = m$) and female ($x = f$) instaneous natural mortality rates at time t, and ϵ_t is the log-normal random walk deviation. Note that both sexes have the same trajectory, and the random walk is only allowed to vary during the projection period.

There were 5 simulations under the **0.1R0** scenario that had non-convergent OM refits, where the number of replicates had to be extended to about 140 to get a significant number of samples. The conditions where the OM had the most trouble refitting to simulated data were the highest total age sample sizes (1600 or 2000 samples) and where the refit OM estimated a time-varying M in the projection period. It is not surprising that the OM had a hard time refitting under these conditions, as the *tvM* model was mis-specified, given that the cause of the decline was a recruitment failure, and the simulated age data is likely in conflict with that mis-specification.

Table 2. Simulated data and operating model M hypotheses used when re-fitting the Sablefish operating model in closed loop simulations. Experiments differ by scenario and are defined by the total age composition data sample sizes (N), allocation of samples (StRS, Trap, Longline), and M hypothesis used when the OM is refitted in years 11 and 21.

R Scenario	Sample Size (N)	StRS	Trap	Longline	Re-fit M hypoth.	Label
baseR	2000	60%	40%	0%	conM	2k_60:40:0
	1600	60%	40%	0%	conM	Baseline_1.6k_60:40:0
	1200	60%	40%	0%	conM	1.2k_60:40:0
	800	60%	40%	0%	conM	.8k_60:40:0
	400	60%	40%	0%	conM	.4k_60:40:0
0.1R0	2000	60%	40%	0%	tvM	2k_60:40:0_tvM
	600	60%	40%	0%	tvM	.6k_60:40:0_tvM
	2000	60%	40%	0%	conM	2k_60:40:0_conM
	600	60%	40%	0%	conM	.6k_60:40:0_conM
	1600	60%	20%	20%	tvM	1.6k_60:20:20_tvM
	800	60%	20%	20%	tvM	.8k_60:20:20_tvM
	1600	60%	20%	20%	conM	1.6k_60:20:20_conM
	800	60%	20%	20%	conM	.8k_60:20:20_conM

Effects of sample size on bias and precision - baseR

The Sablefish operating model reacts to changes in sample size for the **baseR** recruitment scenario and 60:40:0 sample design as expected. As sample size is increased, there is a reduction in bias and variance for the StRS selectivity parameters (Figure 4), while most other parameters had a less direct or weaker relationship with increased sample size. For StRS selectivity parameters, there appeared to be a diminishing return on age sampling effort as total sample size exceeded 1200 samples (combined for StRS and trap fleet male and female samples). This combined sample size was less than the 2011-2022 average of 1557 samples aged, but closer to the most recent 3-year average of 1275 samples aged. Above the 1200 sample threshold, the response of bias and variance for the length-at-100% selectivity (selA) was relatively flat, while the response of the selectivity standard deviation (selB), which controls the width of the dome-shaped selectivity curve, was still noticably changing, albeit at a declining rate.

In contrast, some parameters had increased bias and variance with increasing sample size, such as trap fleet selectivity standard deviation (Figure 4, selB_Trap) and terminal biomass (Figure 4, BT). This result was counter to expectations for a simulation experiment, where a correctly specified selectivity model should approach an unbiased estimate at higher sample sizes. There were two possible sources for this bias. First, as explained above, there were some distributional differences between simulated trap data and the observed trap data, possibly driven by fits with higher age sampling variance given the difference in the distribution of age composition data and the population vulnerable age structure. Recall that Figure 3 showed divergence in both dispersion and skewness of the simulated age distribution from the observed distribution. The observed sample sizes were in the range of 155 to 614, but most years were under 280. While such relatively low sample sizes may suggest that trap age composition data has low statistical power and may be unable to accurately detect the population age structure, the years with the largest skew/dispersion differences between observed and simulated data were those with large plus-group observations, which often had higher sample sizes from 200 to 319. Ultimately, issues related to sample size are one possible explanation for the difference between observed and simulated data, but alternative explanations include biased sampling, or immigration from neighbouring areas. The divergence between observed and simulated data was much less for StRS age compositions, which had roughly 5 times the total sample sizes overall but also did not include many of the years with large male plus group observations. Nevertheless, the difference between observed and simulated trap fishery age composition data was influencing the refit OM's estimates of selectivity, recruitment, and mortality parameters for years 11 and 21 of the projection. Therefore, as sample sizes increase above 800 for the trap fishery, new simulated data was able to outweigh the combined influence of informative priors and historical data, something that cannot be replicated by smaller sample sizes. As such, estimated parameters defaulted to the prior distribution, which was unbiased relative to the OM. The same phenomenon was likely the source of the increased bias in average recruitment during the projection years (Rbar) and terminal biomass (BT) in response to increasing sample size.

A second possible explanation for some parameters having increased bias with higher sample sizes could be the influence of the very large year class(es) from 2016-18. For example, selectivity could be ‘stretched’ as the model simultaneously attempts to fit to the future age composition data which had a second persistent mode passing through the population (e.g., Appendix Figure 11). This may have occured as the refit OM attempted to capture the large cohorts as they age, while also modeling incoming new cohorts. Such an effect would also depend on bias in recruitment and mortality estimates, as model parameters interacted non-linearly via the population process model. Additionally, OM refits would have updated historical fishing mortality based on the updated trap fishery selectivity curves. Updated selectivity then has follow-on effects on biomass and recruitment, generating similar biases in terminal biomas (BT) and projection period average recruitment (Rbar). Such patterns in the bias with respect to sample size may have been eliminated if a range of recruitment patterns that reflect both low and high recruitment series had been simulated, instead of a single future recruitment series.

