Gulf of Alaska Pollock Model Updates

Cole Monnahan

Alaska Fisheries Science Center

September, 2024

Executive Summary

A CIE review was conducted for this stock in May 2024. Several proposed changes to the model resulted from the review. First, input index CVs and age composition input sample sizes were reworked to better reflect interannual variation in information. Second, a covariate link to the catchability for the Shelikof index was formally incorporated (Rogers et al. 2024). Third, the age 1 and 2 indices from the winter Shelikof survey were deemed unreliable and removed from the model. Finally, the Dirichlet-multinomial was tested under the concern that the Francis approach to iterative tuning of effective sample sizes was down weighting the age data unrealistically and to the detriment of the model performance. Model 23c includes all changes except the Dirichlet-multinomial, and σ_R reduced from 1.3 to 1.0. Model 23d is 23c with the Dirichlet-multinomial added. I propose both 23c and 23d for consideration this year. **Overall, I recommend moving to 23d this year.**

Table 1. Management implications for proposed models using data from 2023.

Model	SSB (2023)	B0	B40	B35	FOFL	FABC	OFL (2024)	ABC (2024)
23: 2023 final	274,141	505,000	202,000	177,000	0.307	0.26	269,916	232,543
23c: -Shelikof1&2s	298,600	508,000	203,000	178,000	0.325	0.274	363,464	312,257
23d: +Dirichlet-Mult	292,172	517,000	207,000	181,000	0.316	0.267	307,749	264,903

Proposed models

The GOA pollock ADMB model was converted to Template Model Builder (TMB; Kristensen et al. 2016) in 2023 and named model 23 (Monnahan et al. 2023). TMB extends ADMB's functionality to allow for a state-space formulation, specifically the estimation of process errors by minimizing the Laplace approximation to the marginal likelihood. It can be configured to closely match the ADMB approach of penalized maximum likelihood.

Several minor updates to the model code were incorporated in 2024, which had a negligible impact and thus incorporated in all proposed models below. I added priors on all selectivity parameters (fishery and survey) using an approach called prior pushforward checking (e.g., Monnahan 2024). These so-called regularizing priors stabilize estimation by providing information on the parameters where data are uninformative and would otherwise lead to flat portions of the likelihood. Monnahan (2024) detail the philosophy and process of building regularizing priors in detail for selectivity and growth in stock assessments. The priors are particularly important for retrospective peels for the summer AT survey, which has a short time series, Bayesian integration, jitter analyses, and for self-test simulation. I also incorporated the previous internal projection module into the TMB model, switched to using standard statistical functions like 'dnorm' and 'dmultinom' and added TMB internal simulation capabilities.

Finally, new diagnostic capabilities like jitter analyses, self-test simulation, likelihood profiles, OSA residuals, and standard deviation of normalized residuals (SDNR) were added to the R package. These diagnostics are presented at the end of this document for the proposed models.

The GOA pollock TMB assessment model, an accompanying R package, files to reproduce past accepted models, and the SAFE markdown files are developed on a public repository at https://github.com/afsc-assessments/GOApollock.

The overall approach here is to apply cumulative changes, building to two proposed models that incorporate all the previous additions.

Model 23a: Updating input survey CVs and multinomial input sample sizes

A variety of approaches have been taken to determine the observation errors on the input data, including cases where they are assumed to be constant. Constant age composition input sample sizes ignore interannual variability in the quality of the data and updating these aligns with best practices.

The NMFS bottom trawl and ADF&G bottom trawl indices use index coefficients of variation (CV) that come from a design-based estimator and model-based estimator, respectively, and thus vary over time. I do not propose updating these in 2024. The winter Shelikof and summer acoustic surveys have model-based estimates of CV (Table 1.11 of the 2023 SAFE), but these are deemed too small because they ignore important contributors of uncertainty, and instead constant values of 0.2 and 0.25 are used, respectively. MACE staff have ongoing research on a new “total uncertainty” approach to improve CV estimates for use in stock assessment; however, this research is ongoing and not ready for operational use. As an interim step, I propose to take the current “1D geostatistical” annual estimates of CV and rescale them to have the same average CV as has been used in the past (Fig. 1) .

