Explorations of alternative stock assessment models for Eastern Bering Sea Pacific cod

Authors: Steven J. Barbeaux, Pete Hulson, Ingrid Spies, and James Thorson

September 11, 2023

1 Introduction

For 2023 the authors wished to examine outstanding problems common to all four of the eastern Bering Sea (EBS) Pacific cod ensemble models accepted for management in 2022. The two main issues with the ensemble models were: 1) for the length composition data the Dirichlet multinomial log(theta) values approach the upper bound and therefore needed to be fixed for the models to converge, 2) failing residual runs tests for length and age composition data in all ensembles indicating autocorrelation in the residuals pointing at poor residual behavior, 3) potential confounding of aging bias, annually varying growth, and annually varying selectivity result in the models being highly unstable with considerable tuning of the annual devs. on growth and selectivity required for model convergence, and 4) the models are highly sensitive to changes in catchability and natural mortality with small changes in either resulting in substantial changes in management advice with only small changes in negative log likelihood. For Model 22.2 there was a ~150,000 ton range in 2023 ABC with a less than 10 -LL change from the MLE in likelihood profile run over catchability.

For 2023 we conducted a series of model explorations in an attempt to fix these issues. After initial investigations to move to a more standardized way of setting input sample sizes it became apparent that a different approach was needed for model exploration as the updated model did not consistently converge on the MLE and the complex base model made investigations difficult. We chose to move to a more simplified model and then add components to the model sequentially to evaluate the impacts of different model assumptions. All models examined in this paper are built in Stock Synthesis version 3.30.21 and parameterized the same as Model 22.2 (Barbeaux et al. 2022) except for changes specified. For the simplified model we reverted to a standard multinomial for the composition data and removed all time varying components, we also fixed aging bias based on previous model results and in line with isotope analyses validating aging methods (Kastelle et al. 2017). A full description of changes made to simplify the model is provided below. For model explorations we examined a wide range of model alternatives however for this analysis we have limited the changes to the impacts of allowing growth to be time varying, allowing survey selectivity to be time varying, reducing the maximum age from 20 to 12, adding catch data from 1964 to 1976 while removing the regime change parameter on recruitment, and adding conditional age-at-length. Although all the models were generally well fit, the results in terms of both stock size and management implications from these models remained highly variable and individual models remained sensitive with small changes in fit resulting in large changes in management advice. The models explored displayed notable variation in survey catchability spanning a range from 0.74 to 1.10 and was highly negatively correlated (R^2 = -0.92) with natural mortality ranging from 0.33 to 0.44. Likelihood profiles on catchability showed that for many of the seemingly reasonable individual models ABC recommendations could vary by more than 100,000 t with changes in log likelihood of less than 2 points. A major finding therefore of this work is that for Bering Sea Pacific cod very disparate outcomes in terms of management advice could be generated from models with very little difference in fit to data and retrospective bias.

For 2023 the authors recommend the following:

1) Moving away from the ensemble approach currently employed. The authors believe that the current ensemble of models are too similar in nature. A better ensemble approach would be to include models with more varied structures such as the multispecies model (CEATTLE) and simplified random effects models. This effort would require a much larger team of researchers to evaluate individual model performance. As it stands the evaluation of individual model performance in the ensemble may have been hindered by the volume of work required of a single author to complete an ensemble of models given their unstable nature in the limited time between the September and November Plan Teams. In addition it is the opinion of the assessment authors that if the ensemble approach is to continue the SSC should consider setting up a working group outside the normal Plan Team process to develop a science-based and transparent model selection and weighting scheme for new ensembles.
2) Changing the current modeling approach to a simpler model with fewer and/or more constrained, annually varying parameters on growth and selectivity as these parameters are confounded. Generally we add this type of complexity to deal with poor residual behavior and retrospective bias issues, here the residuals of the complex models don't appear to differ substantially from the more complex models and retrospective patterns are within reasonable bounds even with the simplest models considered.
3) Fixing one or more key parameters (e.g. M, Q, L_{min}, L_{max}, K, etc.) in the model or using more constrained priors would provide improved model stability. Constraining natural mortality or growth parameters would be better candidates than survey catchability as more refined aging methods developed recently should better inform these parameters. Either fixing or providing a strong prior on natural mortality would have the added effect of constraining catchability within a model as these parameters are highly correlated.

If a single model is to be used for management this year, of the models presented the authors would recommend Model 23.1.0.d be considered as an option for further exploration for management in 2024. Model 23.1.0.d is the simplified model with the addition of constrained annual variability in growth (Figure 1 and Figure 2) and survey selectivity (Figure 3 and Figure 4). The model employs bootstrapped input sample sizes, the fit to the survey index is improved over Model 22.2 (Figure 5), retrospective bias remains within acceptable limits, and residual runs analyses imply that the fit to the composition data are better with less autocorrelation in the residuals while maintaining similar effective sample sizes. The mean absolute scaled error (MASE) analyses show the predictive skill of Model 23.1.0.d for the survey index was improved over Model 22.2 (0.40 vs 0.69) and remains similar for the fishery mean length and survey mean age.

Model 23.1.0.d has improved performance while reducing the model by 86 parameters, however this reduction in complexity comes at a cost of increased uncertainty (higher standard deviations) in some key parameter estimates (M, Q, and R_0; Figure 6) which translates into increased uncertainty in derived quantities such as B_0, $F_{40\%}$, current recruitment (Figure 7), current spawning biomass (Figure 8 and Figure 9), and future catch recommendations. Another potential red-flag in Model 23.1.0.d is that natural mortality at 0.429 is higher than most methods external to the model indicate and catchability at

0.765 is lower than most other models examined previously. Profiles over catchability show little change in likelihood over a wide range of natural mortality and catchability suggesting little information in the data to inform these quantities (Figure 10).

One solution may be found in applying a maximum age-based method for deriving a prior for natural mortality (Sulliven at al. 2022; Thorson et al. 2023), which suggests a lower value with M at 0.387. When natural mortality is fixed in Model 23.1.0.d to this value, catchability increases to 0.972 and may be a reasonable alternative to allowing natural mortality to be freely fit. Fixing M in this model has a minor cost of degrading the overall model performance by +1.4 negative log likelihood (-LL) with an improvement to the fit to the age composition of -3.95 -LL, but a poorer fit to the survey index at +2.56 LL and length composition at +2.40 -LL.

2 Model 22.2 updated changes in input sample size

Hulson et al. (2023) found that there was not a consistent approach to setting input sample sizes for composition data in assessment models at the Alaska Fisheries Science Center. They proposed a unifying bootstrap approach that would evaluate the variance and autocorrelation within the survey composition data collections to appropriately calculate annual input sample sizes. For the 2022 Pacific cod ensemble models the input sample sizes for the survey size and age composition data were set at the number of surveyed hauls for each year, and the fishery size composition data were set at the number of hauls sampled standardized to the mean sample of hauls from the survey over all years surveyed which is unique to Eastern Bering Sea Pacific cod. This method led to an average input sample size of 369 for both survey and fishery length and age compositions (Table 1). As noted the ensemble models were fit using the Dirichlet-multinomial (DM), which as coded in Stock Synthesis uses a parameter (log theta) to re-weight the data and in effect reduce the size and age composition input sample sizes, if needed, to appropriately weight data components within the model. In all of the 2022 ensemble models the DM log theta parameters approaches the upper bound for both the fishery and survey size composition data and must be fixed in order for the models to converge. Having the DM theta parameter fixed at the upper bound is not optimal and may indicate that input sample sizes for the size composition data are too small and therefore underweighted in the model compared to other data components. Using a bootstrap approach (Hulson et al. 2023) for calculating input sample size for the survey length and age composition data resulted in an on average smaller age composition sample size of 250 (in agreement with the fitted DM theta value of -0.47) and a much larger on average input sample size of for the size composition data of 1661 (Table 1). A bootstrap approach is not yet available for the fishery composition data and therefore for the fishery size composition data input sample size we used the annual number of hauls sampled standardized to the mean survey size composition input sample size so that both means were equal for the two size composition data sets. As in previous years it was assumed that the raw numbers of hauls were far too high as they numbered in the tens of thousands for some year, far higher than the survey input sample size. Model 22.2 was then fit using these new sample sizes. Although not useful for judging differences in model fits the overall negative log likelihood (-LL) increased from 10,875 to 18,362, with a sharp increase in the length composition -LL from 9,990 to 17,383 which does indicate a shift in the model weighting with more weight being given to the composition data. The DM log theta values for the survey age composition data changed from -0.47 to -0.60 and the survey size composition changed from near the bound at 10 in the old model to 1.32 in the updated model. However the fishery size composition DM log theta remained at the bound suggesting the sample sizes remained too low in comparison with the other data components or potentially an additional issue with model misspecification. The change resulted in substantially more weight on the composition data than in the old Model 22.2 and a degradation in fit to the survey (Table 2 and Table 3) from -6 -LL to 68 -LL. Retrospective and mean absolute scaled error (MASE) values were the same between the two model configurations (Table 4) and the fishery mean length residuals remained significantly autocorrelated as determined by the residual runs test (ss3diags; Winker et al.2023). However, the change in input samples sizes resulted in better residual behavior for the survey mean length and age (Table 5). Convergence was impacted with a large number of jitter runs failing to converge at the MLE for the updated model. The profile over catchability for the updated Model 22.2 (Figure 1) shows a highly irregular profile resulting from the models not converging to the MLE for each of the fixed catchability value runs. In conducting likelihood profiles over catchability for both the old and updated Model 22.2 survey catchability (Q) and natural mortality (M) are strongly negatively correlated (R^2 = -0.999) with a slightly higher M (Table 10) in the model with updated input sample size. The updated Model 22.2 growth parameters were different from the old Model 22.2 driven entirely by the different input sample sizes (Table 7) and changes in relative weighting of the data components. However, changes in growth parameters had little impact on the overall size at age (Figure 1). Additional changes in influential parameters are shown in Table 8. Despite changes in important parameters such as catchability and natural mortality both model configurations resulted in similar estimates of recruitment over time (Figure 7), spawning stock biomass over time (Figure 8 and Figure 9), reference points (within 5%; Table 9), and management advice (2024 ABC within 2%; Table 9).

We explored changing from the DM to standard multinomial and implementing the Francis TA1.8 weighting method for Model 22.2 to improve model performance. However, we found that when iteratively fitting using Francis TA1.8 methods the model the suggested correction to the fishery size composition value continued to increase without settling until the model no longer converged. This suggests model misspecification in Model 22.2, potentially due to the same issue contributing to the DM log theta approaching the upper bound for the fishery length composition data despite rather large input samples sizes. We theorized that the model misspecification could be due to confounding among the freely fit aging bias, annually varying growth, and annually varying selectivity.

3 Model 23.1.0.a description

To allow an easier understanding of the interaction of model components on model results and model sensitivities we created a simplified version of Model 22.2 (Barbeaux et al. 2022). The new simplified Model 23.1.0.A had the following changes:

1. Removing length composition data for years with age composition data (1994-2021) which were duplicated in the age comps.
2. Reconfiguring both survey and fishery selectivity to be static instead of including annually varying parameters.
3. Reconfiguring the Richard’s growth to be static instead of including annually varying L_{min}.
4. Reconfiguring the survey double normal selectivity (Stock Synthesis pattern 24; Methot et al. 2023) to estimate parameters 1-4 and using new asymptotic option for parameter 6.
5. Fixing the pre-2007 aging bias to Model 22.2 values.
6. For the growth model fixing CV at older ages at 0.06 and fixing CV at younger ages at 0.2 based on the previous ensemble model fits.
7. Changing from the Dirichlet-multinomial to standard multinomial for length and age composition data.
8. Using the iterative Francis TA1.8 weighting method to tune the model.

3.1 Duplicate composition data

For the 2022 ensemble models both survey and age composition are included for all years in which they are available, resulting in 1994-2021 having both survey age and survey size composition data included in the model. Therefore the survey composition data for these years are potentially more highly weighted in the models than other data components. In the exploratory models for the years with bottom trawl survey age composition data available (1994-2019, and 2021) the bottom trawl survey size composition data were removed.

3.2 Selectivity and growth

For the 2022 ensemble models both fishery and survey selectivity and L_{min} in the Richard's growth model were set to be annually varying. This is likely confounded as the model would likely not be able to discern between annually varying growth and selectivity at the smaller sizes potentially leading to issues with model convergence and inability to settle on appropriate Francis weighting for the fishery length composition data. This may have also led to issues with the DM log theta approaching the upper bound for these data. Whether the annual variability is attributed to growth or selectivity has impacts on model results affecting management advice. For the simplified model, Model 23.1.0a, we set both growth (Figure 2) and selectivity (Figure 3) to be static over time reducing the number of devs by 220 (Table 2).

In addition, we implemented a selectivity feature new to stock synthesis for the survey that simplifies the double normal function (selectivity option 24 in Stock Synthesis; Method et al. 2023) where the values past a set length are meant to be static, here we set all selectivity values at lengths greater than 40 cm to be fixed. It should be noted that although this feature was meant to fix the survey selectivity to be asymptotic and not allow the shape to become dome-shaped, this was found not to be the case when conducting profiles over catchability. In cases where catchability was high and natural mortality was very low, the shape of the survey selectivity curve did unexpectedly become dome-shaped.

3.3 Aging bias

Aging bias was fit for all of the 2022 ensemble models as a two parameter linear vector from ages 2 to 20. The two parameters are the aging bias at age 2 and at the max age. These two parameters tended to vary considerably depending on assumptions of growth and selectivity as model configurations were explored leading to some fits that were improbable suggesting that these parameters were likely confounded with growth and selectivity. Changes in estimated aging bias had substantial impacts on model results and some fits were well outside what would be expected given isotope analysis (Kastelle et al. 2017). For the models explored this year in order to stabilize model explorations we fixed the two parameters based on the 2022 Model 22.2 accepted values.

3.4 Composition distribution from Dirichlet multinomial to standard multinomial

As described above when fitting Model 22.2 with the updated bootstrap input sample sizes the fishery size composition DM log theta continued to approach the upper bound, which in effect reverts the distribution to the standard multinomial. Although it has been common practice to fix the log(theta) parameter near the upper bound when this occurs, the fit may indicate that the input sample sizes continue to be inadequate or some other model misspecification.

In order to investigate this phenomenon and evaluate other options, we changed the presumed distribution of the composition data to the standard multinomial. We then iteratively adjusted the model as per the Francis reweighting scheme TA1.8 (Francis, 2011) as implemented in the R library r4ss (Taylor et al. 2021), a technique previously utilized in Pacific cod models prior to 2018.

3.5 Model 23.1.0.a Results

Model 23.1.0.a performed well overall with standard metrics for model fits similar and in some cases improved over the more complicated Model 22.2 with substantially fewer fit parameters (82 vs. 306). Iterative Francis reweighting for Model 23.1.0.a settled on consistent values and resulted in considerable down-weighting of the length and age composition data from the initial input sample sizes with a multiplier values of 0.03 and 0.06 for the fishery and survey length composition data and 0.25 for the survey age composition data. This resulted in the survey index having more influence on the model then in the 2022 ensemble models. This is evidenced with the improved likelihood and RMSE on the survey abundance index fit (Table 3 and Table 4) and shown as a much tighter fit of the model to the survey abundance index (Figure 5). Due to the change to the Francis weighting versus the DM between the two models, no direct comparison of likelihoods can be used for comparison, however looking at the effective sample size shows a drop in the effective sample size between the updated Model 22.2 and Model 23.1.0.a for the fishery size and survey size and age compositions (Table 4). The visual fits to the fishery length composition data reveal little difference in the fitted values (Figure 11) or residuals (Figure 12). However, Model 22.2 fits the survey size composition better, specifically Model 23.1.0.a tends to overestimate large incoming small fish even more than Model 22.2 when large recruitments are present (Figure 11). Visual inspection of the age composition fits (Figure 13) show Model 23.1.0.a fitting the age data less closely than Model 22.2 with lower effective same size, however visual inspection of the age composition Pearson residuals from both models shows similar patterns (Figure 14).

The retrospective (Table 5) and residual runs tests (Table 6) results were similar between updated Model 22.2 and the simplified Model 23.1.0.a with Mohn's rho retrospective values on spawning bias at 0.07 and passing runs test for all but the fishery size composition data component for both models. Mohn's rho tests show a small positive bias (0.08) for Model 23.1.0.a while the updated Model 22.2 had a slight negative bias (-0.06). Examination using the mean absolute scaled error (MASE) provided in the ss3diags R Package (Winker et al. 2023) showed a marked improvement in the prediction skill of Model 23.1.0.a over the updated Model 22.2 for the survey index (Table 5), a slight improvement for the prediction skill of mean survey age, and a slight degradation for the prediction skill of the mean fishery length.

Despite being different from the updated Model 22.2 the growth parameter estimates between the old Model 22.2 and Model 23.1.0.a are similar (Table 6), however the standard deviation of the parameter estimates for L_{max} and the Richard's parameter are nearly double in the simpler model. Similar fitted values are expected given the down-weighting of the composition samples through the Francis re-weighting method and lower sample sizes in the old Model 22.2 compared to the updated model. The increase in the variance of these parameters in Model 23.1.0.a over Model 22.2 was due to removal of annual variability in growth and selectivity where some of the variability was attributed to the annual devs.

Model 23.1.0.a results in an increase in the estimated catchability over the update Model 22.2 to 1.097 from 0.974. This has the impact of scaling down recruitment (Figure 7) and spawning stock biomass overall (Figure 9). For Model 23.1.0.a both the jitter analysis and likelihood profile over catchability were well behaved with all jitter runs converging and the majority arriving at the MLE. Although the likelihood profile over survey catchability for this model is well behaved it shows very little change in the overall likelihood over a wide range of survey catchability values (Figure 10 and Table 11). For catchability ranging from 0.9 to 1.28 there is a change in negative log likelihood of less than 2 -LL from the maximum likelihood estimate (MLE) resulting in a 100,000 t difference in 2023 recommended ABC across that range. This is of similar concern in Model 22.2. This may be an issue with environmentally driven fluctuations in growth and recruitment contributing substantially to the overall biomass's variability, leading to limited insights into the consequences of fishery removals. The ability of a model to fit catchability is influenced by the degree to which catch impacts changes in survey abundance. Given the considerable impact of environmental drivers on cod abundance and mortality, there's a possibility of insufficient data for accurately determining survey catchability without a better understanding of the environmental drivers of this stock.

4 Further model explorations

Alternative models with increasing complexity were developed to further explore model sensitivity and performance with the following:

1. Allowing annual variability in both the L_{min} and Richards K parameters.
2. Adding constrained annually varying selectivity for the survey.
3. Integrating catch data from the period 1964-1976 into the model while eliminating the recruitment adjustment parameter linked to the 1977 environmental regime change.
4. Adjusting the upper age group for model dynamics from age 20 to age 12 to more accurately represent available data and speed up model runs.
5. Introducing survey conditional age-at-length (CAAL) data.

Model	Npar. +Ndevs	Annually varying growth	Annually varying survey selectivity	Max age to 12	Catch to 1964 no regime	CAAL
23.1.0.a	82					
23.1.0.b	176	X				
23.1.0.d	218	X	x			
23.1.0.g	217	X	x	x	x	
23.1.0.h	217	X	x	x	x	x

4.1 Model 23.1.0.b

4.1.1 Annually varying growth

For the 2022 ensemble models all growth was fit as a 4 parameter Richard's growth relationship with L_{min} fit as an annually varying deviation. All parameters in the 2022 models were fit with an uninformative prior. For the simplified Model 23.1.0.a although the four parameters were fit within the model with uninformative priors both growth and selectivity were set to be time-invariant. It has been long understood that environment, particularly temperature, is influential in the growth of Gadus species (Taylor 1958) and annual variability in growth should be expected. Growth in Pacific cod specifically has been found to be rather elastic and dependent on environmental conditions particularly for young fish (Laurel et al. 2008, Barbeaux et al. 2021). To evaluate this elasticity we explored including annually varying growth in Model 23.1.0.b. Here we used the growth parameters and standard deviations as posteriors from Model 23.1.0.a (Table 7) as 'priors' for all four growth parameters and added a mean tending random walk for L_{min} and Richard's parameter. This is option 5 described in the Stock Synthesis manual (Methot et al 2023) as a mean reverting random walk with rho and a logit transformation to stay within the minimum and maximum parameter bound. In developing Model 23.1.0.b allowing all the growth parameters to vary annually was evaluated, however only varying L_{min} and the Richards parameter provided substantial improvements to the model fit. The authors understand that using the values from Model 23.1.0a are not true priors, but given their large standard deviations it was considered only as a means of providing soft bounds on the parameters and kept the parameters from wandering to extreme values during fitting. The standard deviation of the devs were tuned following the methods of Thompson et al. (2021) where the standard deviation was tuned to set the variance of the estimates plus the sum of the estimates' variances equal to unity. This resulted in standard deviation of 0.4416 for L_{min} and 0.2995 for the Richards parameter annual deviations.