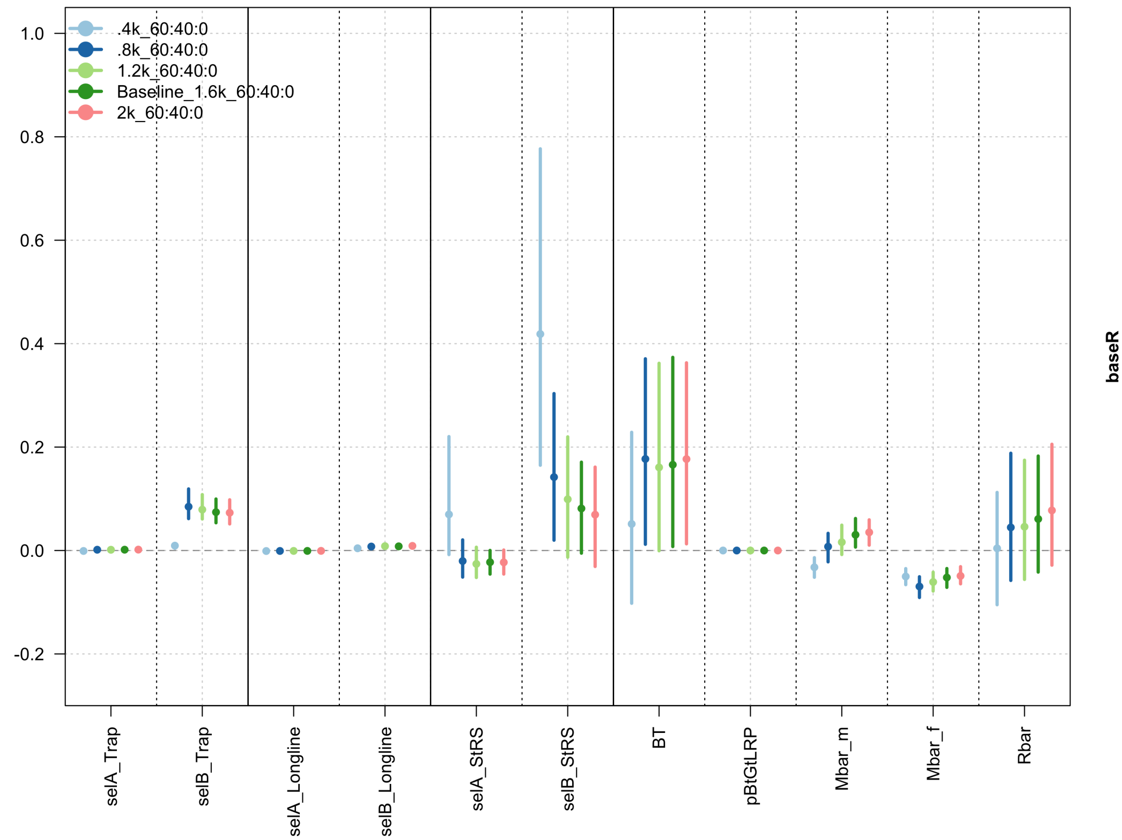

*Figure 4. Relative error distributions for selected OM quantities in the final projection year (2042) under the **baseR** scenarios with a baseline sampling design (60:40:0 for StRS:Trap:Longline). Quantitites include: selectivity mode or age-at-50% (selA) and selectivity standard deviation or age-at-95% (selB) parameters for commercial trap fishery, commercial longline fishery, and Stratified Random Survey (StRS); terminal biomass (BT), stock status measured as the probability of biomass in year T=21 exceeding the LRP of $0.4B_{MSY}$ (pBtGrLRP), mean M in projections (Mbar_f and Mbar_m for females and males, respectively), and mean recruitment in projections (Rbar). Line segments show the central 95% of each relative error distribution (variance) while points show the median relative error (bias).*

Detecting a persistent recruitment failure - 0.1R0

The **0.1R0** scenario was tuned so that recruitments were reduced by 90% in the projection period, thereby driving the MPD estimate of female spawning biomass to the LRP of 40% of B_{MSY} in 2042. For the purpose of those simulations and the relative errors derived from them, we assumed $P(B_{2042} \geq 0.4B_{MSY}) = 0.5$. This scenario was intended to evaluate whether age sampling designs could detect the change in stock status should the stock decline markedly for 20 years into the future.

Results indicated that age composition sample size was not the most significant factor when estimating whether the LRP had been breached due to persistent recruitment failure. In practice, there were two components of the system that are meant to detect and react to such a situation. First, the MP's SSPM trend should respond to a declining index with a corresponding reduction in harvest rates via the harvest control rule, promoting stock growth and protecting the stock from collapse. Second, frequent updates of the more complex OM at 3-5 year intervals acts as a back-stop to the MP by producing a revised characterization of stock status using the age-structured

OM and updated estimates of management parameters that were used for simulating stock and fleet monitoring data for MP performance evaluation. In the **0.1R0** scenario, neither of those components functioned as intended because the actual cycle of OM update and MP evaluation was not mimicked in our simulations.

The SSPM component does not detect the magnitude of the decline in spawning biomass for two reasons. As is common with retrospective analyses, there is a 1-2 year lag for the SSPM to react to new data points (Figure 5, lower panels). Part of the delayed reaction is that the spawning biomass is near the SSPM estimate of B_{MSY}, so the underlying deterministic logistic model estimates the production to be high. Therefore, while estimates do eventually react to the StRS index data trend, most of the decline in SSPM biomass is driven by process error estimates. During the last 1-2 years of each biomass time series, estimates often reverse the decline as the SSPM reverts to deterministic behaviour largely driven by highly informative priors placed on the SSPM parameters to obtain the desired MP performance. As a result, catches are not reduced quickly enough, so the median legal harvest rate exceeds U_{MSY} in 2036, and overfishing is sustained for 3 years before the fishery is abruptly shut down completely (Figure 6).

The OM was also challenged to detect the persistent low recruitment and resulting decline in biomass under both *conM* and *tvM* hypotheses. While our simulations did not follow the OM update schedule (every 3-5 years), the decadal refits at years 11 and 21 showed that both the *tvM* and *conM* models over-estimated the spawning biomass by about a factor of two at year 11 (Figure 5, top row), with the *tvM* model performing worse than the *conM* model. For the final year, both models did eventually detect a decline, but the terminal biomass was always overestimated by the *conM* model, while the *tvM* model was able to more closely estimate the true biomass in the example replicate shown (Figure 5, top row). There are a few reasons that the OM cannot fit to the simulated data. First, there is the statistical probability of estimating the scaled down recruitments. Specifically, the scalar on recruitment shifts absolute recruitment to the lower tail of the process error distribution. Therefore, a long run of large negative recruitment deviations are required to estimate recruitments with low bias. A second, somewhat related reason is that there is model mis-specification that occurs where (a) the stock-recruitment model is mis-scaled, or (b) where for the *tvM* model the mortality is aliasing a recruitment failure, and affecting all age classes instead of just age-1s. Finally, the overfishing driven by the MP may be biasing the productivity of the stock, which compounds these two sources of error, possibly leading to the high overestimate in year 11.

While the **0.1R0** scenario was of very low probability, it was not completely unprecedented that recruitment could fail for an extended period. In fact, Sablefish recently exited a period of below average recruitment with predominantly negative recruitment deviations during the 2000 - 2015 period. However, it should be noted that in practice the MSE process would likely have reacted much sooner than the simulated system, as there are human actors who would notice one of many indicators that the system is not functioning as intended. For example, in projections the simulated StRS survey produced indices near or below the minimum values observed in the 2010s, which would have signalled the need to intervene to diagnose the departure from expected performance if such values occurred in practice.