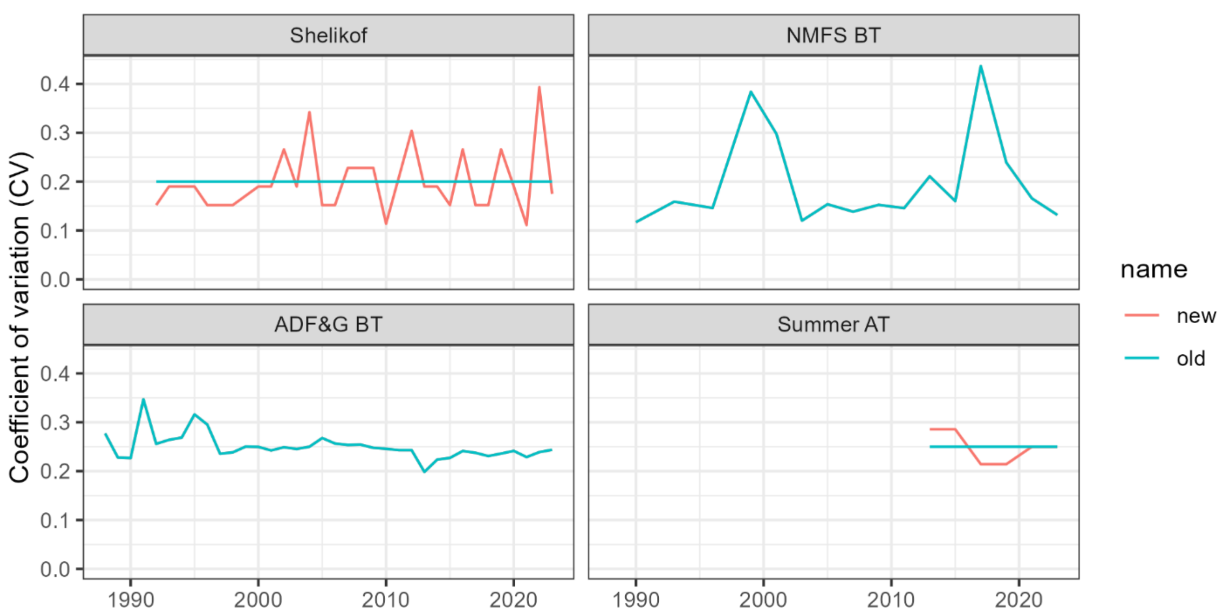

Figure 1. Proposed updates to input CVs for the two acoustic survey indices.

Multinomial likelihoods are used for the age composition data for the fishery and surveys. The input sample size (ISS) is tuned using the Francis (2011) approach to get effective sample size (ESS). The surveys are assumed to have a constant ISS, which again ignores interannual variability in the quantity of information in the age compositions. I therefore propose updating the ISS for the surveys as follows.

For the two acoustic surveys (Shelikof and summer AT) the total number of hauls (midwater + bottom trawl) are used as a proxy for ISS (see Table 1.11 of the 2023 SAFE). The NMFS BT survey ISS was determined using new bootstrap methods (Hulson and Williams 2024).

The fishery ISS historically has been set as the minimum of the number of trips or 200, which recently has led to a constant ISS of 200. The catch at age was estimated using an external Fortran program that stratifies by season and area using age-length key methods described in (Kimura 1989). This Fortran program is custom-built in an outdated version of Fortran and is not sustainable into the future. The AFSC has a similar approach written in ADMB (called “sampler”) that has additional features (bootstrapping, in particular) that are beneficial for this stock. I propose switching to this framework. Preliminary analyses suggest that catch at age (CAA) and weight at age (WAA) match closely between the software platforms (Figs. 2,3).