4.1.2 Results of adding annually varying growth

Model 23.1.0.b performed well overall with standard metrics for model fits improved over Model 23.1.0.a. The results of allowing annually varying growth (Figure 2) was an overall improvement to fits to all data components over the static growth model with lower negative log likelihood across all data components, lower RMSE for the survey index, and higher effective N for all of the composition data (Table 3, Table 4, and Figure 5). As was expected with the use of the 'prior' and annual variability the standard deviations of the growth parameters were also reduced (Table 7) from Model 23.1.0.a. Retrospective bias remained similar to Model 23.1.0.a with a small positive bias for spawning stock biomass (Table 5). The MASE evaluation showed improved predictive skill on the index and size composition and a slight degradation in the predictive skill on age composition (Table 5). Residual runs tests resulted in an improvement for the fit to the survey index and both survey and fishery length composition fits, but a slight degradation in the survey age composition fit from Model 23.1.0.a. However, unlike all the previous models described, all data components passed the residuals runs test (Table 6). Although fits to the length composition (Figure 15) and age composition (Figure 13) were discernably better than Model 23.1.0.a, patterns in the residuals for both length (Figure 15) and age (Figure 14) composition were visually similar.

While adding annually varying growth improved model fit, it also increased uncertainty in model estimates of management reference points (Table 8 and Table 9), annual estimates of recruitment (Figure 6 and Figure 7) and annual estimates of spawning biomass (Figure 8 and Figure 9). There was an overall increase in uncertainty in model results compared to Model 23.1.0.a and Model 22.2. The estimate of bottom trawl survey catchability was lower and natural mortality higher in Model 23.1.0.b than in Model 23.1.0.a (Table 8). However, the likelihood profile on catchability for Model 23.1.0.b continues to show low certainty in this value as large changes in catchability continue to result in small changes in likelihood (Table 11 and Figure 10).

It is interesting to note that in likelihood profiles over catchability, natural mortality is always highly negatively correlated with catchability, however in the new models growth and selectivity parameters appear to have tipping points where the trajectory of the parameter values with catchability change suddenly (Figure 16). When examining the likelihood profiles by model component this manifests as a change in the trajectory of the index data likelihood (Figure 10). This tipping point becomes even more pronounced when annually varying growth is introduced with priors on the main growth parameters. In the 23.1.0.x series models, lower natural mortality, correlated with higher catchability values above 1.0, are compensated for by forcing the selectivity curve to be increasingly dome-shaped (Figure 17) making a subtle trade-off of fit between the survey age and size composition data and the survey index. Specifically, this trade-off is made for the initial year 1982, 1994, and 2001 where the survey abundance appeared to be higher than expected (Figure 5), out of line with adjacent years, and poorly fit in all models. This tipping point is due to the parameterization of the survey selectivity in which parameter 6, selectivity at last bin, is set to -1040 which was meant set selectivity constant for bins greater than bin number 40, but appears to allow selectivity become dome-shaped in some cases.

An increase in catchability and decrease in natural mortality translated into a scaling up of the population with higher estimates of annual recruitment (Figure 7) and annual spawning biomass (Figure 9). Although unfished spawning biomass (B_0) increased over Model 23.1.0.a (Table 2) it did not increase proportionally to the annual spawning biomass and remained well below that estimated in Model 22.2. This resulted in an overall increase in the estimated status of the stock to over $B_{50\%}$ in 2023 compared to Model 23.1.0.a and Model 22.2, where it was estimated to be below $B_{40\%}$.

4.2 Model 23.1.0.d

4.2.1 Annually varying survey selectivity

In all of the 2022 ensemble models both survey and fishery selectivity was modeled as annually varying. This variability was removed for the 2023 simplified model but reinstitution of annually varying selectivity on survey selectivity was examined (Figure 4). The general parameterization of selectivity remained the same with a six parameter double normal as described for the Model 23.1.0.a, however an annual additive deviation (Stock Synthesis option 2; Methot et al. 2023) was added to the ascending width of the curve with a standard error of the deviation restricted to 0.2, this added 42 dev. parameters to the model, one for each survey year. Although the apparent variation in the fitted curves was minor (Figure 4), the additional flexibility improved the model fit by -10.19 -LL. We had examined models with annually varying selectivity in the fishery as well, however adding annually varying fishery selectivity made only a modest improvement in model fit over a Model 23.1.0.b with a substantial increase in the number of annual dev. parameters (141 additional dev parameters gaining a 4 point reduction in -LL). For brevity the model with fishery annually varying selectivity was dropped from consideration and not fully presented in this document.

4.2.2 Results of adding annually varying survey selectivity

The addition of restricted annually varying selectivity for the bottom trawl survey improved the overall fit compared to Model 23.1.0.b (-10.19 -LL). The largest changes were an improved fit to the survey abundance index (-5.49 -LL), and both length composition data sets (-10.17 -LL) and a degraded fit to the survey age composition data (+2.21 -LL) (Table 2, Table 3, and Table 4). Retrospective Mohn's rho and MASE values were the same as Model 23.1.0.b (Table 5). All modeled components passed the residual runs tests (Table 6) with the same p-values for the index and fishery length composition and slightly lower p-values for the survey length and age composition (Table 6). There was little difference in the growth parameter estimates with nearly identical standard deviations (Table 7). There was an increase in the estimated R_0 and natural mortality and a decrease in the survey abundance index catchability with standard deviation remaining near those of Model 23.1.0.b (Table 8). Similar to all of the models examined the log likelihood profile over catchability again showed the model to have low certainty in this value as large changes in catchability continue to result in small changes in likelihood (Table 12 and Figure 10).

The increase in catchability and decrease in natural mortality once again translated into a scaling up of the population from Model 23.1.0.b with higher estimates of annual recruitment (Figure 7) and annual spawning biomass (Table 2, Table 9, and Figure 9). Although this model is very near in fit to Model 23.1.0.b the management advice provided would have increased the projected 2024 ABC by approximately 24 thousand tons.

4.3 Model 23.10.0.g

4.3.1Catch data 1964-1976

Regarding the ensemble models for 2022, the catch series begins in 1977, and there is a presumed shift in the climate regime in the same year, positively influencing recruitment of Pacific cod from that point onward. The steady-state catch level for all four ensemble models was set at 42,500 tons, reflecting the average catch from 1964 to 1976. Examining the catch data available for the years 1964 to 1976 (Table 12), it's evident that the catch fell notably below this average prior to 1967. Despite the considerable catch recorded during the 1920s and 1930s, anecdotal evidence suggests that the catch levels from the 1940s to 1967 were lower compared to that of 1968 onward (Mackovjak, 2019). To test the influence of this initial catch on management reference points we added the 1964-1976 data and set an equilibrium catch value to a low 10,000 t, an approximation of potential catch levels prior to the development of the modern fishery in the late 1960's.

The notion of a regime change in the North Pacific, leading to altered recruitment patterns in groundfish, was initially proposed by Francis et al. in 2003. While the climatic regime shift of 1976-77 is well-documented (Hare and Mantua, 2000), the sustained and consistent positive influence of this warmer regime on Pacific cod recruitment lacks comprehensive documentation. Consequently, we undertook an investigation encompassing models that incorporate the catch data from 1964 to 1976, excluding the regime change parameter, and assuming an equilibrium catch of 10,000 tons. This exploration aimed to assess the sensitivity of the reference points to these initial assumptions regarding the impacts of the regime change on Pacific cod recruitment and 10,000 tons was a best guess rough estimate of catches prior to 1964.

4.3.2 Maximum age from age 20 to age 12

In 2022, the age plus group for model dynamics was maintained at 20 across all four ensemble models. This decision was made despite the fact that only one Pacific cod was observed in the Bering Sea shelf survey since 1993 with an age greater than 14, and only 34 fish were aged above 12 out of a total of 32,050 ages recorded. Starting from 2017, only two fish aged over 10 were identified out of a pool of 5,524 total ages collected from the bottom trawl survey. It's worth noting that due to processing limitations, the age composition analysis from the VAST survey had to be confined to an age 12 plus group.

In the near future, a shift in aging techniques is anticipated, moving towards the utilization of Fourier transform near-infrared spectroscopy (FT-NIRS) as detailed by Benson et al. (2023). Early findings pertaining to Pacific cod indicate notable discrepancies in age predictions beyond age 12 using FT-NIRS (communicated by Helser). Given these circumstances, we undertook an assessment to gauge the model's responsiveness to the transition to an age 12 plus group for model dynamics.

4.3.3 Results of changing catch years and decreasing the maximum age to 12

Two bridging models had been completed treating these two changes separately, but for brevity we chose to only present the combined model, Model 23.1.0.g. Neither of the changes to the model made much difference in the overall fit to the data (Table 2) with only a minor degradation in overall fit (+4.9 -LL). All three composition data components were only minutely impacted negatively with these two changes (+6.1 -LL) offset slightly by an improvement to the survey index (-1.1 –LL; Table 3 and Table 4).

Retrospective Mohn's rho for Model 23.1.0.g, although slightly positively increased over that of Model 23.1.0.d, remained within acceptable bounds at 0.11 (Table 5). The MASE analysis showed a slight improvement of predictive capability for the survey index and a slight degradation in predictive capability for all of the composition data sets (Table 6) consistent with the model component specific likelihoods, RMSE, and effective N results. These slight changes within the model did however impact survey catchability increasing it to 0.79 from 0.77 of Model 23.1.0.d. This increase in catchability was accompanied by an increase in natural mortality to 0.435 from 0.429, contrary to the trend of natural mortality decreasing with increasing Q. The log equilibrium recruitment $\ln(R_0)$ also increased to 13.74 from 13.69 consistent with the increase in natural mortality. A likelihood profile over survey catchability indicated that these model alterations did not improve fitting catchability as large changes in catchability continued to result in only small changes in likelihood (Table 13 and Figure 10).

The main difference in growth was an increase in L_{max} over Model 23.1.0.d due to the switch to the lower maximum age (Table 7). There was also a sharp increase in size and weight at the 12+ group and somewhat longer fish at ages 5 and 6 for the terminal year (Figure 1). The weight-at-age was nearly identical to Model 23.1.0.b (Figure 1) which is slightly lighter at age than Model 23.1.0.d. The trend in annually varying growth remained similar to Model 23.1.0.d despite restricting the age to 12+ (Figure 2) with a stable trend for young fish ages 0 to 6, increasing trend for fish ages 7 to 11, and a decreasing trend in size for the age 12+ group. The annual growth pattern in Model 23.1.0.d was consistent except the recent trend in increasing size was apparent through age 15 before turning to a recent decreasing trend for ages 16 to 20. This difference in the trend by ages between Model 23.1.0.g and Model 23.1.0.d was due to the shortening of the modeled ages. This did not substantially impact the trend in spawning stock biomass with only slightly smaller values (~-3%) after 1982 (Figure 9 and Table 9) consistent with the slightly higher survey catchability.

There was a substantial shift in the B_0 reference point from 623 kt from Model 23.1.0.d to 543 kt in Model 23.1.0.g (Table 9) and therefore despite the small decrease in overall female spawning biomass (Figure 9) the status of the stock in 2023 was higher at $B_{61\%}$ than the Model 23.1.0.d value of $B_{55\%}$ (Table 2 and Figure 8). The change in B_0 can be attributed to the extension of the catch time series and lower equilibrium catch. There was also a reduction in the uncertainty of B_0 to a CV of 0.057 from 0.087 from Model 23.1.0.d, however the uncertainty around the current spawning biomass and ABC projections remained the same (Table 9). $F_{40\%}$ also increased from 0.47 to 0.49 from Model 23.1.0.d to 23.1.0.g due to the addition of the 1964-1976 catch and removal of the regime change parameter. Despite this increase in allowable F the 2024 ABC projection was reduced from 244kt to 239kt due to lower spawning biomass in Model 23.1.0.g (Table 2 and Table 9). Therefore both changes resulted in minor adjustment in management advice.

Model 23.1.0.g continues to have the same issue as all the other models examined, changing the catch start year and reducing maximum age did not improve the profile on catchability. The log likelihood profile over catchability still showed the model to have low certainty in this value as large changes in catchability continue to result in small changes in likelihood (Table 13 and Figure 10). Catchability can be changed between 0.6 and 1.2 and still be within ±2 -LL from the MLE, resulting in a range in management advice for the 2024 ABC of ± 100,000t from the MLE.

Changing to the lower maximum age for the data will likely need to happen as the AFSC switches to FT-NIRS aging however changes within the model dynamics as explored here will not. Impact to the model fit and results looks to be minor with a slight increase in L_{max} and allowable fishing mortality. Including the earlier catch series may not be as clear and further work should be conducted to investigate the influence of these earlier catches and assumed equilibrium catch level. There is work currently underway by Dr. Catherine West, a zooarcheologist from Boston University, through a National Science Foundation grant (NSF award # 2220552) to better account for Pacific cod fishery catch from pre-1964 which may better inform the equilibrium catch level used in the model.

4.4 Model 23.1.0.h

4.4.1 Survey conditional-age-at length

Annually varying growth in the 2022 ensemble models is driven by length and age composition data. In one set of alternative models we explored the inclusion of survey conditional-age-at-length (CAAL) to determine if this improved model estimates of annually varying growth (Figure 20). We employed the same method for calculating CAAL and input sample sizes for the CAAL used in the Gulf of Alaska Pacific cod stock assessment and documented in Barbeaux et al. (2021). In theory the CAAL should provide an improved estimate of the growth parameters including annual devs. Note that because we changed the data in the model the total likelihood cannot be directly compared to previous models however the RMSE, effective N, and other individual likelihood components can be.

4.4.2 Results of adding survey conditional-age-at-length

The addition of the survey CAAL resulted in the fits to all be degraded overall (Table 3 and Table 4). Here we see an increase in the survey length composition of +12 -LL and fishery length composition of +15 –LL while the index had an increase of +9 -LL. There was a minor improvement to the age composition data with the addition of the CAAL 76.74 in Model 23.1.0.g to 75.45 in Model 23.1.0.h. Residual patterns remain similar to previous models (Figure 13, Figure 14, and Figure 21). Fits to the CAAL data were good for most years with an exception for 1992 and 1993 where the model predicted smaller fish at age than were observed (Figure 22). The Mohn's rho for spawning stock biomass showed an increase in positive retrospective bias over Model 23.1.0.g to 0.15. MASE evaluation of predictive skill were similar for the index and survey age, but somewhat worse for fishery length composition data. Model 23.1.0.h fails the residual runs test for both the fishery length composition and survey age composition suggesting significant autocorrelation in the residuals in fitting these data sets. A likelihood profile over survey catchability (Table 14 and Figure 10) shows little improvement with the introduction of CAAL continuing to generate small changes in likelihood over large changes in catchability. This is reflected in a large range in management advice as spawning biomass and ABC are scaled with catchability.

The main impact of the addition of CAAL in the model was to fit a smaller L_{min}, a higher K, and reduce the standard deviation of all the growth parameters (Figure 23 and Table 7), as well as a small reductions in uncertainty for R_0, natural mortality, and catchability (Table 8). The reduction in variance in parameters translated into lower uncertainty in the derived quantities such as fishing mortality, unfished spawning biomass, spawning biomass, and projected ABCs (Table 9). The trends in annually varying length-at-age were similar to Model 23.1.0.d and 23.1.0.g, but with slightly lower interannual variability for those years (1990-2022) with CAAL data. At ages 3 to 11 there is an increasing trend in size-at-age from 2000 forward, and a decreasing trend in size-at-age at age 12+. Random walk devs on L_{min} are relatively consistent for Models 23.1.0.g and Model23.1.0.h (Figure 22) with an overall increasing trend over the time series suggesting an increase in size of juvenile fish in the Bering Sea. Richards parameter impacts the rate of fish growth and is consistent between models prior to the introduction of CAAL from 1977-1986, after which the series diverges with low values for Model 23.1.0.h that then increase over time.

The forecast for Model 23.1.0.h was set to average biological parameters back to 1964, this resulted in some aberrant behavior for the forecasted weight-at-age resulting in these values being substantially higher than the other models (Figure 1). As projections had not been considered in developing these models, this issue went unnoticed until very recently and the authors have not had to time to correct this issue. This would impact model projections including the 2024 ABC which was incorrect and has been removed from all tables. This issue would not impact the time series fits or model performance.

In summary the addition of CAAL reduces uncertainty throughout the model, but at a cost of degrading fits to all of the other model components (Table 3 and Table 4).

5 A Case for Fixing Natural Mortality

One way that could be used to alleviate model sensitivities and that has been explored in the past is to fix key model parameters or provide informative priors, e.g. natural mortality or catchability at 'reasonable' values. It's crucial to emphasize that at its base, fitting catchability within the model is influenced by the degree to which catch impacts changes in survey abundance. The fluctuations in growth and recruitment due to environmental factors can significantly contribute to the overall biomass's variability independent of catch, leading to limited insights into the consequences of fishery removals and therefore little information within the data to inform catchability. There has been substantial debate in the past over catchability with an equal amount of work going into studies to try to better understand this value, the results of which have been equivocal. Given the considerable impact of environmental drivers on cod abundance and mortality, there's a possibility of never having sufficient data to accurately determine survey catchability. Issues with aging Pacific cod in the past have made estimating natural mortality unreliable with estimates varying from 0.20 to 0.96 across the spectrum of Pacific cod stocks (Thompson 2018). However recent improvements in methods may provide a more reliable means of estimating natural mortality outside the model.

5.1 Estimating a life-history-based prior for natural mortality

The parameter M representing natural mortality is difficult to estimate in many stock assessment models. When total removals are fitted and information exists to estimate the fishing mortality rate, estimates of M are typically correlated with estimates of survey catchability, q, such that including a Bayesian prior on M can provide information about population scale and resulting catch limits.

Substantial empirical and theoretical evidence suggests that natural mortality is lower for large bodied individuals (Andersen, 2019). Asymptotic body length L_inf is negatively correlated with the von Bertalanffy growth parameter k, such that these two growth parameters are sometimes used to predict M (Hoenig, 1983). In fact, the ratio M/k has erroneously been called a "life-history invariant" (Roff, 1984), despite theory suggesting that higher M/k is associated with lower L_mat/L_inf (Beverton & Holt, 1959). In particular, some taxa evolve behavioral and morphological defenses against predators (e.g., spines) that likely contribute to a lower M/k than otherwise expected (Thorson et al., 2014). These antipredator defenses may in some cases be evolutionarily conserved, such that a lower-than-expected M/k for a related taxa will be informative when predicting the value of M from k for a given species. This intuition gives rise to taxonomic-nested linear mixed models or phylogenetic trait imputation, which have been used to impute missing values for natural mortality (Thorson et al., 2017), recruitment density dependence (Thorson, 2020), or other behavioral and ecological traits (Thorson et al., 2023).

As an alternative to estimating natural mortality from growth parameters, researchers have also compiled estimates of longevity from aged specimens, and research suggests that longevity-based predictions of natural mortality rate are more precise than growth-based estimates (Hamel & Cope, 2022; Then et al., 2015). Longevity can be recorded either as the maximum aged specimen, or the average of the five maximum ages (Sullivan et al., 2022). However, developing separate estimators using longevity and growth parameters then results in multiple estimators for a given species (Sullivan et al., 2022), which presents a challenge in either selecting a single estimator or weighting alternative estimators within an ensemble (Cope & Hamel, 2022).