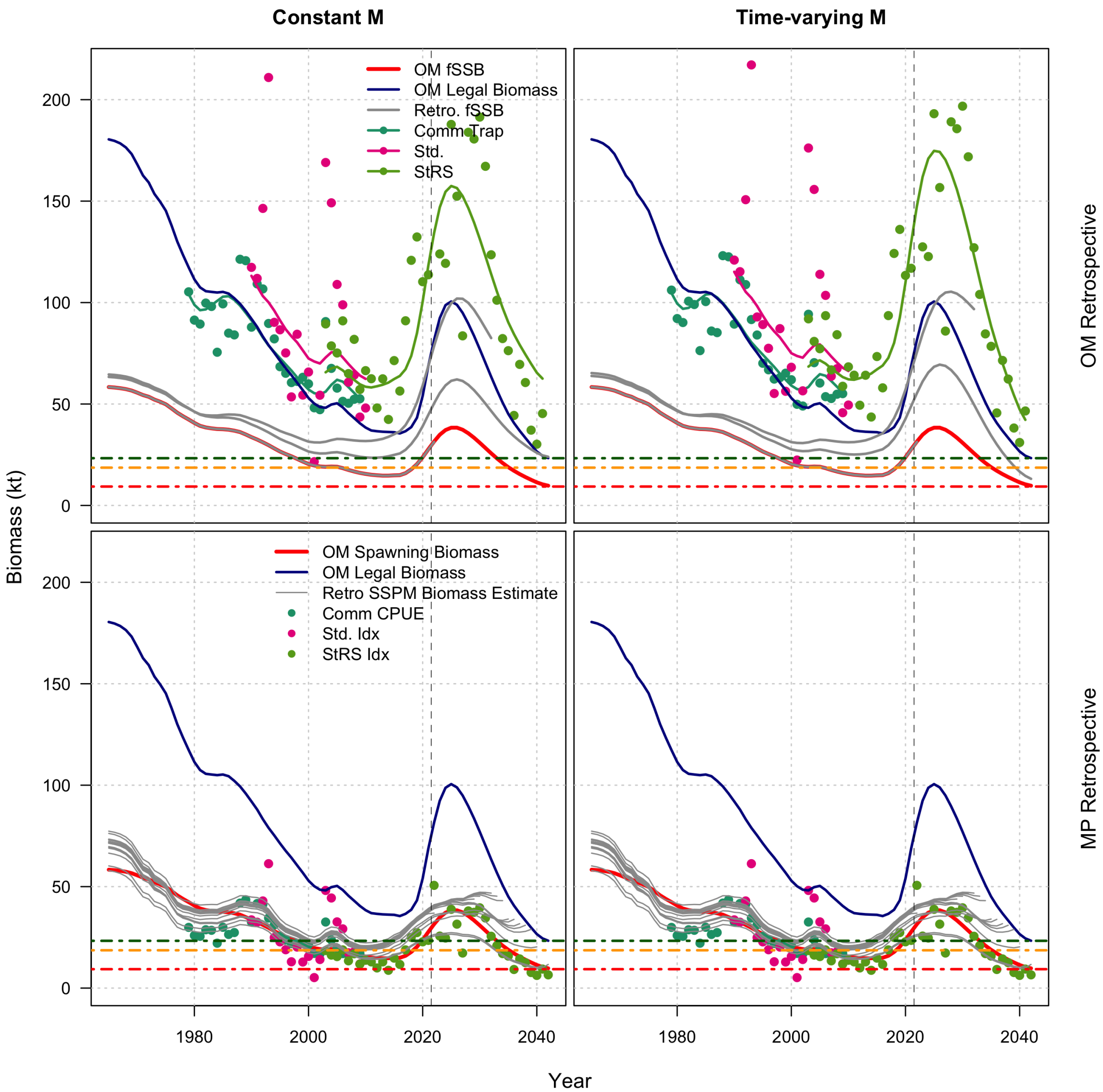

Figure 5. Retrospective estimates of operating model female spawning biomass (top row) and management procedure state-space surplus production model biomass (bottom row) when the 2k_60:40:0_conM (left) and 2k_60:40:0_tvM (right) experiments are applied to the 0.1R0_SQuo scenario. All panels show operating model female spawning biomass (fSSB, red) and legal sized biomass (blue), biomass index data scaled to the most recent model estimate of catchability (coloured points), and model-specific estimated of reference points including the LRP at 40% of B_{MSY} (red), 80% of B_{MSY} (yellow), and B_{MSY} (green). The grey lines either represent OM re-estimates of fSSB (top, 2021 estimate is perfectly aligned with OM fSSB) or MP estimates of biomass (bottom). Finally, the top panel also shows the final year's estimates of vulnerable biomass for each index as coloured lines that match the index points, to visually indicate goodness of fit (only shown for the year range of the corresponding index data).

Baseline sampling design: In the **0.1R0** scenario, increasing sample size reduced the variance in relative errors for the StRS selectivity standard deviation, but produced a non-significant increase on bias (Figure 7, selB_StRS). This effect was virtually identical for both the *conM* and *tvM* versions of the year 21 refit operating model. There was also a smaller effect on the StRS age at 50% selectivity parameter (selA_StRS), with higher sample size reducing bias and variance. The remaining selectivity parameters had less response to sample sizes, given that the tagging priors are very informative and quite closely aligned with parameter estimates given the extensive time series of trap and StRS age composition data.

Age sample size had less of an influence on the OM estimates of terminal biomass (B_T) and stock status (measured as $P(B_t \geq 0.4B_{MSY})$, and abbreviated as pBtGrLRP). Instead, the largest effect was observed between *tvM* and *conM* OMs. The *conM* models consistently over-estimated biomass at the end of the projection by at least 90% (Figure 7, BT), even if they eventually detected a decline (Figure 5). Additionally, *conM* models always over-estimated stock status, incorrectly assigning $P(B_t \geq 0.4B_{MSY}) = 1$ in over 50% of replicates. The *tvM* operating models still over-estimated biomass on average, but under the **SQuo** and **LLShift** catch distribution scenarios, there were some replicates where biomass was correctly estimated or under-estimated. Despite this, it was still very difficult for the operating models to correctly estimate stock status, with most distributional mass concentrated near a relative error of 1.0 (i.e., estimating $P(B_t \geq 0.4B_{MSY}) = 1$).

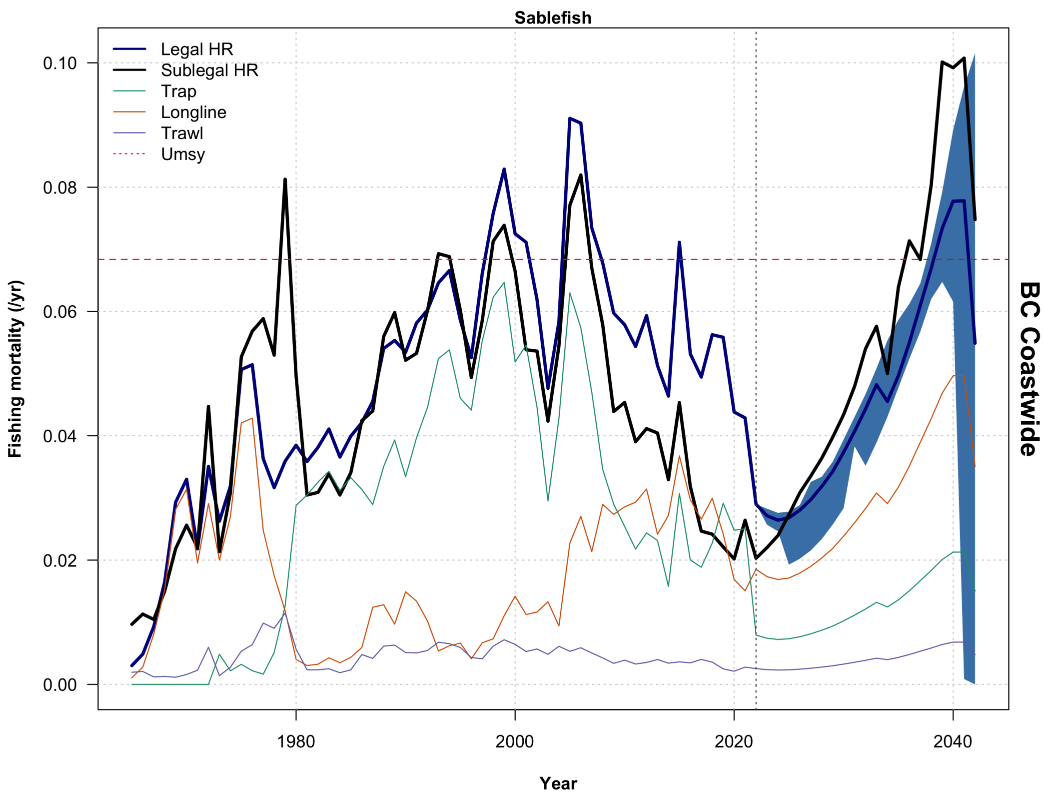

Figure 6. Simulation envelopes of total yearly harvest rates under the 0.1R0 scenario and the .8k_60:20:20_conM sampling design and M setting. Here the MP fails to curtail fishing due to over-estimation of biomass and the stock crashes due to rapidly increasing fishing mortality.

Finally, there were some effects of sample size and M estimation on average mortality and recruitment estimates in the projection period. For mortality, increasing sample size led to increases in relative error variance and had only a small effect on bias (Figure 7, Mbar_m and Mbar_f). The effect was most noticeable under the *SQuo* scenario, and fairly subtle for the scenarios that shift to trap- or longline-dominated catch distributions. The effect of sample size on time-varying mortality rate estimates was not surprising, as age data was the primary source of natural mortality rate information. The effect of catch distribution on the variance of relative errors in average mortality over the projections was not related to sample size for any fleet, as the fleet-specific age compositions were modeled independent of total fleet catch. Instead, the effect of TAC allocation on relative error variance may have been related to differences in size selectivity between the trap and longline fleets, where longline had almost zero selectivity for sub-legal fish, leading to less confounding between natural and fishing mortality for lower age classes. For estimates of average recruitment over the projection period, the OM overestimates by at least 100% of the true value for all sample sizes and M estimation settings (Figure 7, Rbar). While the lowest bias and lowest 2.5th percentiles both occur under the highest sample sizes (2k), bias and variance is slightly lower under the *tvM* models for both tested sample sizes. As expected, bias and variance in average recruitment is negatively correlated with variance (and to a lesser extent, bias) in the average mortality rates, given that recruitment and mortality have opposite effects on the population.