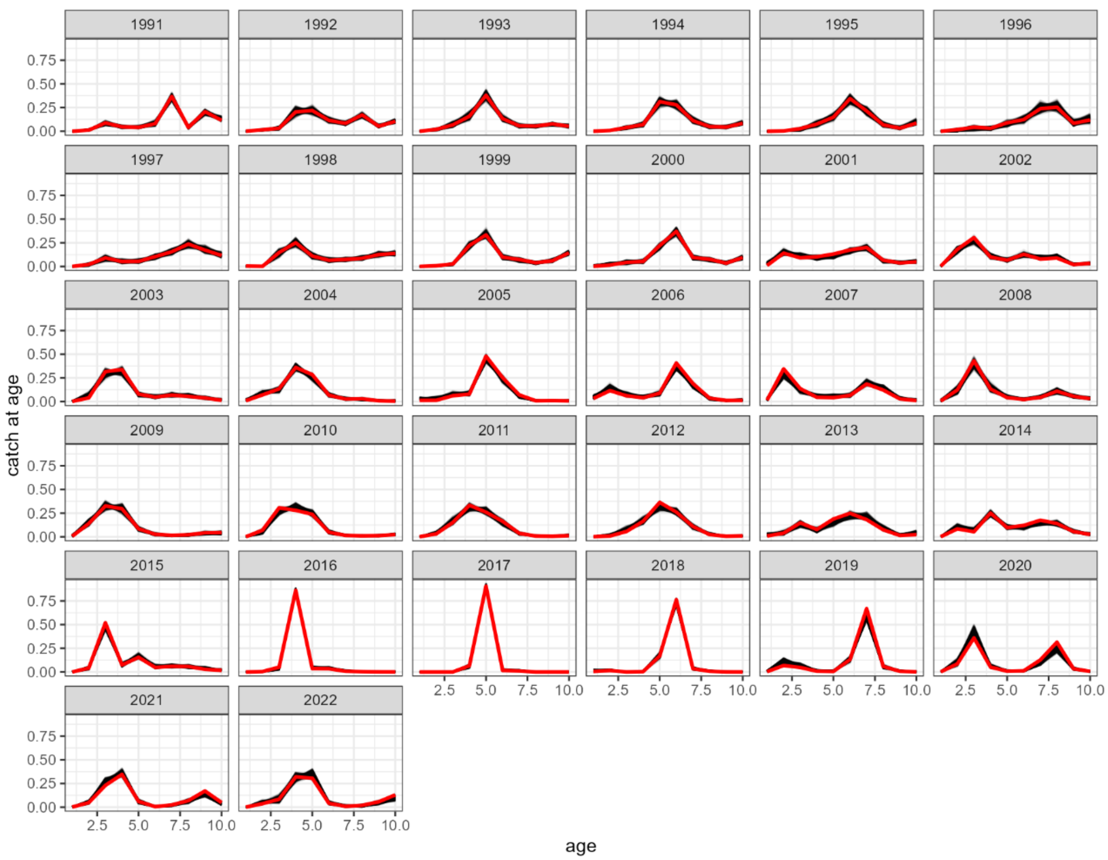

Figure 2. Catch at age estimates using a bootstrap approach (black lines) compared to the historical estimates (red line). Data before 1991 are insufficient to use this approach.

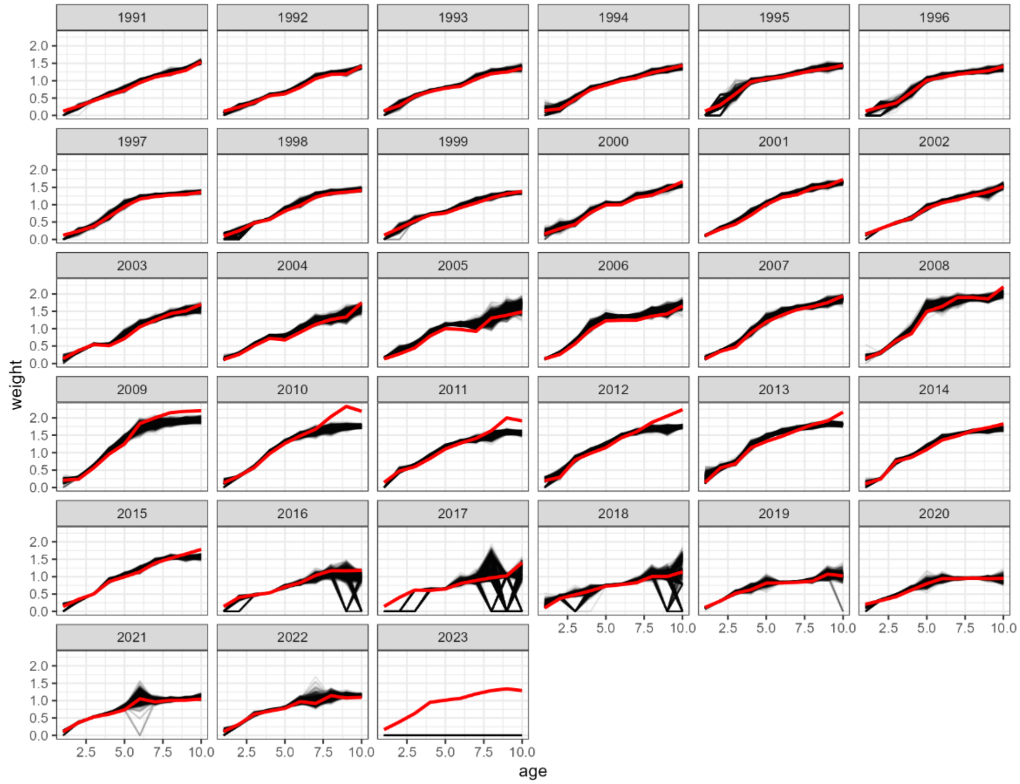

Figure 3. Weight (kg) at age estimates using the bootstrap approach (black lines) compared to the historical estimates (red lines). Note that 2023 had no age data at the time of the analysis and can be ignored. In some cases, bootstraps are missing certain age and year combinations by chance, resulting in estimates of zero that can be ignored.

Bridging from Fortran to ADMB is fairly straightforward due to the similar estimates. For now, the historical CAA and WAA are used, and instead I focus on updating the ISS.

The ADMB sampler also has a bootstrapping procedure that can be used to estimate a realized sample size (RSS) for each bootstrap, similar to the approach taken in bootstrapping survey data to get ISS (e.g., Hulson and Williams 2024). Specifically, the estimated ISS is calculated as the harmonic mean of RSS across the bootstrap samples. This leads to estimates of ISS that vary among years for use in the assessment (Fig. 4).