As alternative to developing separate models using growth or longevity information, recent research has developed phylogenetic structural equation models, which can explicitly represent the dependency among multivariate trait data (Thorson et al., 2023; van der Bijl, 2018; von Hardenberg & Gonzalez-Voyer, 2013). In particular, a user-friendly R-package phylosem can impute missing trait values jointly with estimating complex dependencies among traits (Thorson & van der Bijl, In review). Research confirms that phylosem exactly replicates results from simpler models including structural equation models, phylogenetic linear models, and phylogenetic trait imputation (Thorson & van der Bijl, In review).

Here, we fit a phylogenetic structural equation model (PSEM) to a high-quality database of independent estimates of natural mortality (Then et al., 2015). We specifically use a PSEM that specifies three linear associations $\log(L_{inf}) \rightarrow \log(t_{max}), \log(k) \rightarrow \log(t_{max}), \text{and } t_{max} \rightarrow \log(M)$. A jackknife experiment confirms that this PSEM can explain nearly 50% additional variance relative to a conventional linear model when using growth parameters to predict natural mortality rate, while also providing a simple method to include both growth and longevity information in a single natural mortality estimator (Thorson, In review). We then use either the maximum specimen age, or the average of the maximum ages to predict natural mortality rate for Pacific cod in the eastern Bering Sea since 2008. Both longevity metrics result in the same value t_max=14 years, and this results in a predicted value M=0.3866 and log standard deviation of 0.4.

All of the models considered above were refit with the maximum age derived value of natural mortality of 0.3866 and presented in Table 15. In all of the newly developed simpler models the change in model fit by fixing natural mortality was minimal with likelihoods changing by +2 -LL or less. The fixing of natural mortality also resulted in improved retrospective runs as would be expected (Figure 25). Survey catchability fit in the models changed in the direction one would expect as it is negatively correlated with natural mortality, when the fixed M value was higher than the fit value catchability decreased, when it was lower it increased (Table 16).

6 Recommendation

If a single model is to be used for management this year, of the models presented the authors would recommend Model 23.1.0.d be considered as an option for further exploration for management in 2024. Model 23.1.0.d is the simplified model with the addition of constrained annual variability in growth (Figure 1 and Figure 2) and survey selectivity (Figure 3 and Figure 4). The model employs bootstrapped input sample sizes, the fit to the survey index is improved over Model 22.2 (Figure 5), retrospective bias remains within acceptable limits, and residual runs analyses imply that the fit to the composition data are better with less autocorrelation in the residuals while maintaining a similar effective n. The mean absolute scaled error (MASE) analyses show the predictive skill of Model 23.1.0.d for the survey index was improved over Model 22.2 (0.40 vs 0.69) and remains similar for the fishery mean length and survey mean age.

Model 23.1.0.d has improved performance while reducing the model by 86 parameters, however this reduction in complexity comes at a cost of increased uncertainty (higher standard deviations) in some key parameter estimates (M, Q, and R_0; Figure 6) that translates into increased uncertainty in derived quantities such as B_0, $F_{40\%}$, current recruitment (Figure 7), current spawning biomass (Figure 8 and Figure 9), and future catch recommendations. Another potential red-flag in Model 23.1.0.d is that natural mortality at 0.429 is higher than most methods external to the model indicate and catchability at 0.765 is lower than most other models examined previously. Profiles over catchability show little change in likelihood over a wide range of natural mortality and catchability (Figure 10).

One solution may be found in applying a maximum age-based method for deriving a prior for natural mortality that provides a value of natural morality within reasonable bounds of what has been fit in the simplified models presented above. When natural mortality is fixed in Model 23.1.0.d to 0.387, catchability increased to 0.972 and provides a reasonable alternative to allowing natural mortality to be freely fit. Fixing M in Model 23.1.0.d has a minor cost of degrading the overall model performance by only +1.4 negative log likelihood (-LL) with an improvement to the fit to the age composition of -3.95 -LL, but a poorer fit to the survey index at +2.56 LL and length composition at +2.40 -LL. The change provides a more stable model but makes strong assumptions on the value of natural mortality and by association survey catchability (Table 16).

7 References

Andersen. (2019). Fish Ecology, Evolution, and Exploitation. Princeton University Press. https://press.princeton.edu/books/hardcover/9780691176550/fish-ecology-evolution-and-exploitation

Barbeaux, S.J., Barnett, L., Connor, J., Nielson, J., Shotwell, S.K., Siddon, E., Spies, I., Ressler, H.R., Rohan, S., Sweeney, K. and Thompson, G., 2022. 2. Assessment of the Pacific Cod Stock in the Eastern Bering Sea. Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Bering Sea and Aleutian Islands. North Pacific Fishery Management Council, 1007.

Barbeaux, S., Ferriss, B. Palsson, W., Shotwell, K., Spies, I., Wang, M. and Zador, S. 2021. Assessment of the Pacific cod stock in the Gulf of Alaska. In Stock assessment and fishery evaluation report for the groundfish resources of the Gulf of Alaska. North Pac. Fish. Mgmt. Council, Anchorage, Alaska

Benson, I.M., Helser, T.E., Marchetti, G. and Barnett, B.K., 2023. The future of fish age estimation: deep machine learning coupled with Fourier transform near-infrared spectroscopy of otoliths. Canadian Journal of Fisheries and Aquatic Sciences.

Beverton, R., & Holt, S. (1959). A review of the lifespans and mortality rates of fish in nature, and their relation to growth and other physiological characteristics. In G. E. W. Wolstenholme & M. O'Conner (Eds.), Ciba Foundation Symposium-The Lifespan of Animals (Colloquia on Ageing) (pp. 142–177). J. and A. Churchil Ltd.

Carvalho, F., Winker, H., Courtney, D., Kapur, M., Kell, L., Cardinale, M., Schirripa, M., Kitakado, T., Yemane, D., Piner, K.R. and Maunder, M.N., 2021. A cookbook for using model diagnostics in integrated stock assessments. Fisheries Research, 240, p.105959.

Cope, J. M., & Hamel, O. S. (2022). Upgrading from M version 0.2: An application-based method for practical estimation, evaluation and uncertainty characterization of natural mortality. Fisheries Research, 256, 106493. https://doi.org/10.1016/j.fishres.2022.106493

Hamel, O. S., & Cope, J. M. (2022). Development and considerations for application of a longevity-based prior for the natural mortality rate. Fisheries Research, 256, 106477. https://doi.org/10.1016/j.fishres.2022.106477

Hare, S.R. and Mantua, N.J., 2000. Empirical evidence for North Pacific regime shifts in 1977 and 1989. Progress in oceanography, 47(2-4), pp.103-145.

Hoenig, J. M. (1983). Empirical use of longevity data to estimate mortality rates. Fishery Bulletin, 82(4), 898–903.

Hulson, P-J. F., B. C. Williams, M. R. Siskey, M. D. Bryan, and J. Conner. 2023. Bottom trawl survey age and length composition input sample sizes for stocks assessed with statistical catch-at-age assessment models at the Alaska Fisheries Science Center. U.S. Dep. Commer., NOAA Tech. Memo.NMFS-AFSC-470, 38 p.

Kastelle, C.R., Helser, T.E., McKay, J.L., Johnston, C.G., Anderl, D.M., Matta, M.E. and Nichol, D.G., 2017. Age validation of Pacific cod (Gadus macrocephalus) using high-resolution stable oxygen isotope (δ 18O) chronologies in otoliths. Fisheries Research, 185, pp.43-53.

Laurel, B.J., Hurst, T.P., Copeman, L.A. and Davis, M.W., 2008. The role of temperature on the growth and survival of early and late hatching Pacific cod larvae (Gadus macrocephalus). Journal of Plankton Research, 30(9), pp.1051-1060.

Mackovjak, J., 2019. Alaska codfish chronicle: A history of the Pacific cod fishery in Alaska. University of Alaska Press.

Methot Jr., R. R., Wetzel, C. R., Taylor, I. G., Doering, K .L. and Johnson, K. F. 2023. Stock Synthesis User Manual Version 3.30.21. , NOAA Fisheries, Seattle, WA Available: https://nmfs-stock-synthesis.github.io/doc/SS330_User_Manual_release.html#tvOrder

Roff, D. A. (1984). The evolution of life history parameters in teleosts. Canadian Journal of Fisheries and Aquatic Sciences, 41(6), 989–1000.

Stewart, I.J. and Hamel, O.S., 2014. Bootstrapping of sample sizes for length-or age-composition data used in stock assessments. Canadian journal of fisheries and aquatic sciences, 71(4), pp.581-588.

Sullivan, J. Y., C. A. Tribuzio, and K. B. Echave. 2022. A review of available life history data and updated estimates of natural mortality for several rockfish species In Alaska. U.S. Dep. Commer.,NOAA Tech. Memo. NMFS-AFSC-443, 45 p.

Taylor,C.C. 1958. Cod Growth and Temperature, ICES Journal of Marine Science, 23(3). pp366–370, https://doi.org/10.1093/icesjms/23.3.366

Taylor, I.G., Doering, K.L., Johnson, K.F., Wetzel, C.R., Stewart, I.J., 2021. Beyond visualizing catch-at-age models: Lessons learned from the r4ss package about software to support stock assessments. Fisheries Research, 239:105924 https://doi.org/10.1016/j.fishres.2021.105924

Then, A. Y., Hoenig, J. M., Hall, N. G., Hewitt, D. A., & Handling editor: Ernesto Jardim. (2015). Evaluating the predictive performance of empirical estimators of natural mortality rate using information on over 200 fish species. ICES Journal of Marine Science, 72(1), 82–92. https://doi.org/10.1093/icesjms/fsu136

Thompson, G. 2018. 2. Assessment of the Pacific Cod Stock in the Eastern Bering Sea. Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Bering Sea and Aleutian Islands. Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Bering Sea and Aleutian Islands. North Pacific Fishery Management Council, 1007

Thompson, G., Barbeaux, S., Conner, J., Fissel, B., Hurst, T., Laurel, B., O'Leary, C., Rogers, L., Shotwell, K., Siddon, E., Spies, I., Thorson, J. and Tyrell, 2021. 2. Assessment of the Pacific Cod Stock in the Eastern Bering Sea. Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Bering Sea and Aleutian Islands. North Pacific Fishery Management Council, 1007

Thorson, J. T. (2020). Predicting recruitment density dependence and intrinsic growth rate for all fishes worldwide using a data-integrated life-history model. Fish and Fisheries, 21(2), 237–251. https://doi.org/10.1111/faf.12427

Thorson, J. T. (In review). Trees for fishes: The neglected role for phylogenetic comparative methods in fisheries science. Fish and Fisheries.

Thorson, J. T., Maureaud, A. A., Frelat, R., Mérigot, B., Bigman, J. S., Friedman, S. T., Palomares, M. L. D., Pinsky, M. L., Price, S. A., & Wainwright, P. (2023). Identifying direct and indirect associations among traits by merging phylogenetic comparative methods and structural equation models. Methods in Ecology and Evolution, 14(5), 1259–1275. https://doi.org/10.1111/2041-210X.14076

Thorson, J. T., Munch, S. B., Cope, J. M., & Gao, J. (2017). Predicting life history parameters for all fishes worldwide. Ecological Applications, 27(8), 2262–2276. https://doi.org/10.1002/eap.1606

Thorson, J. T., & van der Bijl, W. (In review). phylosem: A fast and simple R package for phylogenetic inference and trait imputation using phylogenetic structural equation models. Journal of Evolutionary Biology.

Thorson, Taylor, I. G., Stewart, I., & Punt, A. E. (2014). Rigorous meta-analysis of life history correlations by simultaneously analyzing multiple population dynamics models. Ecological Applications, 24, 315–326.

van der Bijl, W. (2018). phylopath: Easy phylogenetic path analysis in R. PeerJ, 6, e4718. https://doi.org/10.7717/peerj.4718

von Hardenberg, A., & Gonzalez-Voyer, A. (2013). Disentangling evolutionary cause-effect relationships with phylogenetic confirmatory path analysis. Evolution; International Journal of Organic Evolution, 67(2), 378–387. https://doi.org/10.1111/j.1558-5646.2012.01790.x

Winker H, Carvalho F, Cardinale M, and Kell L .2023. _ss3diags: What the Package Does (One Line, Title Case)_. R package version 1.10.0.

8 Tables

Table 1. Input sample sizes for composition data, the old based on survey haul numbers and 'New' on a bootstrap approach (Hulson et al. 2023).

	Fishery		Survey		
Year	**Old**	**New**	**Old**	**New Length**	**New Age**
1977	6	26			
1978	10	42			
1979	12	52			
1980	12	53			
1981	14	61			
1982	7	30	481	2432	
1983	26	112	476	1171	
1984	31	135	479	2424	
1985	46	203	364	897	
1986	47	207	481	2139	
1987	87	380	412	2104	
1988	89	387	354	1650	
1989	41	179	373	1176	
1990	42	184	354	1226	
1991	345	1506	400	1200	
1992	340	1485	368	807	
1993	201	880	451	813	
1994	317	1383	360	1265	183
1995	344	1503	381	1999	174
1996	445	1943	368	1343	151
1997	472	2063	354	1389	98
1998	451	1972	360	2196	180
1999	600	2622	422	2078	224
2000	652	2849	363	1396	154
2001	692	3025	402	1829	304
2002	759	3318	366	2159	329
2003	947	4138	355	1040	265
2004	794	3471	336	1887	308
2005	761	3328	362	1164	212
2006	594	2595	369	2487	492
2007	466	2035	359	270	55
2008	551	2409	347	1757	235
2009	488	2134	364	908	201
2010	435	1902	363	1191	150
2011	572	2498	332	1398	127
2012	611	2670	330	865	150
2013	726	3171	329	909	149
2014	793	3467	293	1057	124
2015	733	3202	370	2068	362
2016	621	2715	339	3149	536
2017	544	2377	349	2802	447
2018	418	1827	369	2996	367
2019	301	1316	264	1230	250
2020	231	1008	NA	NA	NA
2021	189	827	255	3167	531
2022	128	1115	320	2388	NA
Mean	369	1626	369	1661	250

Table 2. Results from 2023 model exploration. The table shows Natural mortality (M),bottom trawl survey catchability (Q), unfished female spawning biomass (B_0), female spawning biomass in 2023 (B_{23}), projected Allowable Biological Catch in 2024 (ABC_{24}), number of non-dev parameters (Npars), number of annual devs (Ndevs), and negative log likelihood (-LL) by model.

M	Q	B_0 (kt)	F_{MSY}	B_{23} (kt)	B_{23}/B_0	ABC_{24} (kt)	Npars	Ndevs	-LL	Model
0.347	0.960	661.5	0.326	249.8	0.378	145	20	284	10875	MODEL 22.2 old
0.328	0.974	694.7	0.290	263.2	0.379	141	22	284	18362	MODEL 22.2 updated
0.344	1.097	586.1	0.332	205.9	0.351	132	18	64	251	MODEL 23.1.0.a
0.414	0.822	605.4	0.441	314.1	0.519	220	18	158	143	MODEL 23.1.0.b
0.429	0.765	623.4	0.465	343.4	0.551	244	18	200	133	MODEL 23.1.0.d
0.435	0.792	542.6	0.488	331.8	0.612	239	17	200	141	MODEL 23.1.0.g
0.424	0.808	611.4	0.466	313.1	0.512		17	200	631	MODEL 23.1.0.h

Table 3. Negative log likelihoods by data component and fleet.

Model	Label	All	Fishery	Survey
MODEL 22.2 old	Age_like	817.80		817.80
MODEL 22.2 updated	Age_like	766.34		766.34
MODEL 23.1.0.a	Age_like	88.62		88.62
MODEL 23.1.0.b	Age_like	71.07		71.07
MODEL 23.1.0.d	Age_like	73.27		73.27
MODEL 23.1.0.g	Age_like	76.74		76.74
MODEL23.1.0.h	Age_like	75.45		75.45
MODEL 22.2 old	Length_like	9990.5	4502.5	5487.98
MODEL 22.2 updated	Length_like	17382.5	7682.9	9699.66
MODEL 23.1.0.a	Length_like	184.38	79.03	105.35
MODEL 23.1.0.b	Length_like	130.75	60.29	70.46
MODEL 23.1.0.d	Length_like	120.58	59.78	60.81
MODEL 23.1.0.g	Length_like	123.19	60.46	62.73
MODEL 23.1.0.h	Length_like	150.98	76.04	74.94
MODEL 22.2 old	Surv_like	-5.96		-5.96
MODEL 22.2 updated	Surv_like	67.53		67.53
MODEL 23.1.0.a	Surv_like	-30.05		-30.05
MODEL 23.1.0.b	Surv_like	-83.13		-83.13
MODEL 23.1.0.d	Surv_like	-88.62		-88.62
MODEL 23.1.0.g	Surv_like	-89.74		-89.74
MODEL 23.1.0.h	Surv_like	-80.49		-80.49
MODEL 23.1.0.h	Survey CAAL	445.62		445.62

Table 4. Root mean squared error RMSE and effective N for data components by model. For the fishery length on Models 23.1.1.x gear splits are in order of Trawl/Longline/Pot

		Effective N			
Index RMSE	Recruitment RMSE/SigmaR	Fishery Length	Survey Length	Survey Age	Model
0.13	1.01	2919	852	168	MODEL 22.2_old
0.16	1.22	3474	929	122	MODEL 22.2_updated
0.12	1.06	1700	561	87	MODEL 23.1.0.a
0.07	0.81	2263	813	132	MODEL 23.1.0.b
0.07	0.77	2288	899	132	MODEL 23.1.0.d
0.07	0.82	2242	860	120	MODEL 23.1.0.g
0.08	0.71	1867	691	33	MODEL 23.1.0.h

Table 5. Retrospective results (Mohn's Rho) for a ten-year peal on spawning stock biomass and mean absolute scaled error (MASE) analyses from ss3diags library for components of models assessed.

		MASE		
Model	Mohn's Rho	Index	Fish Length (adj.)	Survey Age
M22.2 old	-0.06 (-0.07)	0.69	0.93 (0.21)	0.35
M22.2 updated	-0.06 (-0.07)	0.69	0.93 (0.21)	0.35
M23.1.0.a	0.08 (0.10)	0.42	1.07 (0.24)	0.32
M23.1.0.b	0.09 (0.07)	0.40	0.97 (0.22)	0.38
M23.1.0.d	0.09 (0.07)	0.40	0.96 (0.22)	0.38
M23.1.0.g	0.11 (0.08)	0.39	0.97 (0.22)	0.41
M23.1.0.h	0.15 (0.14)	0.40	1.05 (0.24)	0.41

Table 6. Residual runs test for models evaluated with combined fishery comp data from ss3diags.