Additional longline sampling: Adding longline samples to the simulated age composition data set under the '60:20:20' sample design induced some changes in relative error bias and precision, but the results were qualitatively very similar to the baseline '60:40:0' sampling design. As before, the most intuitive effect of sample size was the reduction in variance for StRS selectivity parameters.

One major difference from the baseline sampling design was that terminal biomass and stock status was better estimated when longline data was included. Terminal biomass relative error distributions had greater mass below zero, especially for the **SQuo** catch distribution scenarios (Figure 8, BT), with lower bias as well. Interestingly, under the **TrShift** fishery scenario, increased sample size led to higher bias for terminal biomass, which may be related to the bias in longline selectivity parameters, as well as an effect of nearly unbiased mean natural mortality rates under the *1.6k_60:20:20_tvM* case. Stock status relative errors also had a larger proportion of negative relative errors (Figure 8, pBtGtLRP), consistent with better estimates of terminal biomass. Estimates of mean mortality and recruitment in the projection period had the same pattern as above. Higher sample sizes led to lower bias in R estimates on average, and time-varying M increased variance in M estimates on average while also reducing bias in R estimates on average. Overall, allocating more heavily to one fleet or another reduced variance relative to **SQuo**.

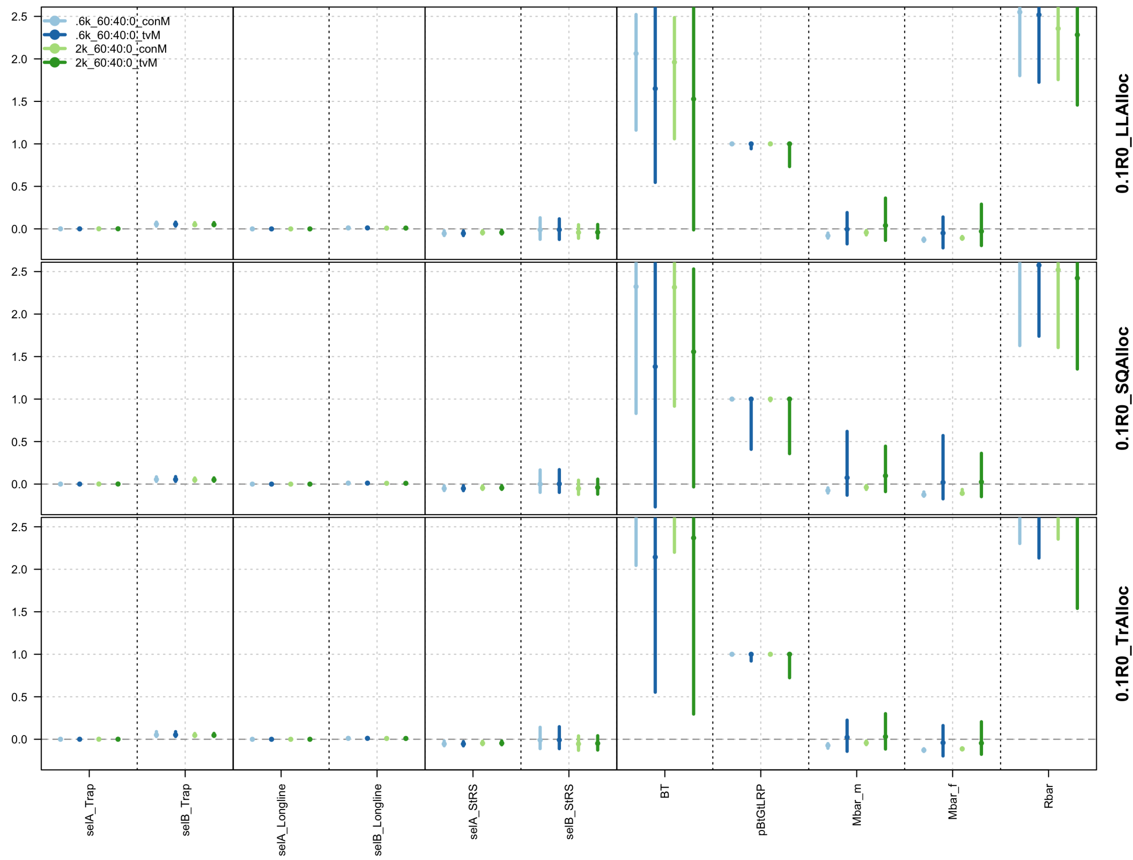

*Figure 7. Relative error distributions for selected OM quantities in the final projection year (2042) under the **0.1R0** scenarios with a baseline sampling design. Each panel shows relative errors for a different directed catch allocation during the projection period. Quantitites include: selectivity mode or age-at-50% (selA) and selectivity standard deviation or age-at-95% (selB) parameters for commercial trap, commercial tongline, and Stratified Random Survey (StRS); terminal biomass (BT), stock status (pBtGrLRP), mean M in projections (Mbar_f and Mbar_m), and mean recruitment in projections (Rbar). Line segments show the central 95% of each relative error distribution (variance) while points show the median relative error (bias).*

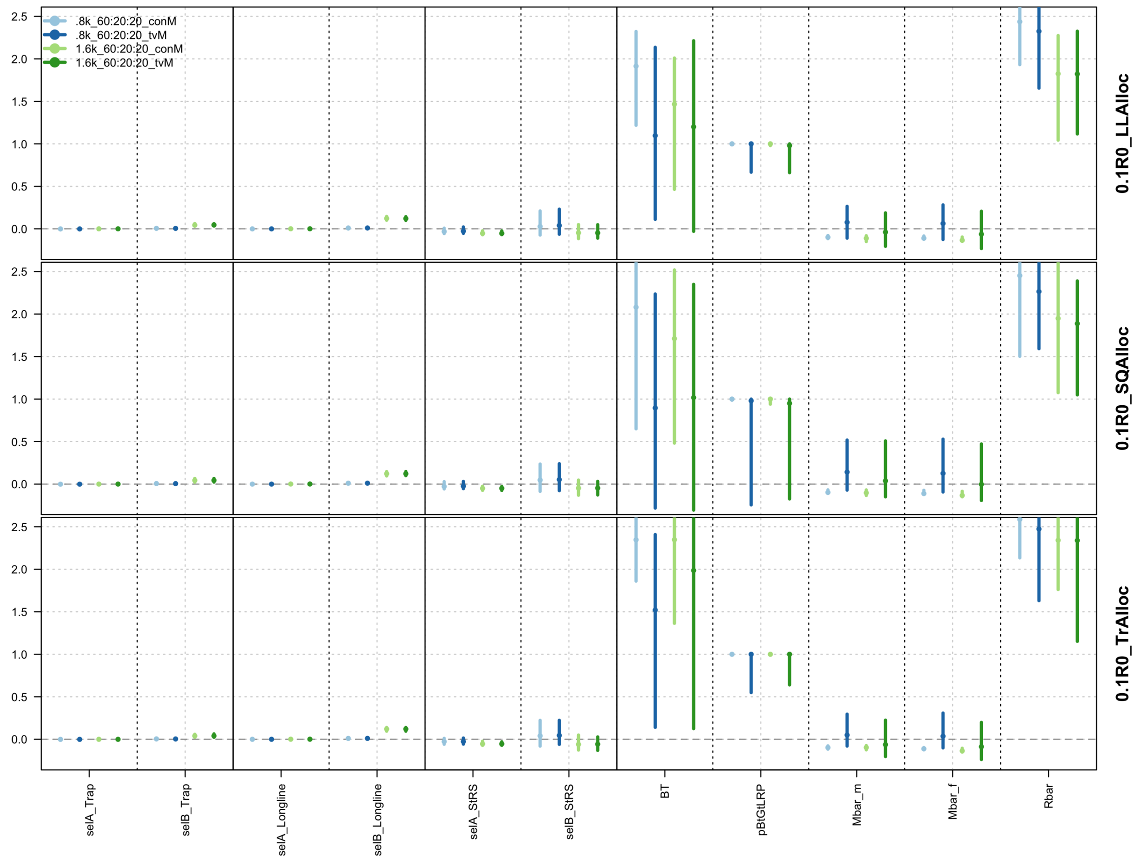

*Figure 8. Relative error distributions for selected OM quantities in the final projection year (2042) under the **0.1R0** scenarios with longline ages added to the sampling design. Each panel shows relative errors for a different directed catch allocation during the projection period. Quantitites include: selectivity mode or age-at-50% (selA) and selectivity standard deviation or age-at-95% (selB) parameters for commercial trap, commercial longline, and Stratified Random Survey (StRS); terminal biomass (BT), stock status (pBtGrLRP), mean M in projections (Mbar_f and Mbar_m), and mean recruitment in projections (Rbar). Line segments show the central 95% of each relative error distribution (variance) while points show the median relative error (bias).*