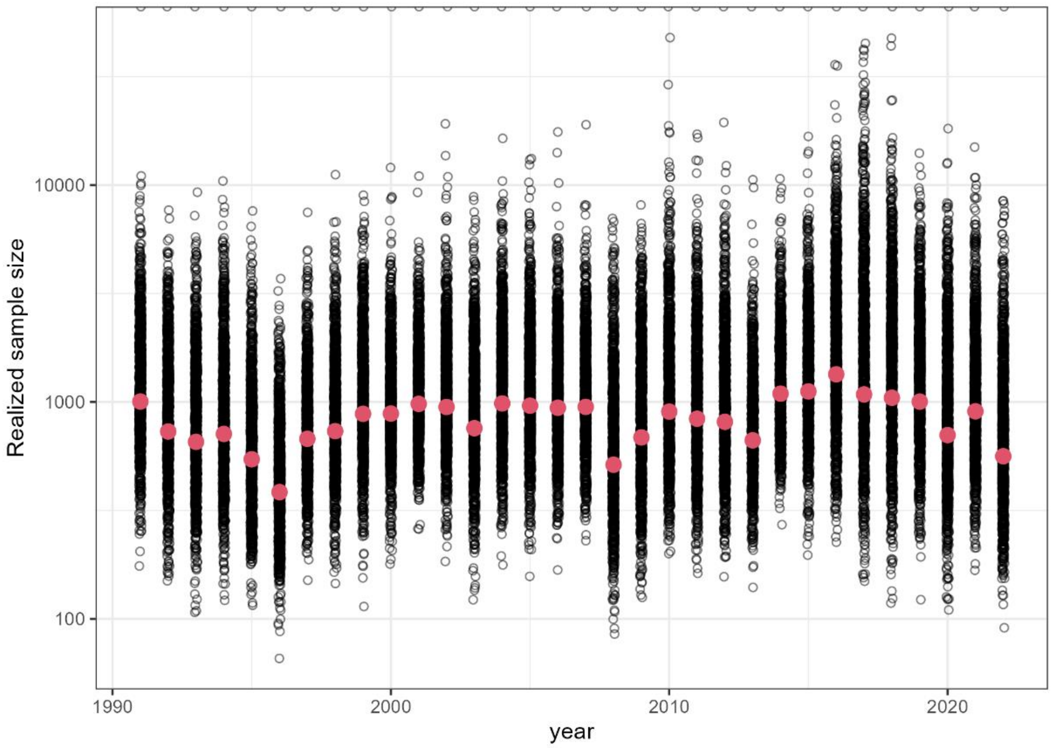

Figure 4. Bootstrap estimates of ISS, called realized sample size (black points) and the harmonic mean (red points) that are used as the ISS in the assessment.

One challenge is that the ADMB sampler approach is only applicable after 1991 due to earlier data limitations. Thus, the new ISS needs to be combined with the old sample size estimates from 1976-1990. I took the new ISS estimates and scaled them to have an average of 200 so that the whole CAA time series had a similar overall weight. Composition data were then tuned via the Francis approach (Fig. 5).