Model	Type	Index	p-value	Test	Sigma3 lo	Sigma3 hi
MODEL 22.2 old	cpue	Survey	0.280	Passed	-0.376	0.376
MODEL 22.2 updated	cpue	Survey	0.261	Passed	-0.433	0.433
MODEL 23.1.0.a	cpue	Survey	0.850	Passed	-0.424	0.424
MODEL 23.1.0.b	cpue	Survey	0.903	Passed	-0.260	0.260
MODEL 23.1.0.d	cpue	Survey	0.903	Passed	-0.222	0.222
MODEL 23.1.0.g	cpue	Survey	0.974	Passed	-0.227	0.227
MODEL 23.1.0.h	cpue	Survey	0.903	Passed	-0.265	0.265
Model22.2 old	len	Fishery	0.002	**Failed**	-0.024	0.024
Model22.2 old	len	Survey	0.000	**Failed**	-0.077	0.077
MODEL 22.2 updated	len	Fishery	0.009	**Failed**	-0.019	0.019
MODEL 22.2 updated	len	Survey	0.122	Passed	-0.090	0.090
MODEL 23.1.0.a	len	Fishery	0.003	**Failed**	-0.066	0.066
MODEL 23.1.0.a	len	Survey	0.625	Passed	-0.100	0.100
MODEL 23.1.0.b	len	Fishery	0.155	Passed	-0.060	0.060
MODEL 23.1.0.b	len	Survey	0.815	Passed	-0.125	0.125
MODEL 23.1.0.d	len	Fishery	0.155	Passed	-0.060	0.060
MODEL 23.1.0.d	len	Survey	0.462	Passed	-0.087	0.087
MODEL 23.1.0.g	len	Fishery	0.155	Passed	-0.061	0.061
MODEL 23.1.0.g	len	Survey	0.815	Passed	-0.083	0.083
MODEL 23.1.0.h	len	Fishery	0.015	**Failed**	-0.075	0.075
MODEL 23.1.0.h	len	Survey	0.625	Passed	-0.083	0.083
MODEL l22.2_old	age	Survey	0.039	**Failed**	-0.160	0.160
MODEL 22.2_updated	age	Survey	0.401	Passed	-0.199	0.199
MODEL 23.1.0.a	age	Survey	0.177	Passed	-0.250	0.250
MODEL 23.1.0.b	age	Survey	0.086	Passed	-0.161	0.161
MODEL 23.1.0.d	age	Survey	0.298	Passed	-0.160	0.160
MODEL 23.1.0.g	age	Survey	0.086	Passed	-0.153	0.153
MODEL 23.1.0.h	age	Survey	0.016	**Failed**	-0.152	0.152

Table 7. Growth parameter values and standard deviations.

Label	*Value*	*StDev*	*Model*	*Label*	*Value*	*StDev*	*Model*
L_{MAX}	112.387	3.05	Model22.2_old	Richards	1.474	0.04	Model22.2_old
L_{MAX}	116.862	1.78	Model22.2_updated	Richards	1.541	0.02	Model22.2_updated
L_{MAX}	112.958	5.92	MODEL23.1.0.a	Richards	1.494	0.11	MODEL23.1.0.a
L_{MAX}	112.380	3.24	MODEL23.1.0.b	Richards	1.529	0.08	MODEL23.1.0.b
L_{MAX}	112.355	3.24	MODEL23.1.0.d	Richards	1.528	0.08	MODEL23.1.0.d
L_{MAX}	113.217	3.23	MODEL23.1.0.g	Richards	1.539	0.08	MODEL23.1.0.g
L_{MAX}	110.918	2.28	MODEL23.1.0.h	Richards	1.535	0.07	MODEL23.1.0.h
L_{MIN}	15.134	0.45	Model22.2_old	VonBert K	0.115	0.009	Model22.2_old
L_{MIN}	15.648	0.44	Model22.2_updated	VonBert K	0.100	0.004	Model22.2_updated
L_{MIN}	14.772	0.24	MODEL23.1.0.a	VonBert K	0.110	0.021	MODEL23.1.0.a
L_{MIN}	14.674	0.20	MODEL23.1.0.b	VonBert K	0.112	0.011	MODEL23.1.0.b
L_{MIN}	14.713	0.21	MODEL23.1.0.d	VonBert K	0.113	0.011	MODEL23.1.0.d
L_{MIN}	14.708	0.21	MODEL23.1.0.g	VonBert K	0.109	0.011	MODEL23.1.0.g
L_{MIN}	14.681	0.20	MODEL23.1.0.h	VonBert K	0.131	0.009	MODEL23.1.0.h

Table 8. Influential parameter values and standard deviations.

Label	*Value*	*StDev*	*Model*	*Label*	*Value*	*StDev*	*Model*
LN(R_0)	13.156	0.100	Model22.2_old	NatM	0.347	0.012	Model22.2_old
LN(R_0)	13.016	0.075	Model22.2_updated	NatM	0.328	0.009	Model22.2_updated
LN(R_0)	13.022	0.141	MODEL23.1.0.a	NatM	0.344	0.018	MODEL23.1.0.a
LN(R_0)	13.601	0.241	MODEL23.1.0.b	NatM	0.414	0.026	MODEL23.1.0.b
LN(R_0)	13.740	0.248	MODEL23.1.0.d	NatM	0.429	0.025	MODEL23.1.0.d
LN(R_0)	13.688	0.240	MODEL23.1.0.g	NatM	0.435	0.025	MODEL23.1.0.g
LN(R_0)	13.669	0.175	MODEL23.1.0.h	NatM	0.424	0.021	MODEL23.1.0.h
LnQ BT Shelf Survey	-0.041	0.064	Model22.2_old				
LnQ BT Shelf Survey	-0.026	0.049	Model22.2_updated				
LnQ BT Shelf Survey	0.092	0.086	MODEL23.1.0.a				
LnQ BT Shelf Survey	-0.196	0.163	MODEL23.1.0.b				
LnQ BT Shelf Survey	-0.268	0.172	MODEL23.1.0.d				
LnQ BT Shelf Survey	-0.233	0.162	MODEL23.1.0.g				
LnQ BT Shelf Survey	-0.213	0.104	MODEL23.1.0.h				

Table 9. Derived quantities values, standard deviations, and coefficient of variation.

Label	Value	StdDev	CV	Model	Label	Value	StdDev	CV	Model
F_{40}	0.33	0.02	0.05	Model22.2_old	B_{2023}	249,809	17,360	0.07	Model22.2_old
F_{40}	0.29	0.01	0.04	Model22.2_updated	B_{2023}	263,189	14,151	0.05	Model22.2_updated
F_{40}	0.33	0.03	0.09	MODEL23.1.0.a	B_{2023}	205,914	19,749	0.10	MODEL23.1.0.a
F_{40}	0.44	0.05	0.10	MODEL23.1.0.b	B_{2023}	314,146	58,787	0.19	MODEL23.1.0.b
F_{40}	0.47	0.05	0.11	MODEL23.1.0.d	B_{2023}	343,431	66,590	0.19	MODEL23.1.0.d
F_{40}	0.49	0.05	0.10	MODEL23.1.0.g	B_{2023}	331,845	62,266	0.19	MODEL23.1.0.g
F_{40}	0.47	0.04	0.09	MODEL23.1.0.h	B_{2023}	313,052	38,688	0.12	MODEL23.1.0.h
ABC_{2024}	144,694	14,664	0.10	Model22.2_old	B_0	661,455	14,493	0.02	Model22.2_old
ABC_{2024}	141,115	11,792	0.08	Model22.2_updated	B_0	694,750	12,587	0.02	Model22.2_updated
ABC_{2024}	131,883	18,010	0.14	MODEL23.1.0.a	B_0	586,050	27,073	0.05	MODEL23.1.0.a
ABC_{2024}	219,817	49,257	0.22	MODEL23.1.0.b	B_0	605,435	50,776	0.08	MODEL23.1.0.b
ABC_{2024}	243,533	56,378	0.23	MODEL23.1.0.d	B_0	623,435	54,253	0.09	MODEL23.1.0.d
ABC_{2024}	239,088	53,953	0.23	MODEL23.1.0.g	B_0	542,635	30,880	0.06	MODEL23.1.0.g
ABC_{2024}			0.20	MODEL23.1.0.h	B_0	611,365	23,726	0.04	MODEL23.1.0.h

Table 10. Likelihood profiles over survey catchability for the old input sample size and updated input sample size Model 22.2. Light shaded rows are ± 10LL from the MLE, dark shaded row is the closest to MLE.

M	Q	B_0	F_{MSY}	B_{2023}	B_{2023}/B_0	ABC_{2024}	Model	-LL
0.408	0.607	702,510	0.493	372,577	0.530	243,473	Model 22.2 old	10893
0.402	0.638	690,370	0.489	355,882	0.515	231,967	Model 22.2 old	10889
0.396	0.670	680,070	0.480	340,092	0.500	220,867	Model 22.2 old	10886
0.390	0.705	671,990	0.468	325,376	0.484	210,231	Model 22.2 old	10884
0.384	0.741	665,730	0.453	311,391	0.468	199,983	Model 22.2 old	10883
0.378	0.779	660,385	0.448	297,939	0.451	190,460	Model 22.2 old	10880
0.370	0.819	659,620	0.427	287,355	0.436	181,953	Model 22.2 old	10879
0.363	0.861	656,310	0.424	273,501	0.417	167,867	Model 22.2 old	10876
0.356	0.905	657,270	0.412	262,236	0.399	155,598	Model 22.2 old	10876
0.348	0.951	660,520	0.399	251,553	0.381	146,301	Model 22.2 old	10875
0.340	1.000	666,530	0.385	241,723	0.363	137,214	Model 22.2 old	10875
0.331	1.051	674,500	0.371	232,258	0.344	128,133	Model 22.2 old	10876
0.323	1.105	685,095	0.357	223,100	0.326	119,010	Model 22.2 old	10878
0.314	1.162	698,290	0.343	213,998	0.306	109,750	Model 22.2 old	10882
0.304	1.221	715,160	0.329	205,686	0.288	100,832	Model 22.2 old	10885
0.294	1.284	735,990	0.316	198,475	0.270	92,520	Model 22.2 old	10889
0.284	1.350	760,510	0.302	190,942	0.251	83,996	Model 22.2 old	10895
0.274	1.419	789,875	0.288	183,723	0.233	75,680	Model 22.2 old	10902
0.263	1.492	824,435	0.275	176,848	0.215	67,701	Model 22.2 old	10911
0.251	1.568	876,375	0.258	171,031	0.195	59,398	Model 22.2 old	10919
0.239	1.649	928,585	0.245	164,735	0.177	51,936	Model 22.2 old	10929
0.387	0.607	721,175	0.457	384,413	0.533	233,327	Model 22.2 update	18423
0.382	0.638	710,870	0.447	367,428	0.517	222,109	Model 22.2 update	18416
0.375	0.670	705,800	0.422	355,832	0.504	213,090	Model 22.2 update	18437
0.369	0.705	695,605	0.413	339,071	0.487	202,348	Model 22.2 update	18428
0.363	0.741	690,235	0.402	325,001	0.471	192,820	Model 22.2 update	18423
0.358	0.779	687,200	0.393	312,083	0.454	184,339	Model 22.2 update	18421
0.355	0.819	686,660	0.384	306,398	0.446	179,781	Model 22.2 update	18396
0.343	0.861	686,650	0.379	285,626	0.416	161,963	Model 22.2 update	18386
0.339	0.905	686,880	0.370	279,310	0.407	155,643	Model 22.2 update	18363
0.329	0.951	689,040	0.365	259,098	0.376	139,004	Model 22.2 update	18370
0.323	1.000	695,900	0.354	249,146	0.358	130,620	Model 22.2 update	18372
0.315	1.051	703,190	0.343	238,344	0.339	121,292	Model 22.2 update	18369
0.304	1.105	716,200	0.325	228,459	0.319	111,728	Model 22.2 update	18379
0.293	1.162	734,860	0.310	219,578	0.299	102,230	Model 22.2 update	18376
0.283	1.221	760,165	0.296	212,295	0.279	93,747	Model 22.2 update	18378
0.279	1.284	781,845	0.289	204,729	0.262	86,500	Model 22.2 update	18381
0.264	1.350	823,400	0.270	196,966	0.239	76,732	Model 22.2 update	18392
0.255	1.419	844,445	0.260	188,304	0.223	69,103	Model 22.2 update	18397
0.245	1.492	888,745	0.247	181,505	0.204	61,267	Model 22.2 update	18406
0.233	1.568	946,240	0.233	175,511	0.185	53,735	Model 22.2 update	18416
0.222	1.649	1,009,605	0.220	169,199	0.168	46,607	Model 22.2 update	18429

Table 11. Model 23.1.0.a and Model 23.1.0.b likelihood profiles over catchability. Light shaded rows are ± 2LL from the MLE, dark shaded row is the closest to MLE. *Note hit the lower bound for natural mortality at 0.3.

M	Q	B_0	F_{MSY}	B_{2023}	B_{2023}/B_0	ABC_{2024}	Model	-LL
0.428	0.607	628,560	0.448	358,028	0.570	266,750	Model23.1.0.a	263.27
0.422	0.638	615,975	0.440	340,813	0.553	253,241	Model23.1.0.a	261.84
0.416	0.670	604,905	0.432	324,522	0.536	240,370	Model23.1.0.a	260.43
0.410	0.705	595,370	0.424	309,117	0.519	228,105	Model23.1.0.a	259.04
0.404	0.741	587,395	0.415	294,561	0.501	216,417	Model23.1.0.a	257.69
0.397	0.779	581,020	0.406	280,816	0.483	205,276	Model23.1.0.a	256.39
0.390	0.819	576,295	0.396	267,847	0.465	194,658	Model23.1.0.a	255.16
0.383	0.861	573,285	0.386	255,623	0.446	184,534	Model23.1.0.a	254.04
0.375	0.905	572,080	0.375	244,109	0.427	174,603	Model23.1.0.a	253.04
0.368	0.951	572,765	0.365	233,276	0.407	160,938	Model23.1.0.a	252.2
0.359	1.000	575,445	0.353	223,096	0.388	149,870	Model23.1.0.a	251.55
0.351	1.051	580,240	0.342	213,531	0.368	140,094	Model23.1.0.a	251.15
0.342	1.105	587,305	0.330	204,553	0.348	130,375	Model23.1.0.a	251.04
0.333	1.162	596,790	0.318	196,137	0.329	120,757	Model23.1.0.a	251.28
0.324	1.221	608,855	0.306	188,270	0.309	111,313	Model23.1.0.a	251.93
0.315	1.284	623,360	0.294	182,351	0.293	103,208	Model23.1.0.a	253.05
0.307	1.350	639,265	0.283	178,992	0.280	97,047	Model23.1.0.a	254.52
0.300*	1.419	652,660	0.276	176,069	0.270	92,037	Model23.1.0.a	256.36
0.300*	1.492	654,240	0.275	173,938	0.266	90,100	Model23.1.0.a	259.23
0.300*	1.568	655,845	0.275	172,085	0.262	88,328	Model23.1.0.a	263.39
0.300*	1.649	657,355	0.275	170,532	0.259	86,749	Model23.1.0.a	268.84
0.452	0.607	673,405	0.515	416,043	0.618	298,565	Model23.1.0b	144.63
0.446	0.638	658,845	0.504	396,918	0.602	284,082	Model23.1.0b	144.26
0.440	0.670	642,510	0.483	377,148	0.587	270,404	Model23.1.0b	144.25
0.434	0.705	634,510	0.479	361,847	0.570	256,961	Model23.1.0b	143.64
0.428	0.741	624,665	0.466	345,823	0.554	244,284	Model23.1.0b	143.40
0.421	0.779	615,195	0.454	330,336	0.537	232,235	Model23.1.0b	143.26
0.414	0.819	606,740	0.441	315,592	0.520	220,737	Model23.1.0b	143.21
0.407	0.861	599,275	0.429	301,553	0.503	209,764	Model23.1.0b	143.25
0.400	0.905	592,795	0.417	288,186	0.486	199,290	Model23.1.0b	143.40
0.392	0.951	587,295	0.405	275,456	0.469	189,292	Model23.1.0b	143.65
0.384	1.000	582,790	0.394	263,335	0.452	179,746	Model23.1.0b	144.03
0.376	1.051	579,350	0.382	251,806	0.435	168,994	Model23.1.0b	144.55
0.367	1.105	578,495	0.371	243,038	0.420	158,466	Model23.1.0b	145.19
0.367	1.162	594,405	0.372	259,770	0.437	170,202	Model23.1.0b	144.71
0.360	1.221	595,900	0.364	254,814	0.428	163,101	Model23.1.0b	145.08
0.354	1.284	597,815	0.356	250,051	0.418	156,223	Model23.1.0b	145.48
0.347	1.350	600,160	0.348	245,405	0.409	149,506	Model23.1.0b	145.91
0.340	1.419	602,040	0.341	240,473	0.399	142,939	Model23.1.0b	146.41
0.333	1.492	606,265	0.333	236,288	0.390	137,854	Model23.1.0b	146.91
0.326	1.568	609,185	0.326	231,425	0.380	132,598	Model23.1.0b	147.50
0.319	1.649	614,605	0.317	227,219	0.370	127,351	Model23.1.0b	148.09

Table 12. Eastern Bering Sea Pacific cod catch for 1964-2022.

Year	Catch (t)	Year	Catch (t)	Year	Catch (t)
1964	13,408	1984	125,103	2004	183,748
1965	14,719	1985	143,447	2005	182,940
1966	18,200	1986	135,605	2006	168,818
1967	32,064	1987	149,903	2007	140,129
1968	57,902	1988	203,071	2008	139,802
1969	50,351	1989	178,323	2009	147,174
1970	70,094	1990	172,067	2010	142,844
1971	43,054	1991	210,241	2011	209,201
1972	42,905	1992	164,210	2012	232,623
1973	53,386	1993	133,186	2013	236,691
1974	62,462	1994	172,263	2014	238,718
1975	51,551	1995	228,498	2015	232,829
1976	50,481	1996	209,067	2016	247,620
1977	33,335	1997	232,601	2017	237,851
1978	42,543	1998	158,529	2018	199,867
1979	33,761	1999	145,867	2019	178,904
1980	35,058	2000	151,376	2020	155,665
1981	56,507	2001	142,542	2021	121,749
1982	61,104	2002	166,555	2022	152,146
1983	94,801	2003	175,443		

Table 13. Model 23.1.0.d and 23.1.0.g likelihood profiles over catchability. Light shaded rows are ± 2LL from the MLE, dark shaded row is the closest to MLE.

M	Q	B_0	F_{MSY}	B_{2023}	B_{2023}/B_0	ABC_{2024}	Model	-LL
0.457	0.607	679,640	0.522	425,216	0.626	307,231	Model23.1.0d	133.82
0.452	0.638	665,475	0.510	405,972	0.610	292,310	Model23.1.0d	133.54
0.446	0.670	644,305	0.490	383,401	0.595	278,523	Model23.1.0d	133.75
0.440	0.705	632,835	0.478	366,254	0.579	264,915	Model23.1.0d	133.56
0.434	0.741	622,430	0.466	349,934	0.562	251,929	Model23.1.0d	133.45
0.427	0.779	613,055	0.454	334,398	0.545	239,537	Model23.1.0d	133.42
0.420	0.819	604,685	0.442	319,606	0.529	227,711	Model23.1.0d	133.47
0.413	0.861	597,300	0.430	305,521	0.512	216,424	Model23.1.0d	133.63
0.406	0.905	590,875	0.418	292,103	0.494	205,655	Model23.1.0d	133.89
0.398	0.951	585,460	0.406	279,330	0.477	195,369	Model23.1.0d	134.28
0.391	1.000	581,080	0.395	267,246	0.460	185,600	Model23.1.0d	134.79
0.382	1.051	580,115	0.384	259,479	0.447	178,331	Model23.1.0d	135.41
0.383	1.105	603,225	0.385	286,700	0.475	192,596	Model23.1.0d	134.26
0.377	1.162	603,805	0.377	281,452	0.466	187,133	Model23.1.0d	134.49
0.370	1.221	604,715	0.368	276,276	0.457	181,787	Model23.1.0d	134.75
0.364	1.284	605,975	0.360	271,122	0.447	176,529	Model23.1.0d	135.05
0.357	1.350	607,595	0.352	265,972	0.438	169,713	Model23.1.0d	135.38
0.350	1.419	609,605	0.344	260,820	0.428	162,337	Model23.1.0d	135.75
0.344	1.492	612,035	0.336	255,666	0.418	155,045	Model23.1.0d	136.17
0.337	1.568	614,925	0.329	250,521	0.407	147,842	Model23.1.0d	136.64
0.330	1.649	618,320	0.321	245,392	0.397	141,201	Model23.1.0d	137.17
0.469	0.607	593,295	0.699	425,484	0.717	313,571	Model23.1.0g	142.10
0.463	0.638	578,390	0.680	404,851	0.700	298,556	Model23.1.0g	141.95
0.457	0.670	570,505	0.664	387,512	0.679	283,536	Model23.1.0g	141.46
0.450	0.705	561,410	0.645	370,082	0.659	269,474	Model23.1.0g	141.22
0.444	0.741	552,870	0.627	353,318	0.639	256,106	Model23.1.0g	141.06
0.437	0.779	545,455	0.610	337,364	0.619	243,336	Model23.1.0g	140.99
0.430	0.819	539,170	0.592	322,179	0.598	231,135	Model23.1.0g	141.00
0.423	0.861	529,465	0.578	306,090	0.578	219,879	Model23.1.0g	141.27
0.416	0.905	523,905	0.558	291,697	0.557	208,847	Model23.1.0g	141.56
0.408	0.951	523,590	0.548	279,700	0.534	197,917	Model23.1.0g	141.73
0.400	1.000	525,145	0.521	268,281	0.511	187,520	Model23.1.0g	142.25
0.392	1.051	523,205	0.513	258,255	0.494	179,501	Model23.1.0g	142.89
0.392	1.105	543,845	0.510	284,780	0.524	193,289	Model23.1.0g	142.05
0.386	1.162	545,660	0.499	279,366	0.512	187,809	Model23.1.0g	142.38
0.379	1.221	548,460	0.487	274,431	0.500	182,550	Model23.1.0g	142.74
0.373	1.284	548,815	0.477	268,000	0.488	177,088	Model23.1.0g	143.24
0.367	1.350	555,755	0.463	265,142	0.477	172,450	Model23.1.0g	143.61
0.360	1.419	559,670	0.452	260,441	0.465	167,525	Model23.1.0g	144.14
0.354	1.492	563,400	0.441	255,520	0.454	162,646	Model23.1.0g	144.74
0.347	1.568	568,140	0.430	250,993	0.442	157,951	Model23.1.0g	145.40
0.341	1.649	573,885	0.419	246,846	0.430	151,632	Model23.1.0g	146.12

Table 14. Model 23.1.0.h likelihood profile over catchability. Light shaded rows are ± 2LL from the MLE, dark shaded row is the closest to MLE. *Note the model hit the lower bound for natural mortality at 0.3.