CONCLUSIONS

In this Science Response, we explored the effects of total age sample size and age sampling design on the statistical performance of the Sablefish operating model over a 20-year projection. Utilizing a model based on a Dirichlet process, we examined the relationship between age composition residual standard error and sample size. The results of this study show a beneficial impact of increased annual age composition sample size on model bias and precision across two recruitment scenarios: one with consistent ‘average recruitment’ and another simulating a substantial, long-term recruitment reduction by about 85%.

Recommendations

Sample sizes in the range of 1200 - 1600 ages / year are sufficient to support the current Sablefish OM

Bias and precision for StRS selectivity parameters did not significantly improve above 1600 total samples from the StRS and Trap fishery, and deteriorated quickly when total sample size drops below 1200 aged fish. Beyond this level, increases in sample size (i.e., 2000 samples) did not yield significant improvements, indicating diminishing returns on uncertainty reduction with higher sampling efforts. Such diminishing returns were consistent with the shape of the residual age sampling variance curve with respect to sample size, estimated from the Dirichlet process based on operating model outputs. While it may seem narrow to focus on the StRS selectivity, observations from the StRS provide the strongest signal of stock trends and year class strength, given that they come from a designed random survey. Additionally, sample sizes less than 800 had insufficient statistical power to overcome the trap selectivity priors, and there was still a large reduction in StRS selectivity standard deviation bias up to 1200 and 1600. Moreover, biases in other parameters are within expectations for age structured models. For example, bias in average recruitment (or unfished rectruitment) is common when simulating age data, which provides information on relative year class strength and not absolute recruitment levels, which must be inferred indirectly via the model fits to other data sources.

Age composition data from longline fisheries should be used to fit the OM

Current biomass and stock status were better estimated when longline data were included. We found that model performance improved by dividing the current ‘target’ 40% sampling effort for the trap fleet equally between the two directed gear types (trap and longline), subject to our assumption that longline fleet samples have a similar Dirichlet process concentration factor to the trap fleet. When longline samples were included, relative error distributions for stock status (pBtGrLRP) and average age-1 recruitment estimates (Rbar) had lower bias and also included zero more often, indicating better coverage of the true value. Factors contributing to improved performance include the reduced model mis-specification from simulated trap age sampling (i.e., less impact of being unable to capture large residuals in the true data) and potentially better age-sampling fidelity of older fish in the longline fishery, given its very low selectivity of sub-legal Sablefish. Based on these results, we recommend that increasing the number of otoliths collected from the longline fishery continues to be a high priority.

The current MP may need to be revised to be more responsive to recruitment failure

While increasing age sample size and expanding the sampling design does not appear to be a sufficient action, on its own, to ensure detection of a persistent recruitment failure, this result should not be interpreted as a reason to drop age sample sizes to half of the baseline level. There were significant improvements in bias and variance for some parameter values under increased sample sizes for the **0.1R0** scenario. Additionally, while no simulated age sample size eliminated the bias in stock status and terminal biomass completely for the **0.1R0** scenario, the situation simulated was quite an extreme case which would likely not occur in practice, and so eliminating the bias further may be impossible without modifying the operating model's structural hypotheses to more closely match the simulated recruitment dynamics. Indeed, given the record high recruitment for the 2016-2018 year classes, an 85% reduction in recruitment strength for 20 years was required to reduce spawning stock to the limit reference point in 2042, a situation that would likely present itself to fishery managers, analysts, and harvesters in multiple ways. For example, signals of recruitment failure may manifest as a decline in the trap survey abundance index or changes to age- or length-composition data. In addition, harvester observations of catch rates and average fish size on the water would supplement the scientific and biological data, and based on historical behaviour those observations would be corroborated by the dynamics of the US stocks in Alaska to the north, and the US west coast to the south. Indices towards the end of the projection were similar to the indices at the stock's lowest points in the mid-2010s, which in practice would be detected first by harvesters and analysts during annual evaluation of survey outcomes, and when updating the OM.

Sablefish biomass within the MP was estimated from a state-space surplus production model, which has the benefit of being simple to understand and apply every year, as well as being fairly slow to react to increases in stock indices. However, the reaction lag that provides a conservation benefit when the stock trend is increasing can also cause slow reactivity when the survey trend is decreasing. The production model's behaviour was related to its symmetric production function, the shape of which was tightly tuned with informative priors. As described in the results, these design elements contributed to a lag in detecting a large decline in biomass to the LRP over a 20 year period. Possible changes to improve the reaction time of the MP include adopting a delay-difference or simple age-structured model which would add some skew to the production function. Alternatively, stock indices could be re-derived to be based on size classes (Regular et al. 2017), enabling faster reaction to changes in size composition. These changes would increase MP complexity, and for an age-structured model or size-based survey indices could also increase data-demand over the production model. However, such increased complexity may be the cost of improved risk mitigation should the stock actually experience persistent low recruitment in future.

Future OMs should consider using scenarios about time-varying processes to test MPs

Given the OM's lag in detecting the recruitment failure, it is advisable to test time-varying quantities in future operating model updates. Specifically, time-varying selectivity, mortality, and recruitment model parameters are fairly easy to implement as alternative structural hypotheses by adding a time-series of random effects, such as we did here. Those random effects tend to stay close to zero in the event that the data have little evidence for time-variation (Stewart and Monnahan 2017), and add fairly little overhead to the model development and fitting cycle.

Precision readings collected by the DFO Schlerochronology Lab should continue to be evaluated to monitor for changes in reader-dependent effects that would affect our recommendations

Our analyses assumed that bias in ageing errors is constant over time (DFO 2020). Time-dependent changes in both age-reading bias and precision were not considered, but could occur as existing Sablefish age-readers gain experience or new age-readers are trained. We recommend that the existing age-correction matrix be updated if possible, and that precision readings collected by the DFO Schlerochronology Lab be evaluated for reader-dependent effects.

Limitations

The Dirichlet process demonstrated a better fit to average residual error from model fits to age compositions rather than annual residual error. This is most obvious in trap fishery data, which may itself be influenced by poor sampling designs in early years and the large residuals in the plus group described above. As such, year-to-year variation in residual error exceeded what the Dirichlet process could simulate, suggesting that further refinement of the Dirichlet process, possibly at a more granular level such as trip or set data, could yield more accurate results.

There are no data on the relationship between longline fishery age composition sample size and age-sampling residual error. Therefore, the operating model's statistical performance with respect to longline age composition samples should be considered as guidance on a relative benefit of including longline samples, and not as a recommendation on absolute sample sizes. When adding longline fishery samples in the future it may be advisable to vary sample sizes higher and lower over time to elicit the relationship between sample size and sampling error faster. Such analysis may result in the revision of guidance for total age sample size presented here as well as the distribution of ages among data sources.