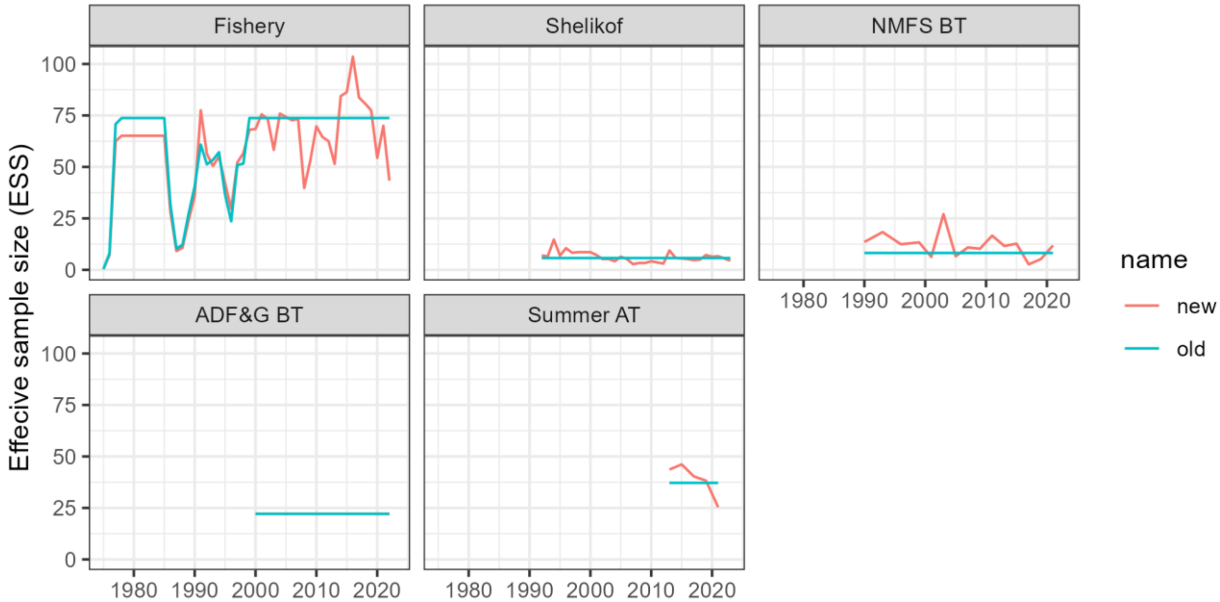

Figure 5. Effective sample sizes before and after (colors) updating the ISS and re-weighting using the Francis approach. No changes are proposed for the ADF&G BT survey this year.

Best practices for developing fishery ISS is ongoing, but moving to the ADMB sampler this year is an important improvement over the status quo.

Model 23a incorporates all proposed changes to the input CVs and ISSs. These were presented to and supported by the 2024 CIE reviewers. Overall estimates of SSB were similar, with slightly lower estimates in recent years and smaller uncertainty (Fig. 6).

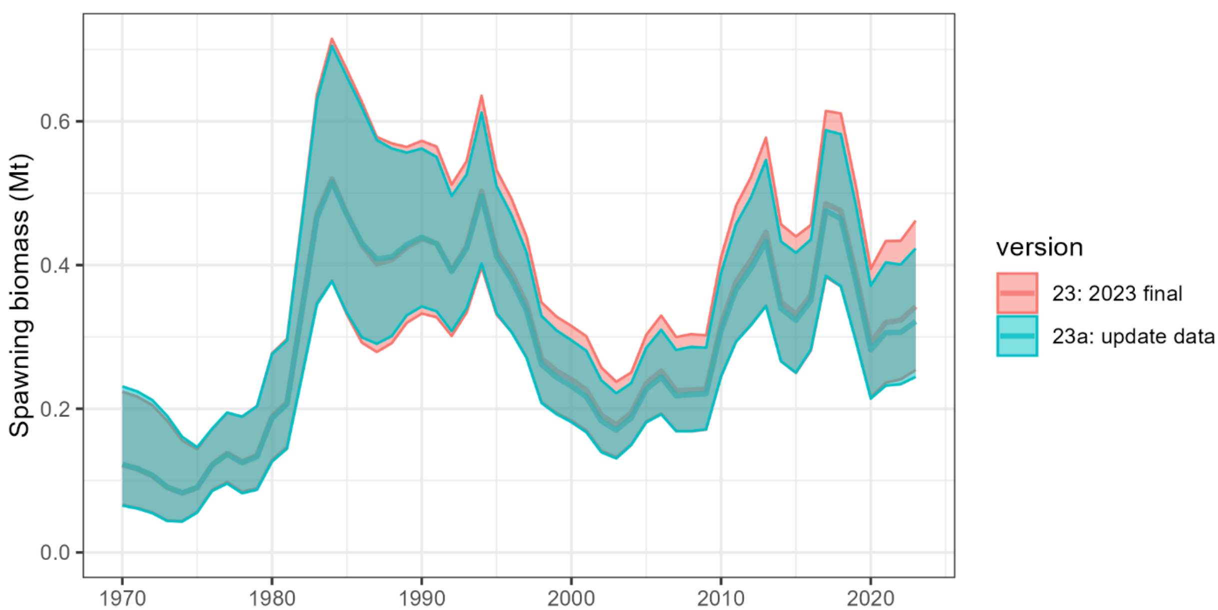

Figure 6. Spawning stock biomass (SSB) estimates between model versions.

Model 23b: Impacts of climate-driven changes in spawn timing

Rogers et al. (2024) found strong statistical evidence of variation in spawn timing on Shelikof catchability using the state-space WHAM modeling platform (Stock and Miller 2021). They found two distinct covariates, logit proportion mature and a metric of mismatch between spawn and survey timing, could both improve fits to the Shelikof index of abundance. This research has been discussed at past Plan Team meetings and has been positively received. Notably, their model helps explain model misfit to the large estimates of biomass from 2016 to 2019, at a time when the bottom trawl surveys were decreasing. In previous years these divergent trends led to a reduction in the ABC via the risk table.

The research is now published and was reviewed and supported by the CIE reviewers in 2024. I therefore propose it for operational use starting in 2024. The proportion mature covariate was selected for operational use despite the better performance of the 'mismatch' one due to its consistent availability. Specifically, whenever there is a Shelikof survey estimates of visual maturity will be available in the same year and can be provided by MACE as part of the survey product pipeline. If no survey is completed the assessment model will smooth over the covariate point, but it will not be used in the assessment as there would be no index to fit to.