M	Q	B_0	F_{MSY}	B_{2023}	B_{2023}/B_0	ABC_{2024}	Model	-LL
0.470	0.607	666,785	0.542	403,687	0.605	[redacted]	Model23.1.0.h	633.52
0.462	0.638	654,040	0.527	386,153	0.590	[redacted]	Model23.1.0.h	632.70
0.455	0.670	642,440	0.514	369,347	0.575	[redacted]	Model23.1.0.h	632.01
0.447	0.705	630,285	0.505	352,520	0.559	[redacted]	Model23.1.0.h	631.44
0.439	0.741	621,895	0.491	337,432	0.543	[redacted]	Model23.1.0.h	631.00
0.430	0.779	615,225	0.477	323,154	0.525	[redacted]	Model23.1.0.h	630.74
0.421	0.819	610,240	0.462	309,616	0.507	[redacted]	Model23.1.0.h	630.69
0.412	0.861	606,980	0.446	296,768	0.489	[redacted]	Model23.1.0.h	630.87
0.402	0.905	605,490	0.430	284,563	0.470	[redacted]	Model23.1.0.h	631.32
0.392	0.951	605,800	0.414	272,945	0.451	[redacted]	Model23.1.0.h	632.08
0.381	1.000	607,970	0.397	261,850	0.431	[redacted]	Model23.1.0.h	633.18
0.371	1.051	610,170	0.381	250,663	0.411	[redacted]	Model23.1.0.h	634.66
0.360	1.105	616,560	0.365	241,103	0.391	[redacted]	Model23.1.0.h	636.54
0.356	1.162	624,225	0.360	245,238	0.393	[redacted]	Model23.1.0.h	638.05
0.347	1.221	632,205	0.349	241,256	0.382	[redacted]	Model23.1.0.h	640.02
0.339	1.284	641,035	0.338	237,399	0.370	[redacted]	Model23.1.0.h	642.18
0.331	1.350	650,730	0.328	233,657	0.359	[redacted]	Model23.1.0.h	644.55
0.322	1.419	661,315	0.318	230,023	0.348	[redacted]	Model23.1.0.h	647.12
0.314	1.492	672,810	0.308	226,494	0.337	[redacted]	Model23.1.0.h	649.92
0.306	1.568	685,235	0.298	223,063	0.326	[redacted]	Model23.1.0.h	652.95
0.300*	1.649	694,810	0.292	220,578	0.317	[redacted]	Model23.1.0.h	656.26

Table 15. Negative log likelihood and derived quantities for assessed models for models with (Free M) M fit with a uninformative prior, (Fixed M) fixed at 0.386623, and the AIC weighted (Burnham and Anderson 2002) values from the likelihood profiles on survey index catchability between -0.5 and 0.5 for the free M models. Negative log likelihood (-LL), unfished female spawning biomass (B_0), female spawning biomass in 2023 (B_{2023}), and projected 2024 allowable biological catch (ABC_{2024}).

	Free M					Fixed M					Change	AIC Weighted			
Model	-LL	B_0	B_{2023}	B_{2023}/B_0	ABC_{2024}	-LL	B_0	B_{2023}	B_{2023}/B_0	ABC_{2024}	in -LL	B_0	B_{2023}	B_{2023}/B_0	ABC_{2024}
M22.2 old	10875	661,455	249,809	0.378	144,694	10,881	653,795	295,111	0.451	192,152	6	663,556	249,862	0.377	144,781
M22.2 up	18362	694,750	263,189	0.379	141,115	18,405	683,985	332,473	0.486	204,657	43	686,923	279,187	0.406	155,540
M23.1.0.a	251	586,050	205,914	0.351	131,883	253	568,340	246,505	0.434	178,060	2	587,786	209,209	0.356	134,930
M23.1.0.b	143	605,435	314,146	0.519	219,817	144	590,270	274,837	0.466	187,374	1	610,263	314,592	0.516	218,929
M23.1.0.d	133	623,435	343,431	0.551	243,533	134	594,955	276,042	0.464	188,263	1	620,013	333,797	0.538	236,289
M23.1.0.g	141	542,635	331,845	0.612	239,088	143	531,915	264,534	0.497	181,473	2	547,403	330,233	0.603	236,739
M23.1.0.h	631	611,365	313,052	0.512		632	613,550	276,694	0.451		1	616,616	318,377	0.516	

Table 16. Survey catchability (Q) estimates for models fit with (Free M) a non-informative prior on natural mortality and (Fixed M) models fit with a fixed natural mortality of 0.386623.

	Survey catchability with	
Model	**Free M**	**Fixed M**
M22.2 old	0.960	0.772
M22.2 up	0.974	0.683
M23.1.0.a	1.097	0.902
M23.1.0.b	0.822	0.953
M23.1.0.d	0.765	0.972
M23.1.0.g	0.792	1.017
M23.1.0.h	0.808	0.949

9 Figures

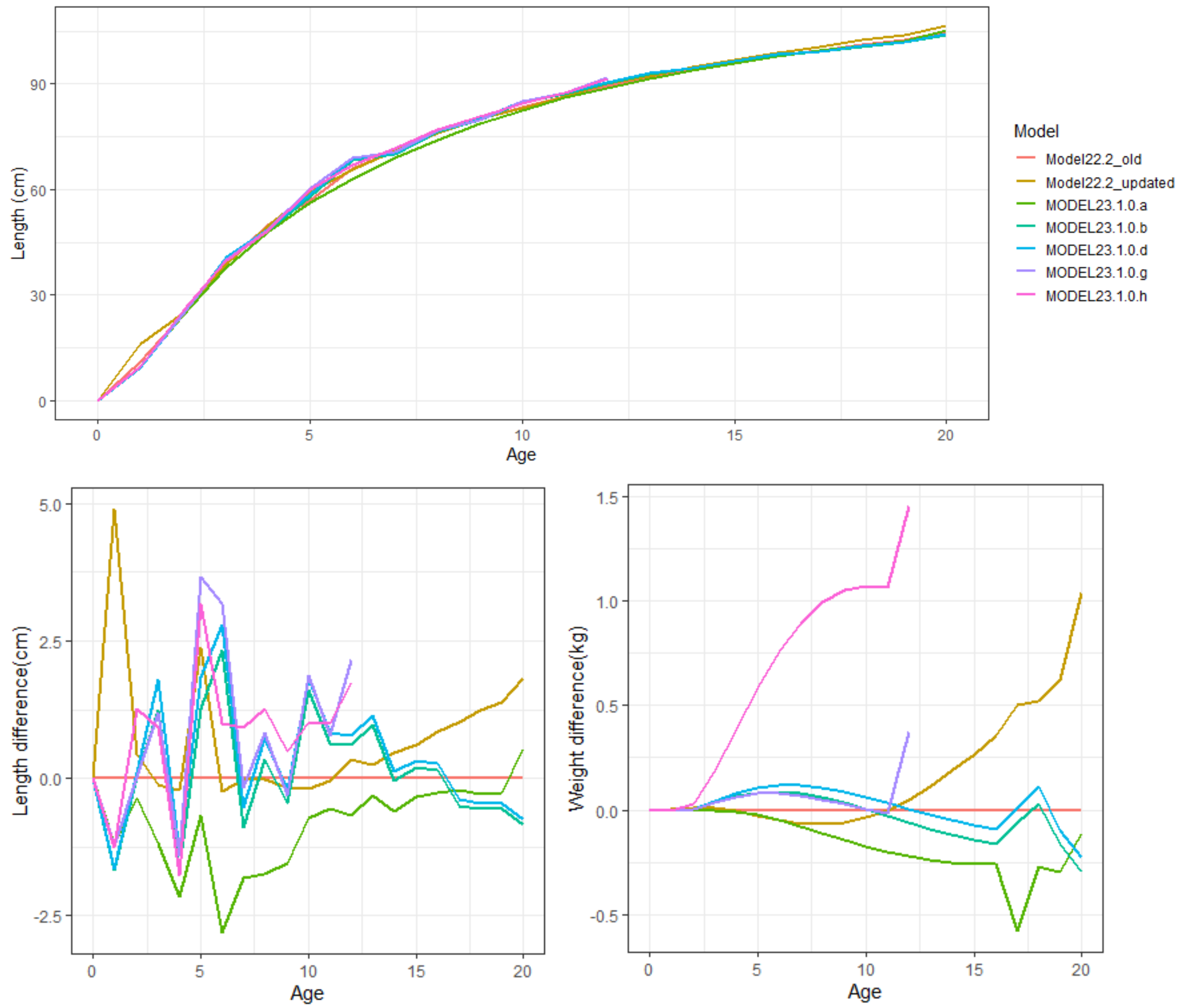


Figure 1. End year (top) length (cm) at age (bottom left) difference in length by age from Model 22.2 old, and (bottom right) difference in weight by age from Model 22.2 old for all models. Please note that the weight difference for Model 23.1.0.h (pink) is in error and should not be considered.

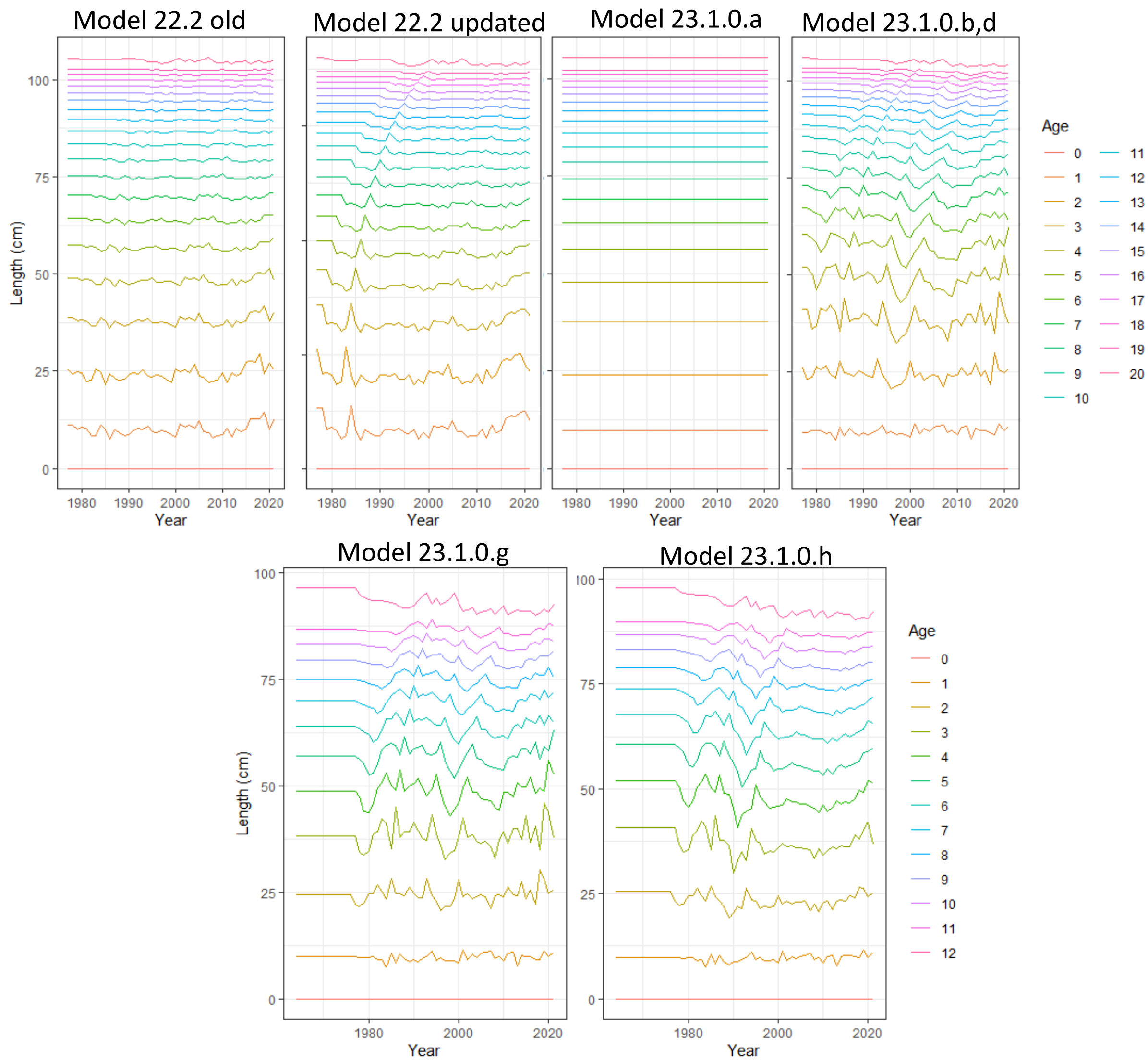


Figure 2. Length at age over time in each of the models examined.

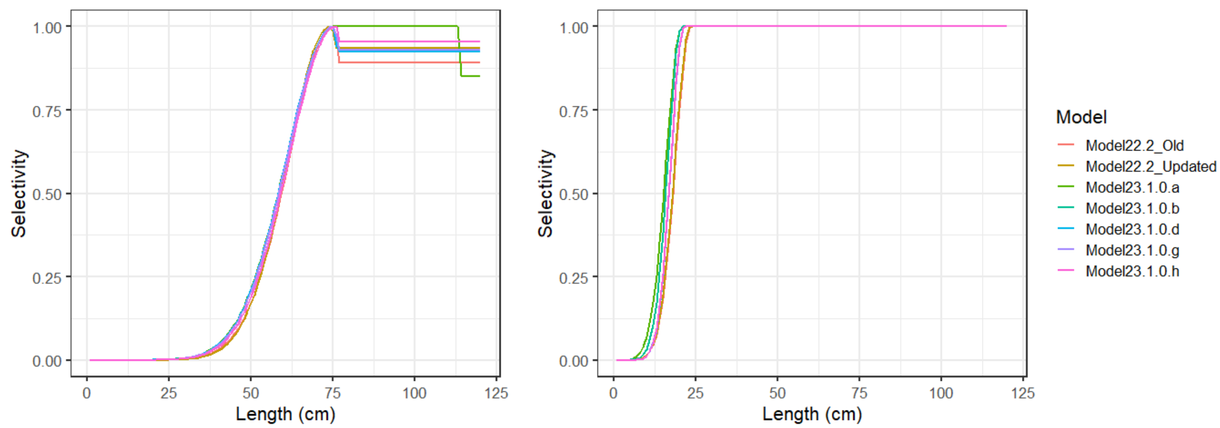


Figure 3. Fishery (left) and survey (right) selectivity for 2022.

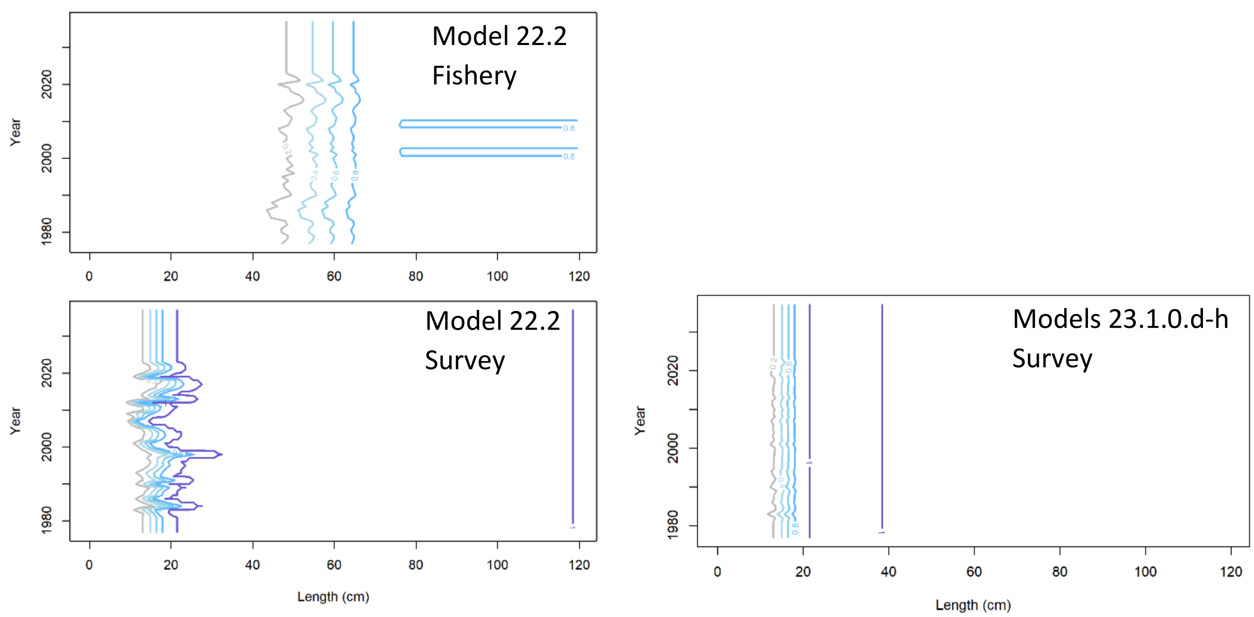


Figure 4. Annually varying selectivity for (top) fishery and (bottom) survey.