Finally, it is important to acknowledge the inherent limitations of simulated data, which can sometimes deviate significantly from real-world observations. This discrepancy was particularly evident in our simulations of trap fishery data and its impact on selectivity parameter estimates at higher sample sizes. Such challenges highlight the need for continual refinement of simulation techniques and model assumptions to enhance the reliability and applicability of our findings in practical fishery management.

LIST OF MEETING PARTICIPANTS

Name	Affiliation
Kendra Holt	DFO Science, Pacific Region
Samuel Johnson	Landmark Fisheries Research
Sean Cox	Landmark Fisheries Research
Rob Kronlund	Interface Fisheries Consulting
Dana Haggarty	DFO Science, Pacific Region (steering committee member)
Audrey Ty	DFO Science, Pacific Region (steering committee member)
Darah Gibson	DFO Fisheries Management, Pacific Region (steering committee member)
Andrew Edwards	DFO Science, Pacific Region (reviewer)

SOURCES OF INFORMATION

Chilton, D.E., and Beamish, R.J. 1982. Age determination methods for fishes studied by the groundfish program at the Pacific Biological Station. Can. Spec. Publ. Fish. Aquat. Sci 60: 102. Department of Fisheries; Oceans Ottawa.

Cox, S., Kronlund, A., and Lacko, L. 2011. Management procedures for the multi-gear Sablefish (*Anoplopoma fimbria*) fishery in British Columbia, Canada. DFO Can. Sci. Advis. Secret. Res. Doc 2011/063: viii + 45 p.

Cox, S.P., Kronlund, A.R., Lacko, L., and Jones, M. 2023. A revised operating model for Sablefish in British Columbia, Canada in 2016. DFO Can. Sci. Advis. Sec. Res. Doc. 2023/023. vii + 127 p.

DFO. 2020. Evaluating the robustness of candidate management procedures in the BC Sablefsh (*Anoplopoma Fimbria*) Fshery for 2019-2020. DFO Can. Sci. Advis Sec. Sci. Resp. 2020/025.

DFO. 2023a. A revised operating model for Sablefish in British Columbia in 2022. DFO Can. Sci. Advis. Sec. Sci. Advis. Rep. 2023/010.

DFO. 2023b. Application of the British Columbia Sablefish management procedure for the 2023-24 fishing year. DFO Can. Sci. Advis Sec. Sci. Resp. 2023/009.

Francis, R.C. 2014. Replacing the multinomial in stock assessment models: A first step. Fisheries Research 151: 70–84. Elsevier.

Hanselman, D.H., Clark, W.G., Heifetz, J., and Anderl, D.M. 2012. Statistical distribution of age readings of known-age sablefish (Anoplopoma fimbria). Fisheries Research 131-133: 1–8.

Johnson, S.D.N., Cox, S.P., Holt, K.R., Lacko, L.C., and Rooper, C.N. 2025. Stock Status and Management Procedure Performance for the BC Sablefish (*Anoplopoma fimbria*) Fishery for 2022/23. DFO Can. Sci. Advis. Sec. Res. Doc. 2024/072. iv+132 p.

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.

Lacko, L.C., Acheson, S.M., and Holt, K.R. 2023a. Summary of the annual 2021 Sablefish (*Anoplopoma fimbria*) trap survey, October 6 - November 21, 2021. Can. Tech. Rep. Fish. Aquat. Sci. 3530: vii + 48 p.

Lacko, L.C., Temple, K.L., Holt, K.R., Supernault, J.K., Kronlund, A.R., Wyeth, M.R., and Connors, B.M. 2023b. Development of methods in support of a head-only sampling program for Sablefish (*Anoplopoma fimbria*) in British Columbia. Can. Tech. Rep. Fish. Aquat. Sci. 3580: vi+19p.

Regular, P., Cadigan, N., Morgan, M., and Healey, B. 2017. A simple SAM-style state-space stock assessment model for greenland halibut in NAFO subarea 2 and divisions 3KLMNO. NAFO Scientific Council Research Document 17(010).

Schnute, J.T., and Haigh, R. 2007. Compositional analysis of catch curve data, with an application to sebastes maliger. ICES Journal of Marine Science: Journal du Conseil 64(2): 218–233.

Stewart, I.J., and Monnahan, C.C. 2017. Implications of process error in selectivity for approaches to weighting compositional data in fisheries stock assessments. Fisheries Research 192: 126–134.

Thorson, J.T., Johnson, K.F., Methot, R.D., and Taylor, I.G. 2017. Model-based estimates of effective sample size in stock assessment models using the dirichlet-multinomial distribution. Fisheries Research 192: 84–93.

APPENDIX

Otolith Sampling History

Table 3. Number of otolith samples collected from field sampling each year, by fleet.

Year	StRS Survey	Trap Fishery	Longline Fishery	Trawl Fishery
2003	5,836	1,348	248	115
2004	5,670	923	513	213
2005	5,671	70	176	336
2006	6,397	181	0	221
2007	5,943	426	0	183
2008	6,096	666	60	60
2009	5,837	997	65	129
2010	6,057	1,036	63	115
2011	4,894	1,272	42	0
2012	4,699	1,483	89	0
2013	3,950	817	24	217
2014	3,445	632	0	426
2015	3,632	833	21	44
2016	3,774	637	0	140
2017	3,675	681	0	756
2018	4,464	423	0	653
2019	4,330	579	182	355
2020	3,587	884	84	6
2021	3,548	685	0	0
2022	4,287	770	197	0

Table 4. Number of otolith samples with available age estimates completed, by year and fleet.

Year	StRS Survey	Trap Fishery	Longline Fishery	Trawl Fishery
2003	1,125	0	0	0
2004	1,061	564	0	0
2005	1,146	0	0	0
2006	1,168	0	0	0
2007	1,305	0	0	0
2008	1,401	0	0	0
2009	1,465	437	65	0
2010	984	552	62	0
2011	1,114	383	42	0
2012	1,210	355	68	0
2013	1,015	704	24	84
2014	1,744	544	0	193
2015	922	802	0	0
2016	1,260	516	0	67
2017	1,165	158	0	160
2018	1,160	0	0	0
2019	1,167	0	0	0
2020	1,152	337	45	0
2021	1,200	0	0	0
2022	919	0	171	0

Simulated Age Composition Data

The following figures compare available yearly age composition data for year prior to 2022 and simulated data for an example simulation replicate in the projection period. Yearly sample sizes used for the projection period were based on a total simulated sample size of 1600 across all fishery and survey sources, to be most in line with the 'Baseline' total sample size assumed for Sablefish age reading.