State-space methods are required to integrate the estimation of the latent covariate (including estimation of missing data) internal to assessments, and hence only practically feasible for TMB models. The operational model is not in WHAM, so functionality to smooth a time series and link it to catchability were added to model 23. Observed data through 2023 were also added to the model and results were very similar to those in Rogers et al. (2024). There are three important decisions to make to implement this approach. The first is how to account for observation error in the observed covariate. I assumed a small observation error (CV=0.02) on the covariate to force it to fit the observations very closely, but estimating it or deriving estimates from the maturity estimates could be explored in the future. I do not anticipate the model being sensitive to this assumption. Second, the type of time-series to assume for the covariate. I assumed an AR(1) process and estimated the process error and correlation. Third, whether, and how, to add additional temporal variance on catchability. Historically log catchability was modeled as random-walk process with a strong penalty. Rogers et al. (2024) discuss how the covariate accounts for temporal variation and that a second unstructured approach can then attempt to account for spatial variation caused by some spawners going to alternative spawning sites. The random-walk component was left on, although it had a minimal effect because the penalty was left at the small value (when estimated it overfits the index) and the covariate explains most of the variation in catchability. This assumption can be revisited in the future, and potentially replaced with a more sophisticated approach that uses MACE's biomass estimates from areas outside of Shelikof to inform variability in spatial availability to the Shelikof survey.

Table 2. Parameter estimates for the covariate link on catchability for the Shelikof survey 3+ biomass index. SE=standard error and CI= 95% confidence interval.

Parameter	Purpose	Estimate	SE	Lower CI	Upper CI
Rho	AR(1) covariate correlation	0.445	0.150	0.151	0.740
SD	AR(1) covariate process error	1.020	0.145	0.737	1.310
Beta	Covariate effect on log catchability	0.339	0.036	0.268	0.409

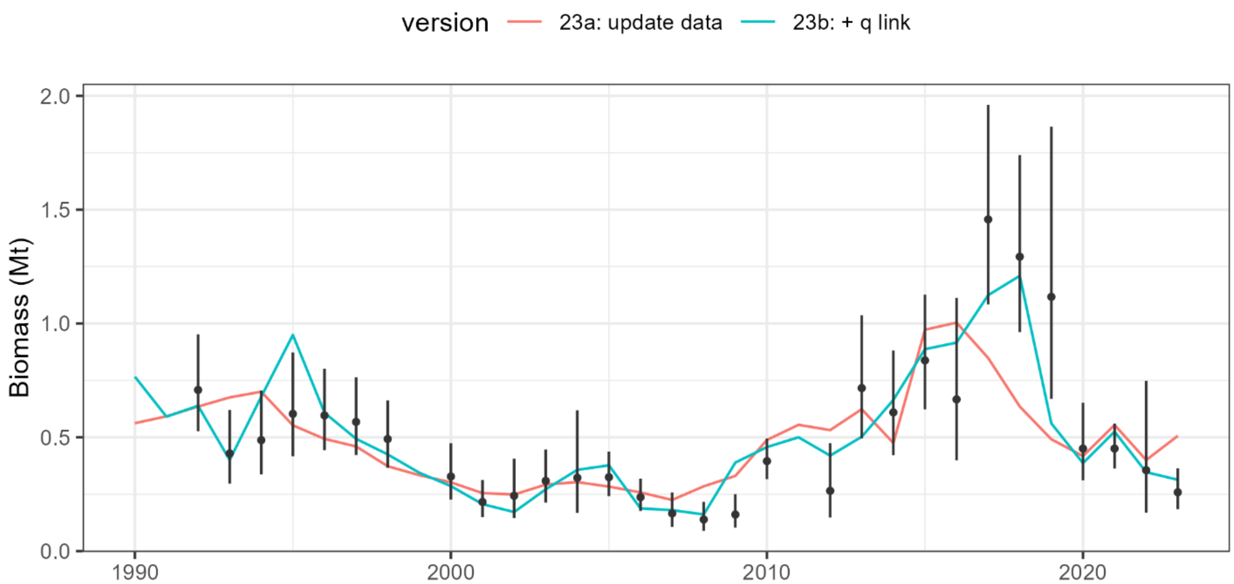

Figure 7. Model fits (colored lines) to the age 3+ index of abundance (points and lines).