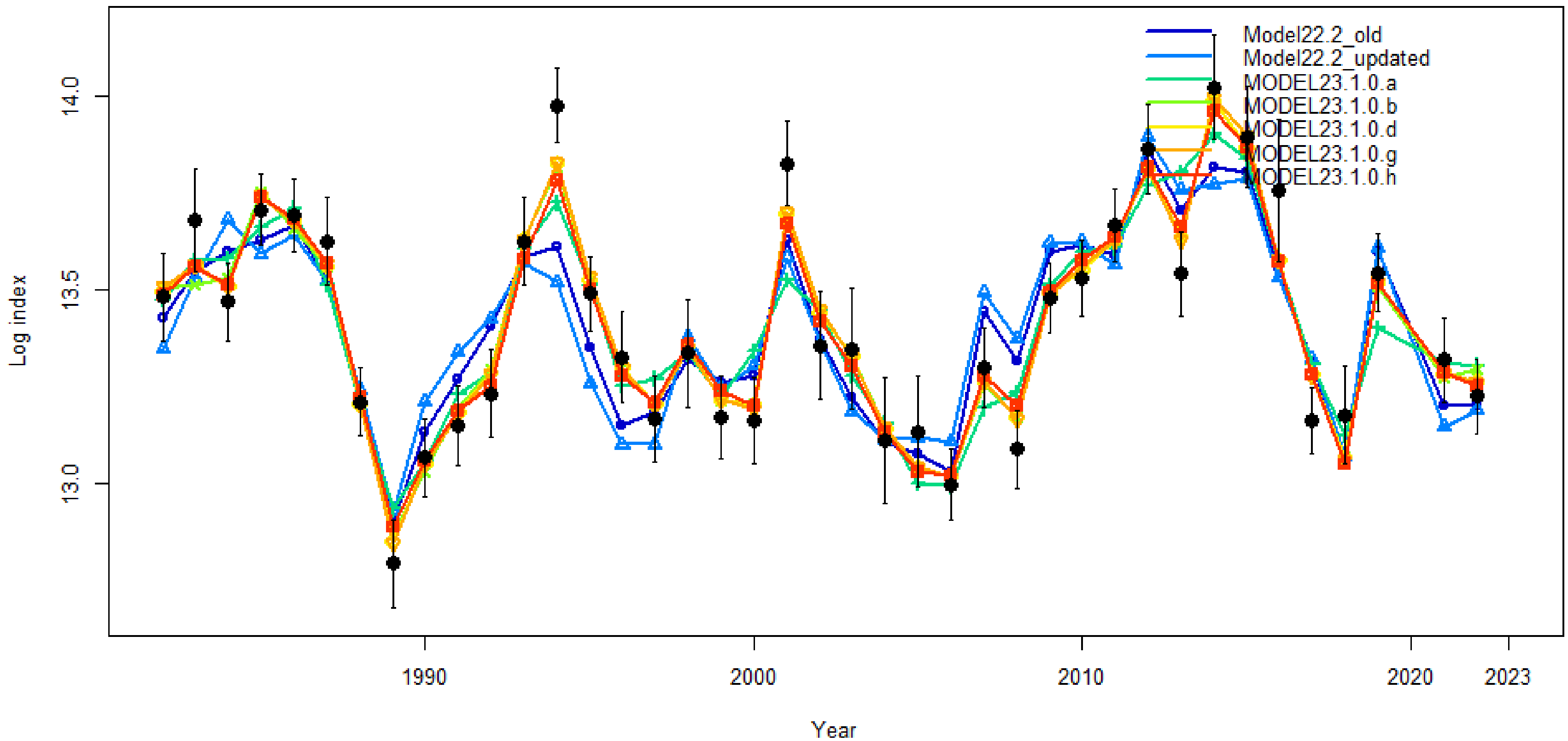


Figure 5. Model fits to the log of the bottom trawl survey index for all models examined.

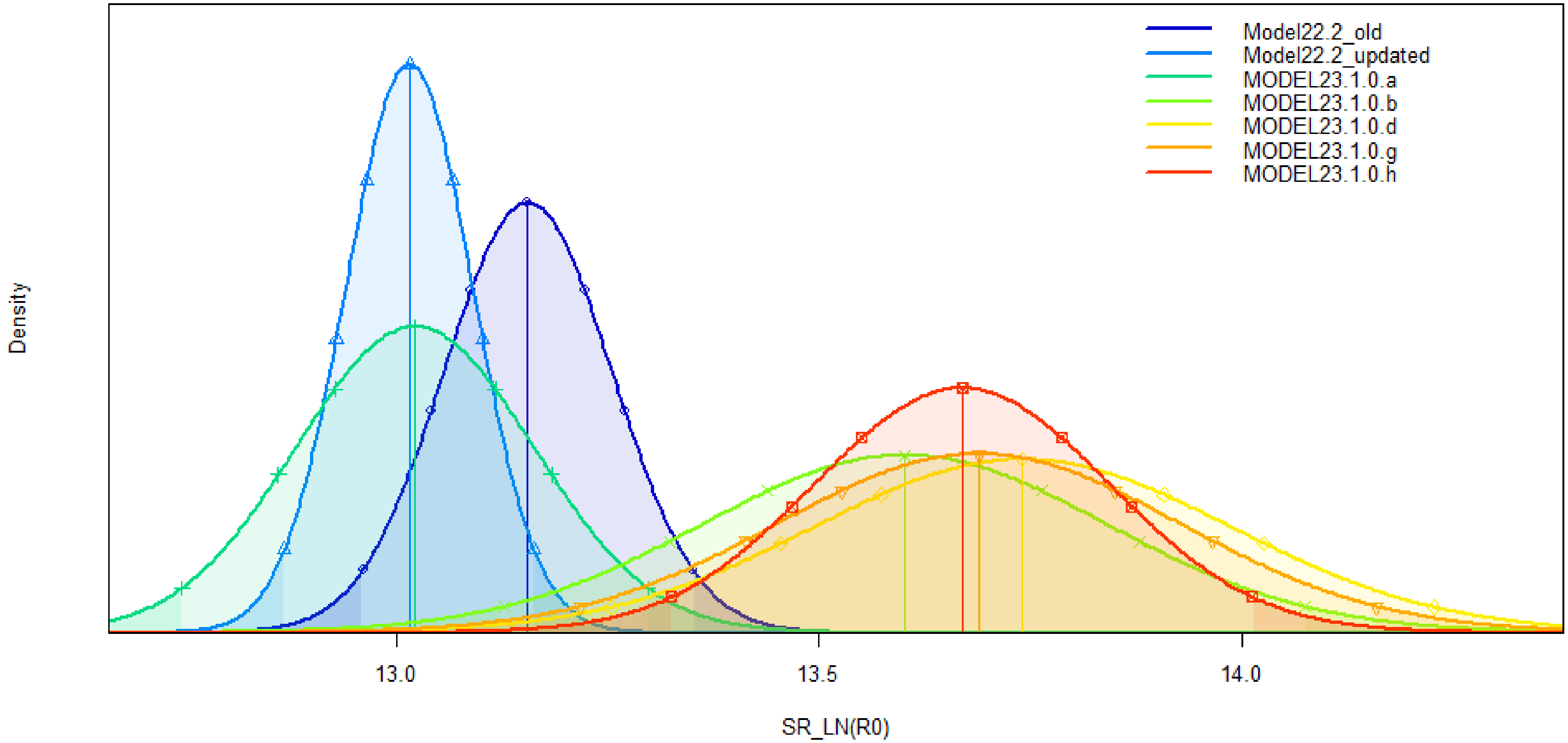


Figure 6. The distribution of the log of virgin recruitment (R_0) for all models.

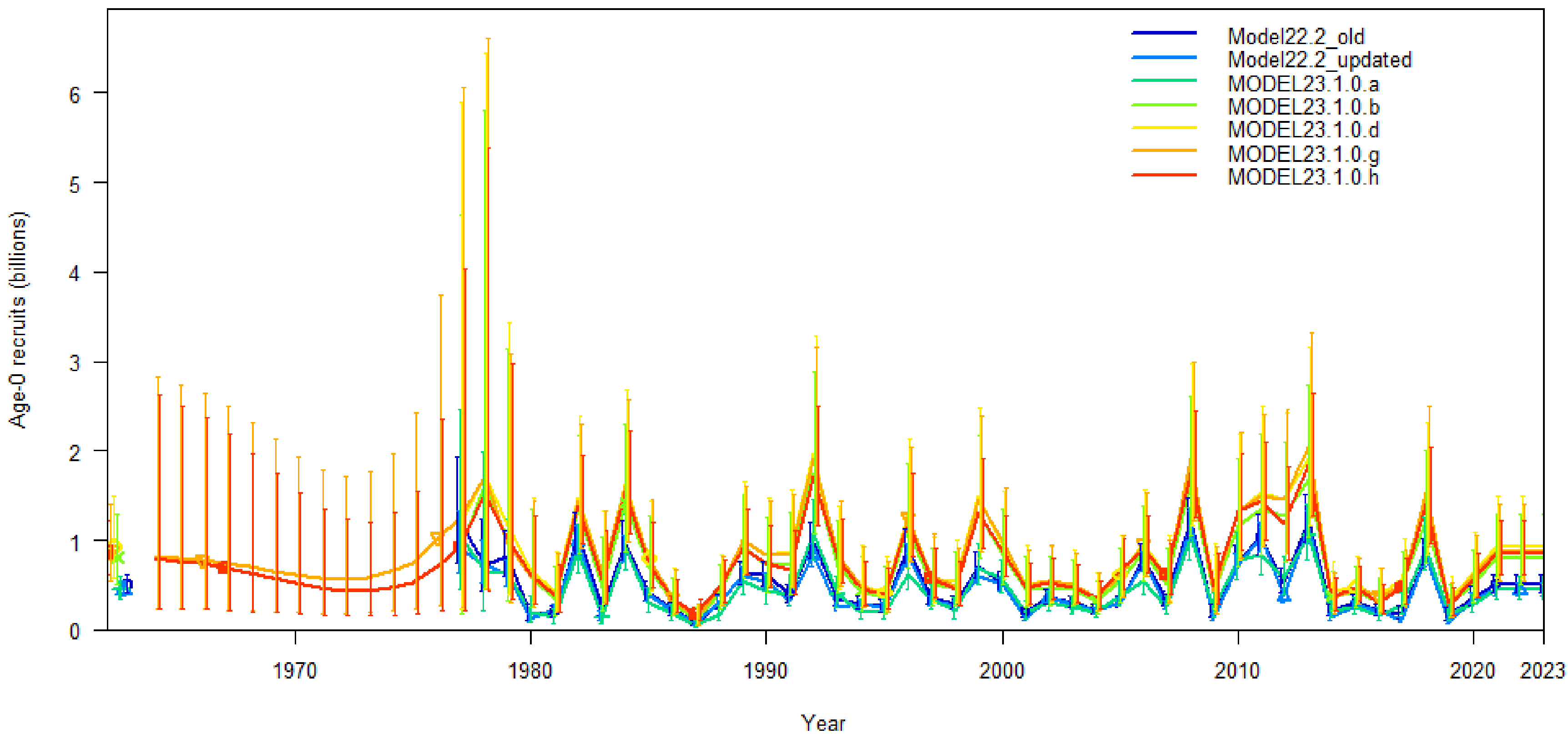


Figure 7. Recruitment in numbers of Age-0 Pacific cod with 95% confidence bounds for all models.

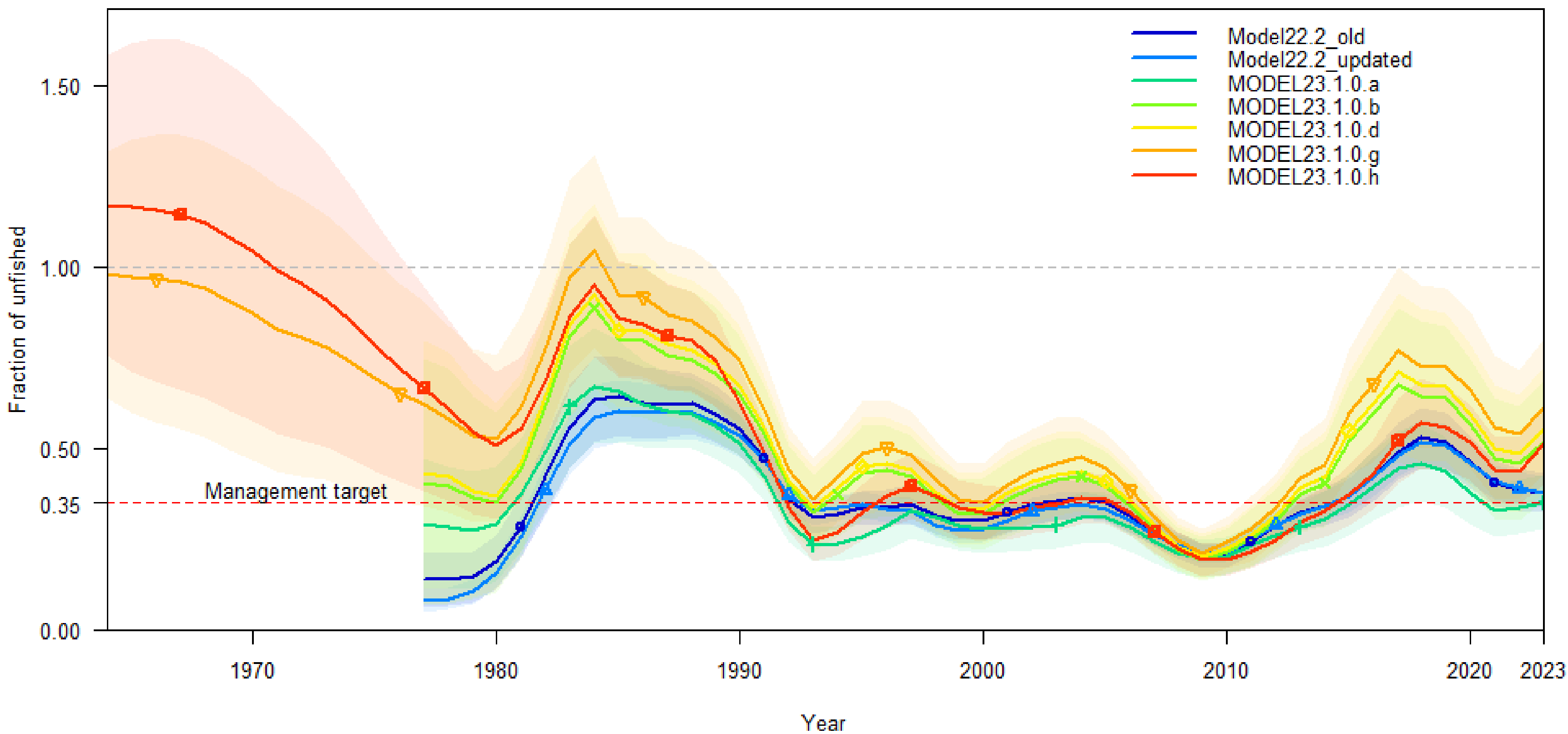


Figure 8. Spawning stock biomass/unfished spawning biomass with $B_{35\%}$ management target and (shaded) 95% confidence bounds for all models.

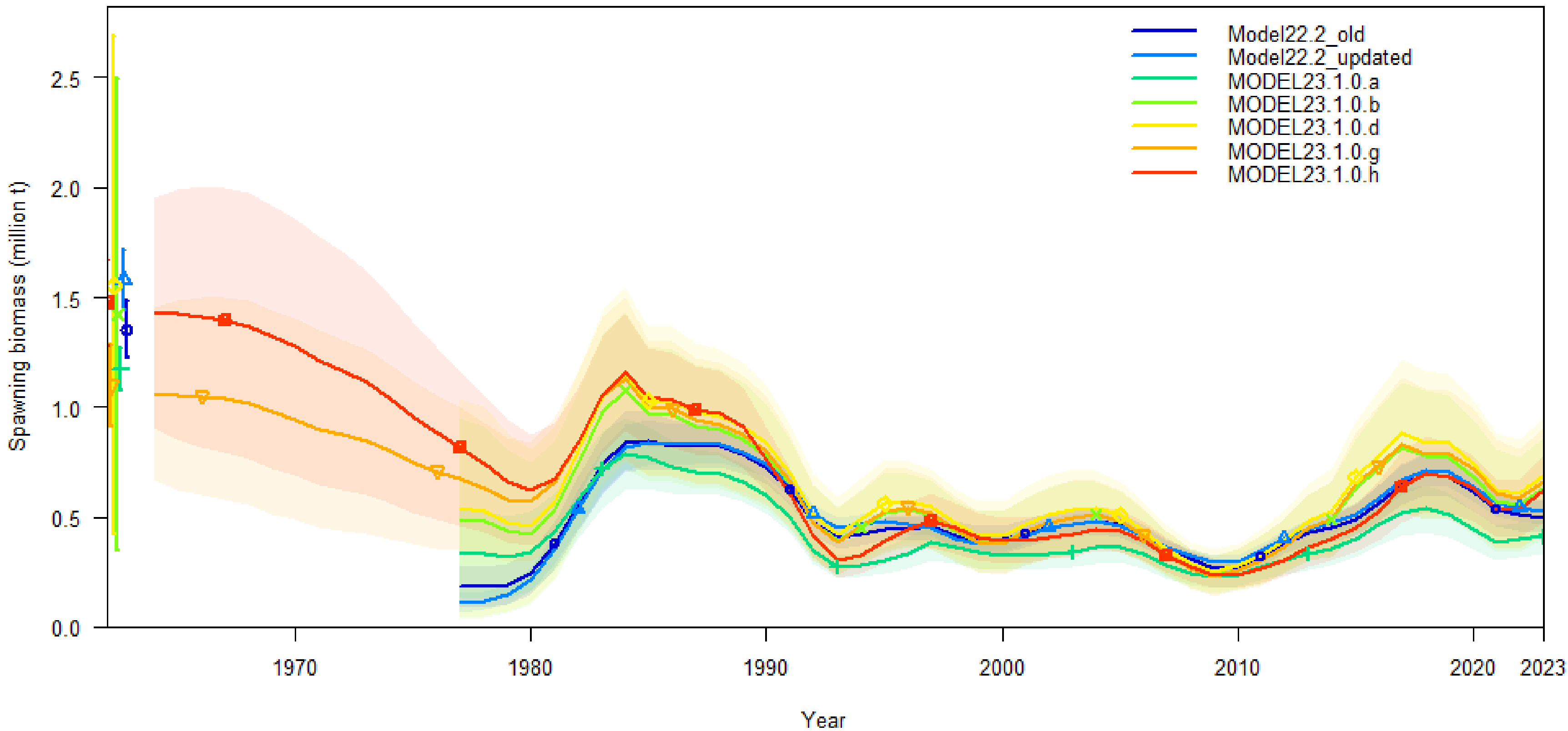


Figure 9. Total spawning biomass (males and females) with (shaded) 95% confidence bounds for all models.