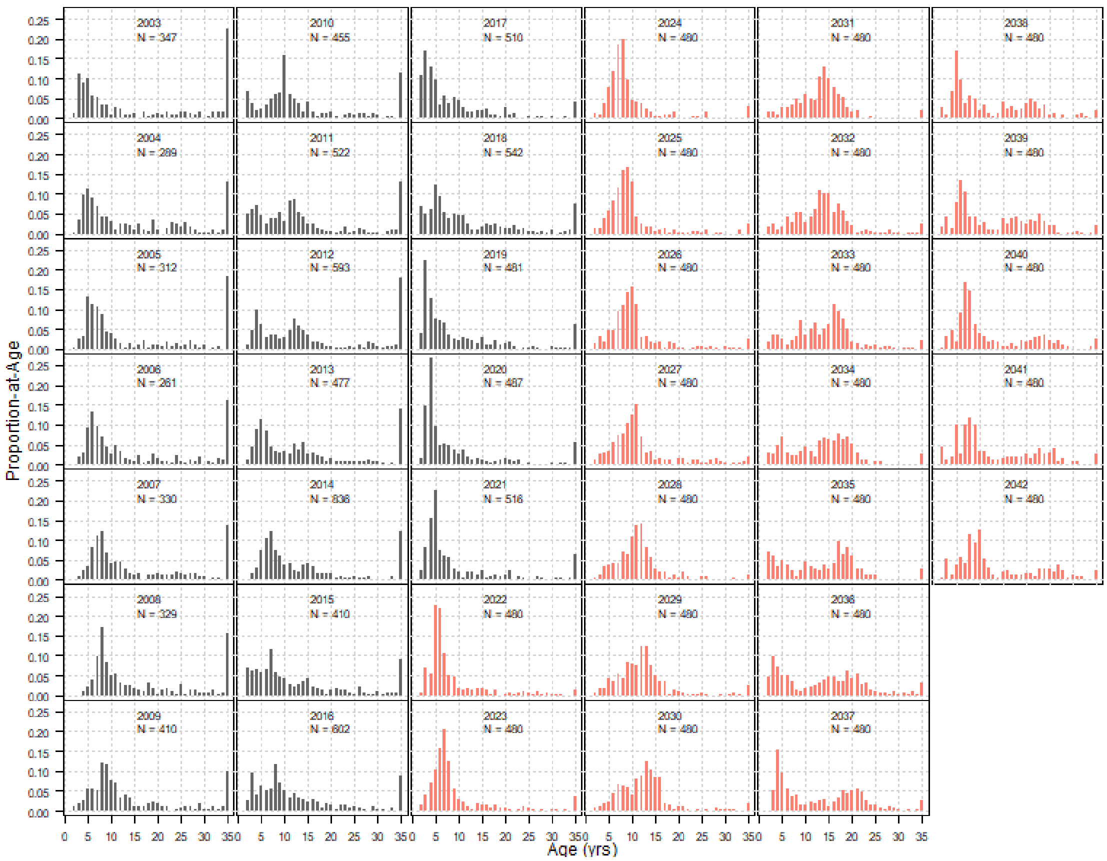

Figure 9. Yearly StRS male age composition data for the historical period (black) and simulated data for an example simulation replicate in the projection period (pink).

Figure 10. Yearly StRS female age composition data for the historical period (black) and simulated data for an example simulation replicate in the projection period (pink).

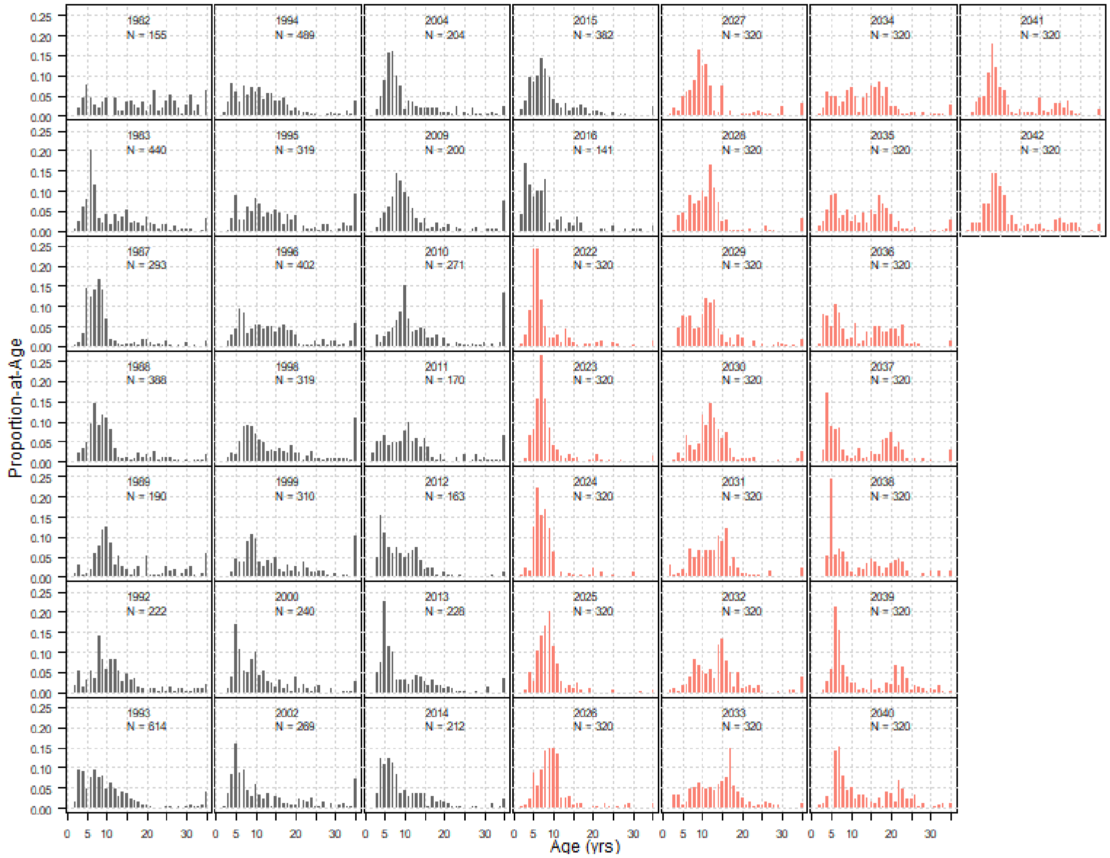

Figure 11. Yearly Trap fishery male age composition data for the historical period (black) and simulated data for an example simulation replicate in the projection period (pink).

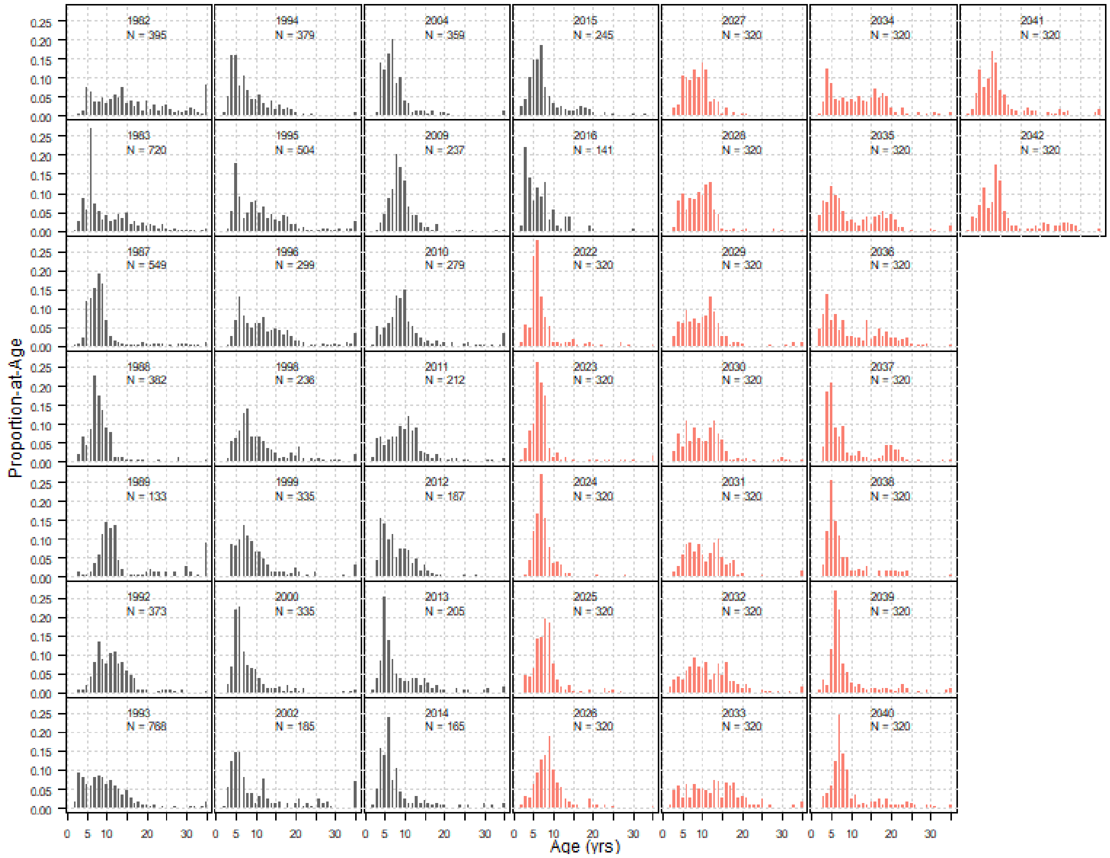

Figure 12. Yearly Trap fishery female age composition data for the historical period (black) and simulated data for an example simulation replicate in the projection period (pink).