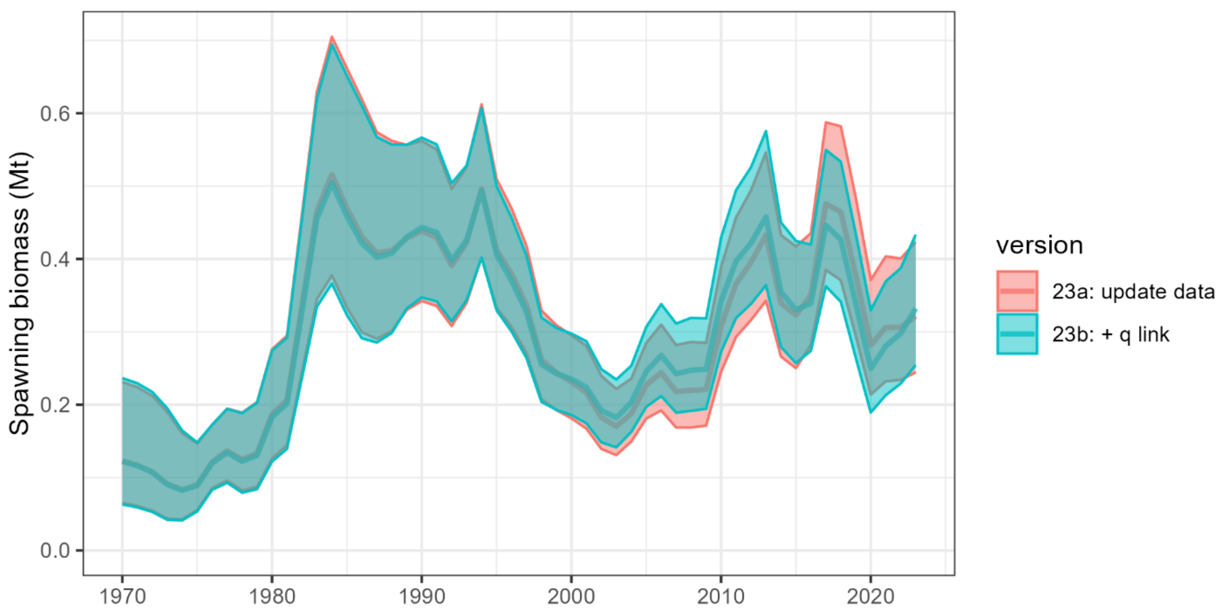

Figure 8. Comparison of SSB between models.

Model 23c: Drop Shelikof age 1 and 2 indices

The 2023 risk table raised the following issue: “A new phenomenon emerged in the last handful of years that is worth highlighting. Many of the estimates of recent cohort sizes are abnormally small compared to previous estimates" (Monnahan et al. 2023). They also noted that these cohorts have an outsized impact on the estimate of variability in recruitment and influence perception of recent changes in productivity. There is no doubt that some recent cohorts are small, as they have been corroborated by subsequent observations of them as older fish. However, the question is whether the estimated cohort sizes more than 7 standard deviations less than average are realistic.

Model exploration determined that these small estimates are driven exclusively by the age 1 and 2 indices from the Shelikof survey. For example, the 2023 estimate of age-1 fish from the Shelikof survey was 52,138 fish. Since this data point provides the primary information about the cohort, model 23 estimated the cohort size of about 1.3 million age-1 fish, which is about 775 times smaller than an average cohort size (as estimated by R0 of 1.02 billion) and the lowest recruitment estimate in the time series by far, despite not fitting the data point well (Fig. 9).

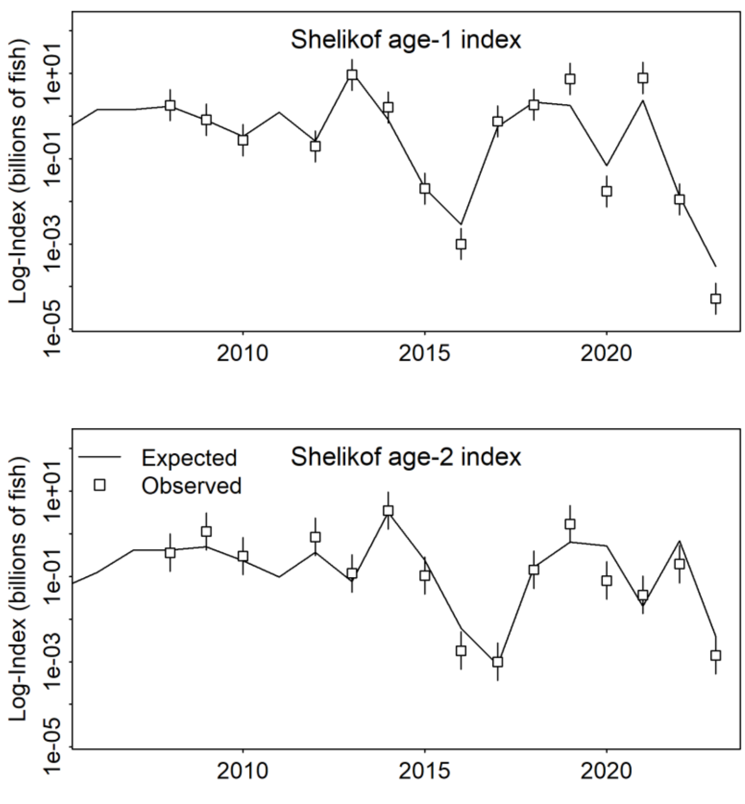

Figure 9. Fits to the Shelikof age 1 and 2 numeric indices from the 2023 accepted model.