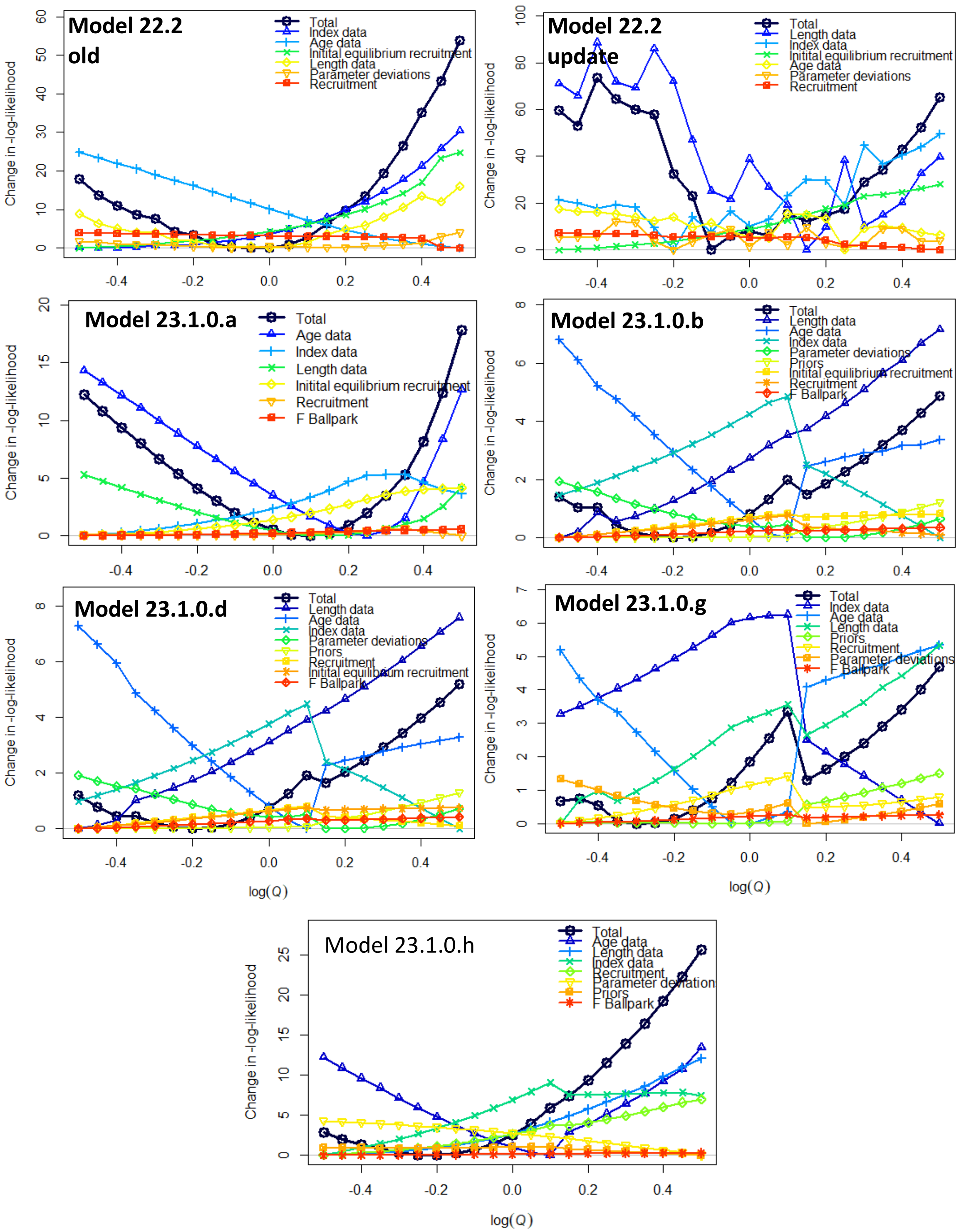


Figure 10. Likelihood profiles scaling the log survey catchability index from -0.5 to 0.5 for the main model components and in total.

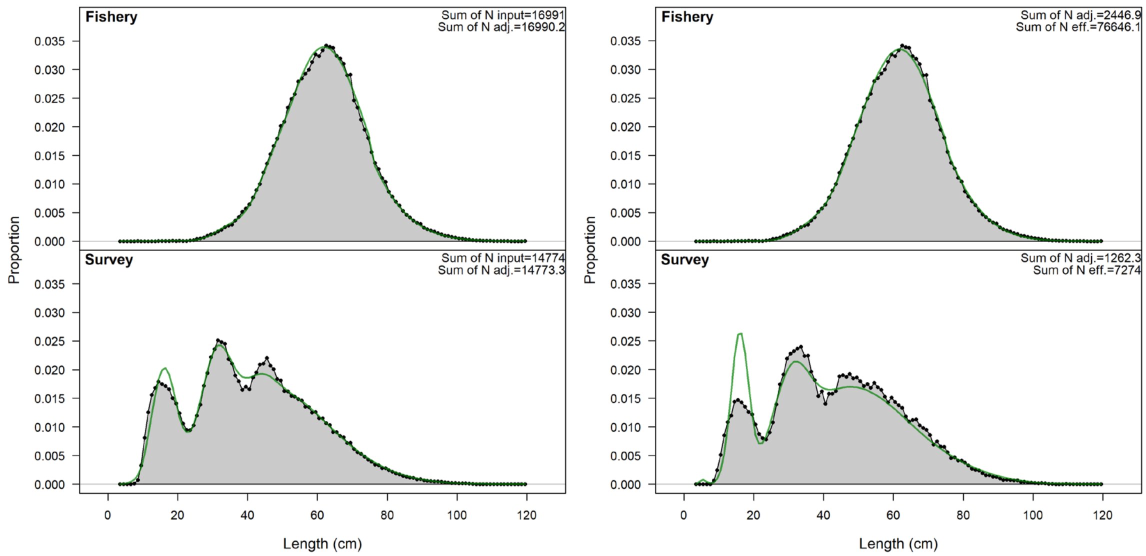


Figure 11. Length comps, aggregated across time by fleet for (left) Model 22.2 and (right) Model 23.1.0a

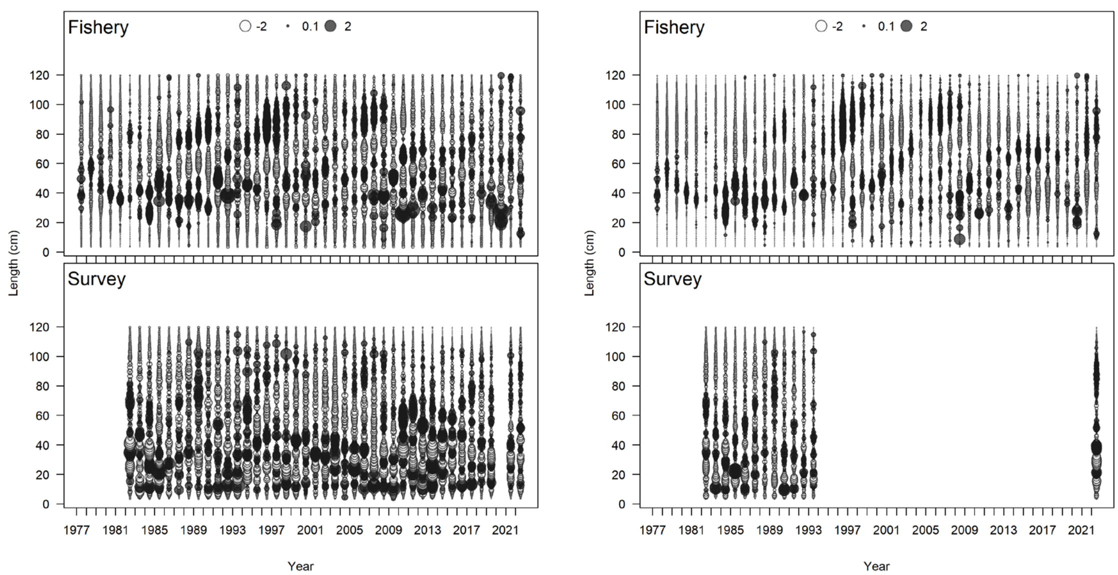


Figure 12. Pearson residuals for length composition, comparing across fleets for (left) Model 22.2 and (right) Model 23.1.0.a. Closed bubbles are positive residuals (observed > expected) and open bubbles are negative residuals (observed < expected).

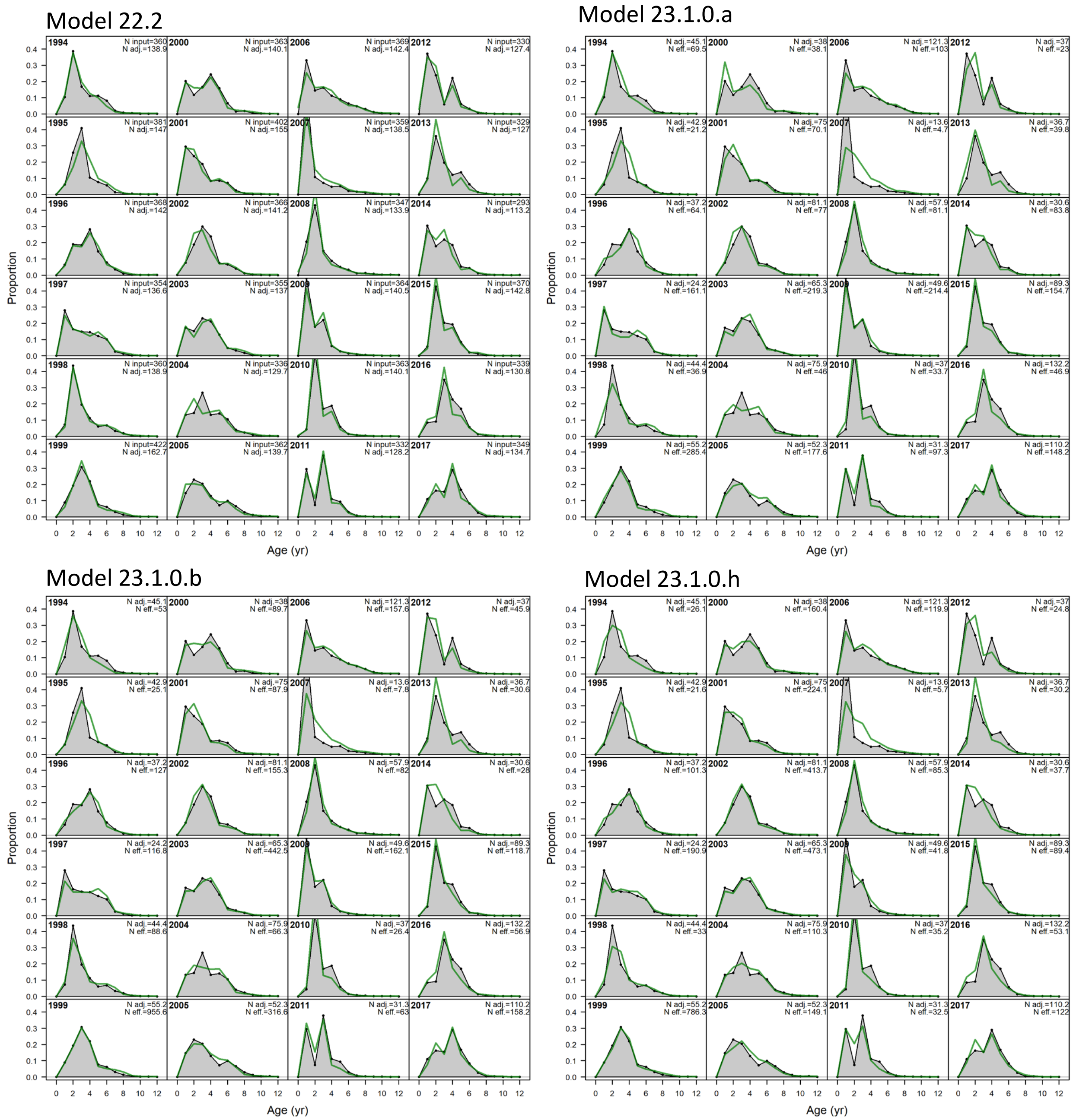


Figure 13. Bottom trawl survey age composition distributions and model fits (green line). Note that models 23.1.0.b, .d, and .g are nearly indistinguishable visually.

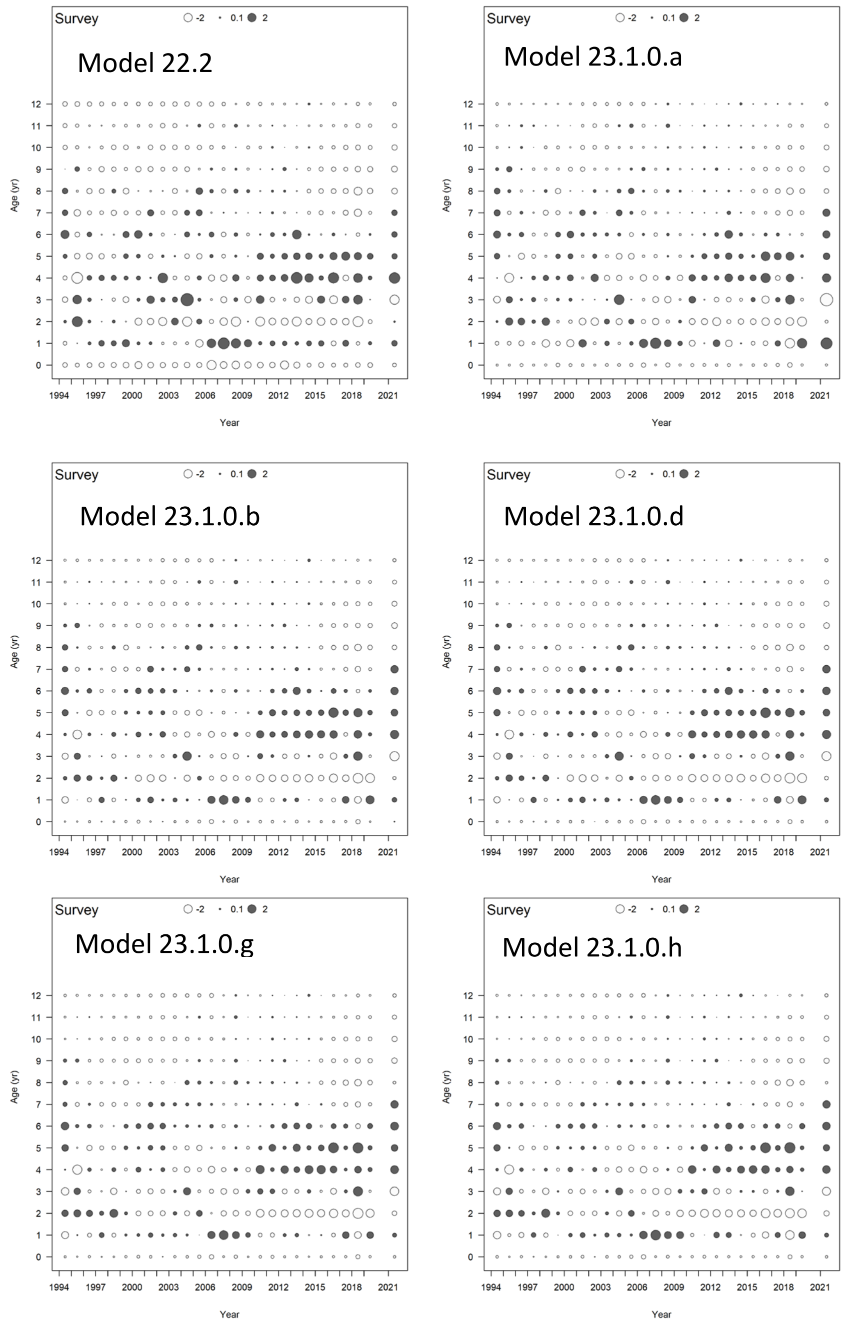


Figure 14. Pearson residuals for survey age composition. Closed bubbles are positive residuals (observed > expected) and open bubbles are negative residuals (observed < expected).

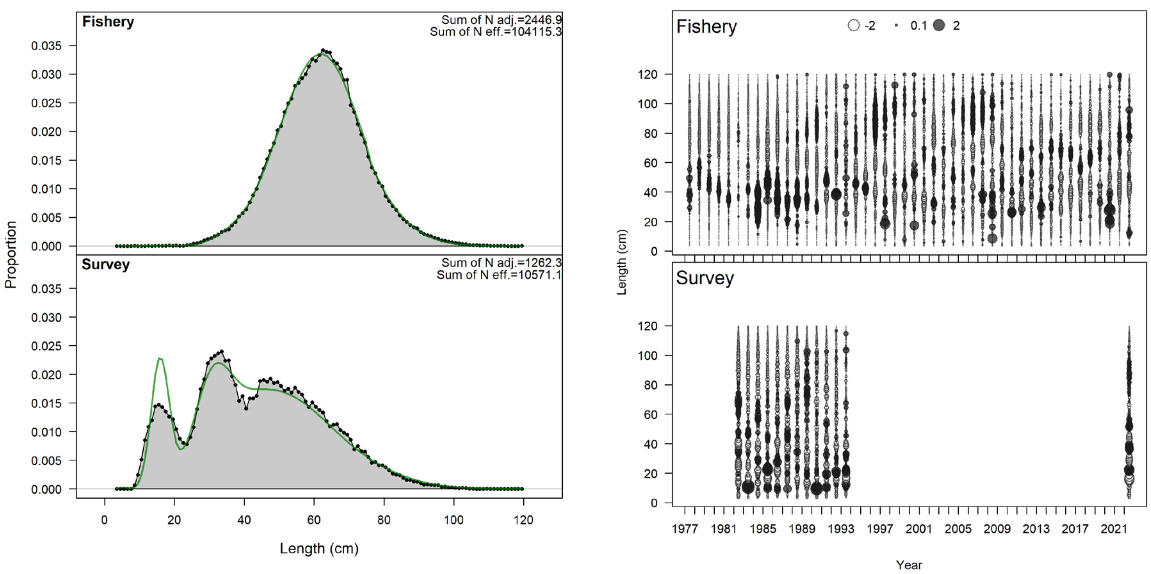
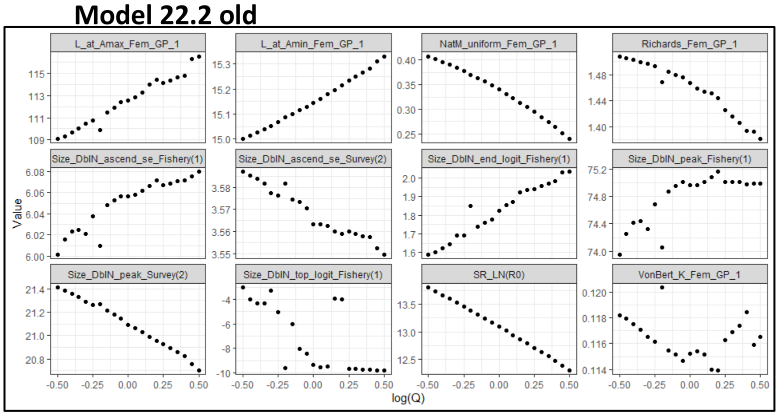


Figure 15. Model 23.1.0.b (Left) length comps, aggregated across time by fleet and (right) Pearson residuals, comparing across fleets for (left) Model 22.2 and (right) Model 23.1.0.a. Closed bubbles are positive residuals (observed > expected) and open bubbles are negative residuals (observed < expected).

Model 22.2 old



Model 23.1.0.b

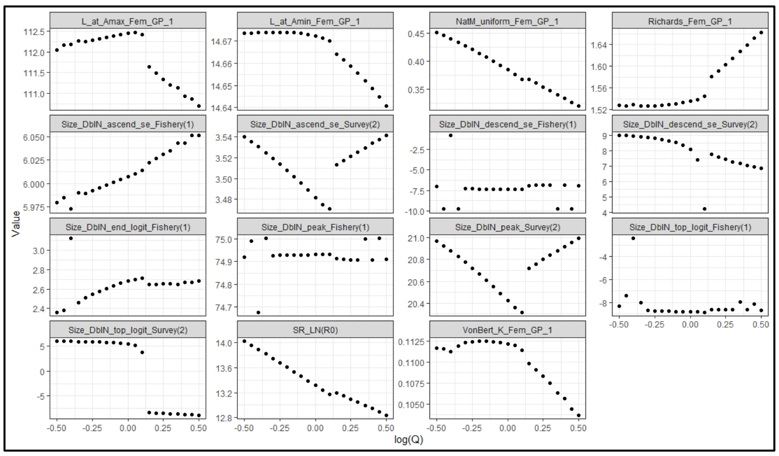


Figure 16. Key parameters fit for the likelihood profile scaling the log survey index catchability from -0.5 to 0.5 for Model 22.2 and Model 23.1.0.b.

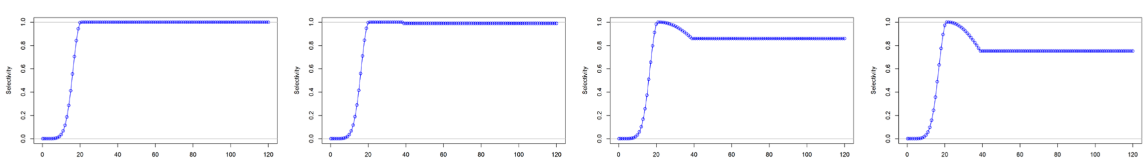


Figure 17. Bottom trawl survey selectivity for Model 23.1.0.b with log catchability fixed at between (far left) -0.25 and (far right) 0.5 showing change to dome-shaped selectivity appearing in the likelihood profile over catchability.