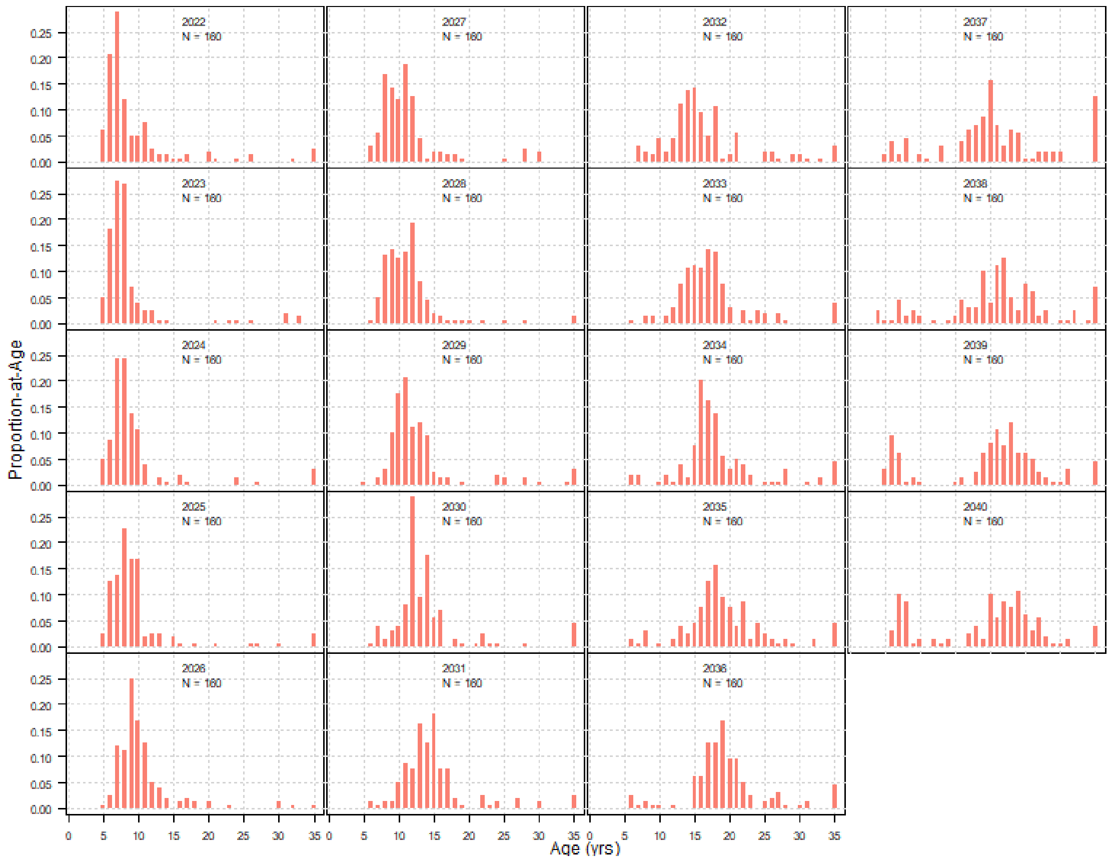

Figure 13. Yearly longline fishery male simulated age composition data in the projection period (pink). There are no historical longline fishery age composition data included in the operating model.

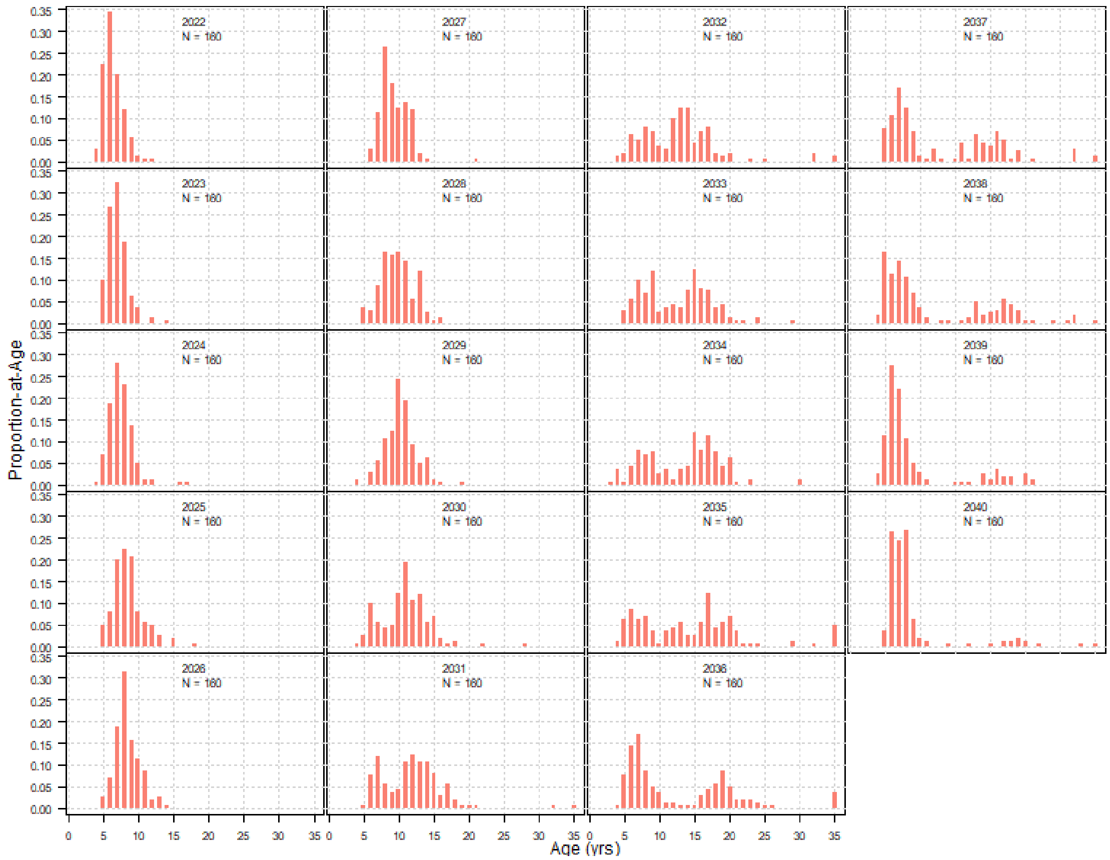

Figure 14. Yearly longline fishery female simulated age composition data in the projection period (pink).There are no historical longline fishery age composition data included in the operating model.

THIS REPORT IS AVAILABLE FROM THE:

Centre for Science Advice (CSA)
Pacific Region
Fisheries and Oceans Canada
3190 Hammond Bay Rd.
Nanaimo, BC V9T 6N7

E-mail: DFO.PacificCSA-CASPacifique.MPO@dfo-mpo.gc.ca
Internet address: www.dfo-mpo.gc.ca/csas-sccs/

ISSN 1919-3769
ISBN 978-0-660-76762-8 Cat No. Fs70-7/2025-013E-PDF

Correct Citation for this Publication:

DFO. 2026. An Investigation of Ageing Requirements to Support the British Columbia Sablefish (*Anoplopoma fimbria*) Operating Model. DFO Can. Sci. Advis. Sec. Sci. Resp. 2025/013.

Aussi disponible en français :

*MPO. 2026. Étude des exigences relatives à la détermination de l’âge à l’appui du modèle opérationnel pour la morue charbonnière (*Anoplopoma fimbria*) en Colombie-Britannique. Secr. can. des avis sci. du MPO. Rép. des Sci. 2025/013.*