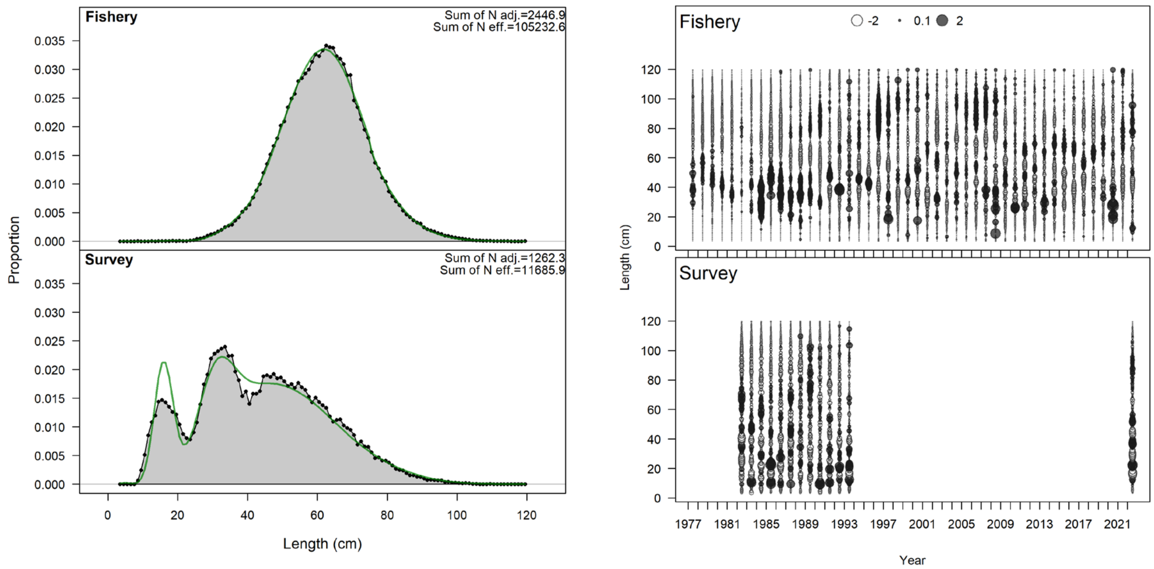


Figure 18. Model 23.1.0.d (Left) length comps, aggregated across time by fleet and (right) Pearson residuals, comparing across fleets for (left) Model 22.2 and (right) Model 23.1.0.a. Closed bubbles are positive residuals (observed > expected) and open bubbles are negative residuals (observed < expected).

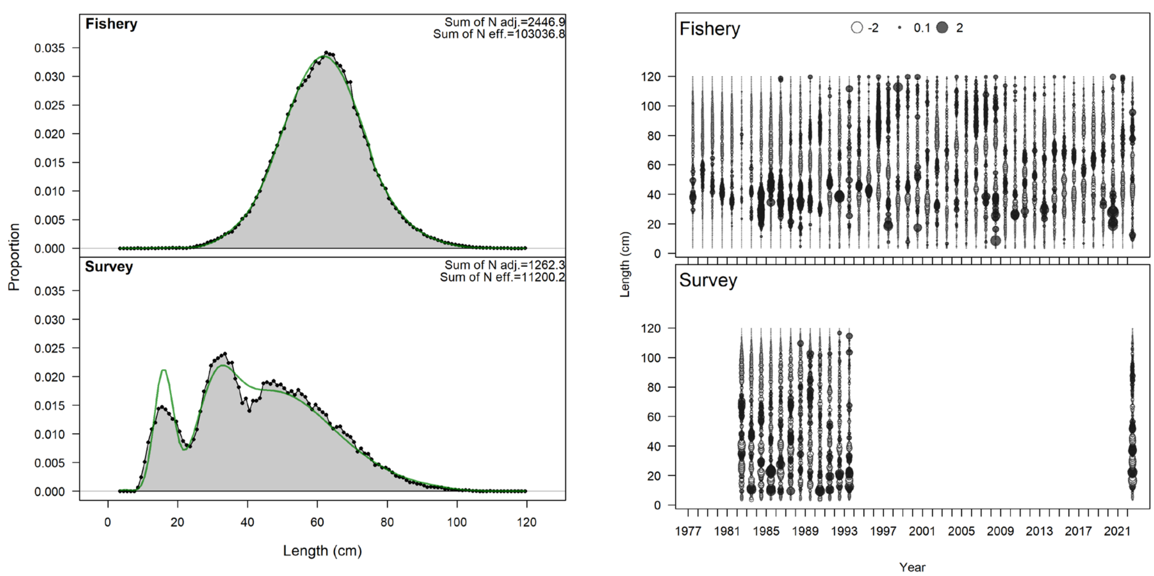


Figure 19. Model 23.1.0.g (Left) length comps, aggregated across time by fleet and (right) Pearson residuals, comparing across fleets for (left) Model 22.2 and (right) Model 23.1.0.a. Closed bubbles are positive residuals (observed > expected) and open bubbles are negative residuals (observed < expected).

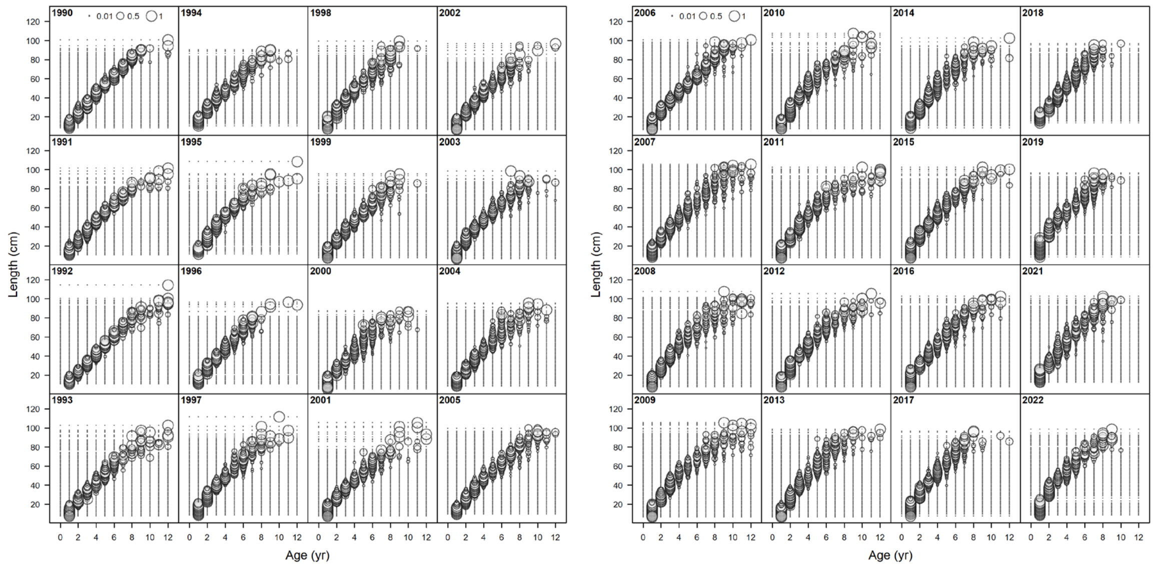


Figure 20. Conditional age-at-length data used in Model 23.1.0.h.

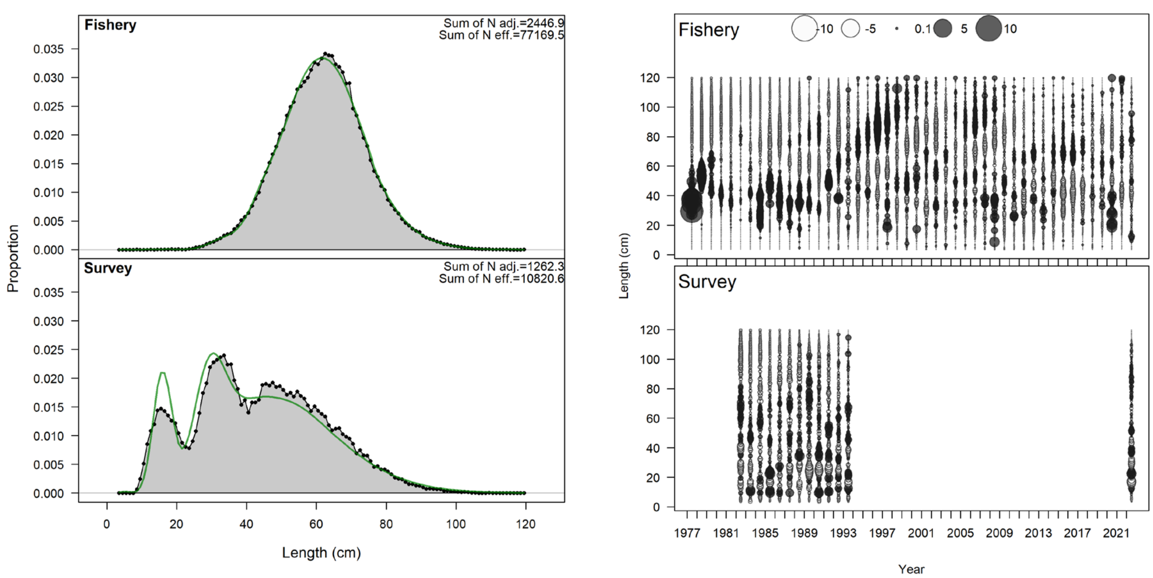


Figure 21. Model 23.1.0.h (Left) length comps, aggregated across time by fleet and (right) Pearson residuals, comparing across fleets for (left) Model 22.2 and (right) Model 23.1.0.a. Closed bubbles are positive residuals (observed > expected) and open bubbles are negative residuals (observed < expected).

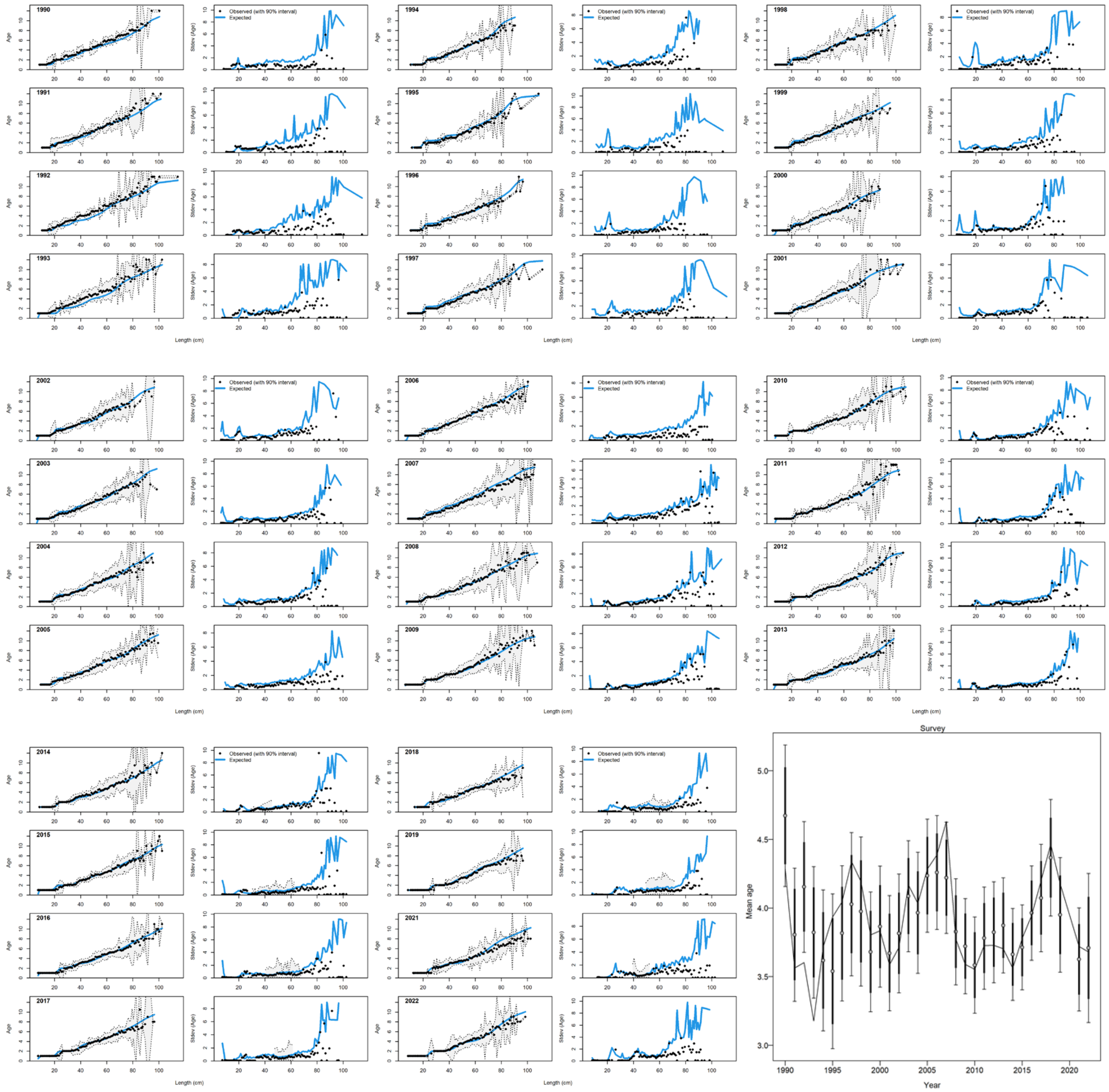


Figure 22. Conditional age-at-length (CAAL) data distributions and Model 23.1.0.h fits to the data including (bottom right) mean age from the CAAL data.

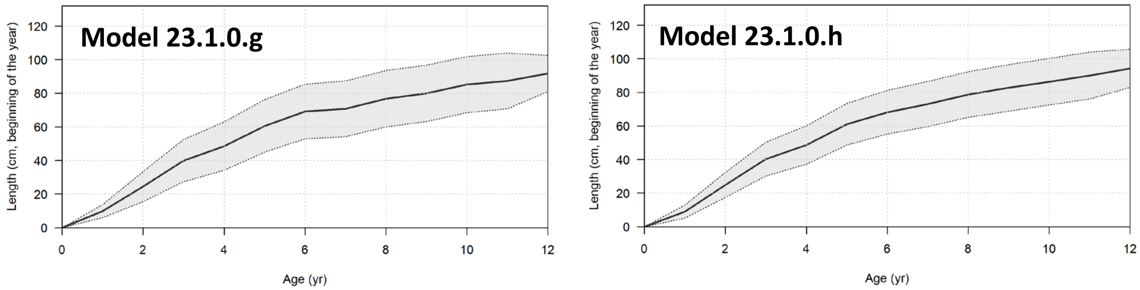


Figure 23. Mean length at age and 95% confidence intervals for (left) Model 23.1.0.g and right (Model 23.1.0.h) showing reduction in uncertainty in growth estimates.

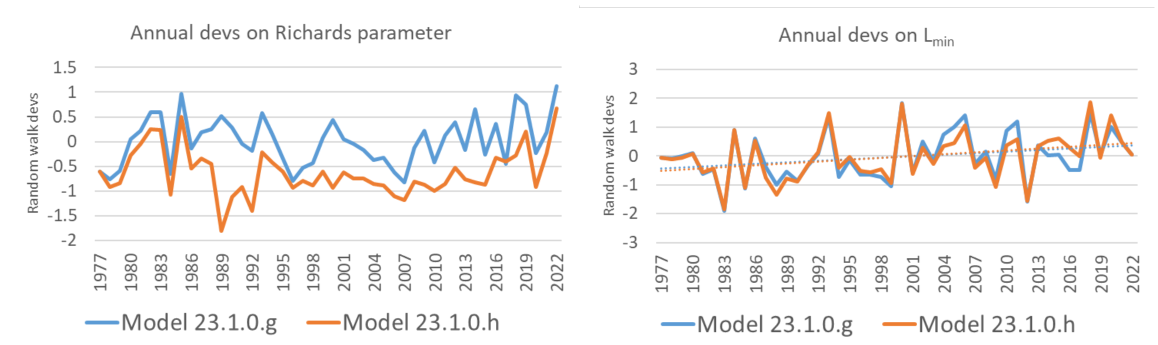


Figure 24. Annual devs on (left) Richards parameter and (right) L_{min} for Model23.1.0.g and Model 23.1.0.h.

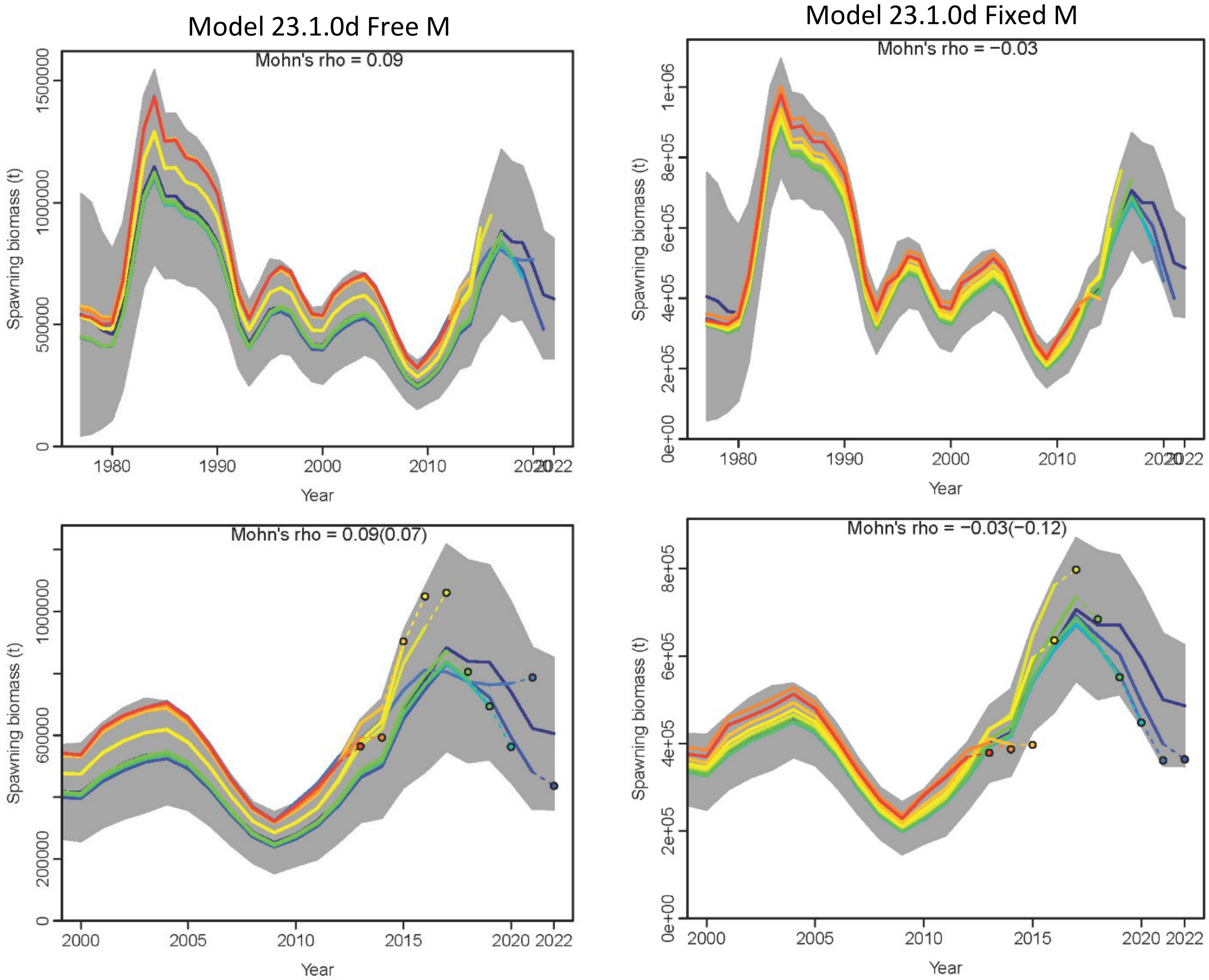


Figure 25 Total spawning biomass from 10-year retrospective peels of Model 23.1.0d with (left) natural mortality fit with an uninformative prior and (right) with natural mortality fixed at 0.387.