These age classes were split off from the 3+ age composition data and modeled as independent numeric indices (billions of fish) based on a CIE review in 2017. Input CVs are 0.45 and 0.55 for age 1 and 2 fish, based on a historical iterative tuning exercise, respectively. One reason for this large uncertainty is that Shelikof is a pre-spawner survey and the immature fish are not expected to consistently migrate into the survey region. Maturity is estimated at 0.000 for age 1 fish and 0.015 for age 2. Thus, small estimates of age 1 and 2 fish could indicate a small cohort, or simply that they were spatially unavailable to the survey in that year. Subsequent surveys when those cohorts are mature can help differentiate between these two scenarios, and historically that seemed to work well. This paradigm does not work well when survey estimates are very, very small, because age compositions in future years will not be able to distinguish between reasonably small and unreasonably small cohort sizes, by nature of the data. Thus, there is no good mechanism in the assessment currently to update knowledge about an exceedingly small cohort estimated from the Shelikof age 1 and 2 data.

These indices are valuable in the stock assessment because they provide the best early signal of cohort sizes, and thus are particularly valuable when estimates are large. Thus, in some sense, we want to trust the large values but not necessarily the very small ones. Throwing out years with small estimates would not be conservative and is not a tenable solution. Several ideas were discussed at the CIE review, such as more flexible catchability options or input CVs that scale non-linearly. Time did not allow for investigating these alternatives during 2024, instead the most straightforward approach in the interim is to remove these data sources from the model for all years. That results in more reasonable estimates of cohort sizes (Fig. 10) while having a minimal impact on SSB (Fig. 11). This proposal was supported by the CIE reviewers. These data will be added back into the assessment model when a suitable approach is determined that can separate issues of changes in availability from cohort size.

Because dropping these data sources affects estimates of recruitment, and specifically decreases variation, σ_R, was estimated and found to be very close to 1. **As such, all subsequent models fixed $\boldsymbol{\sigma_R = 1}$ and used maximum penalized likelihood** to save on run time and because estimation σ_R was stable and had a minimal impact on the rest of the model. That is, fixing it at the MLE vs estimating it leads to negligible differences but saves on run time.

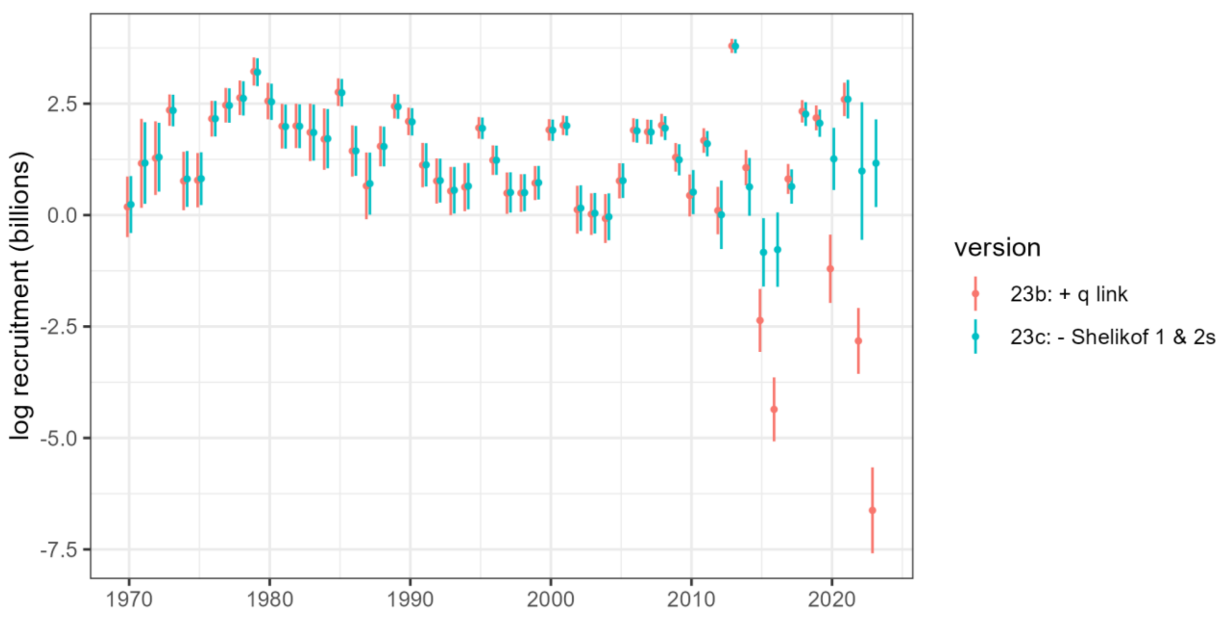

Figure 10. Estimates of recruitment in log space with (23b) and without (23c) the age 1 and 2 indices in the model.

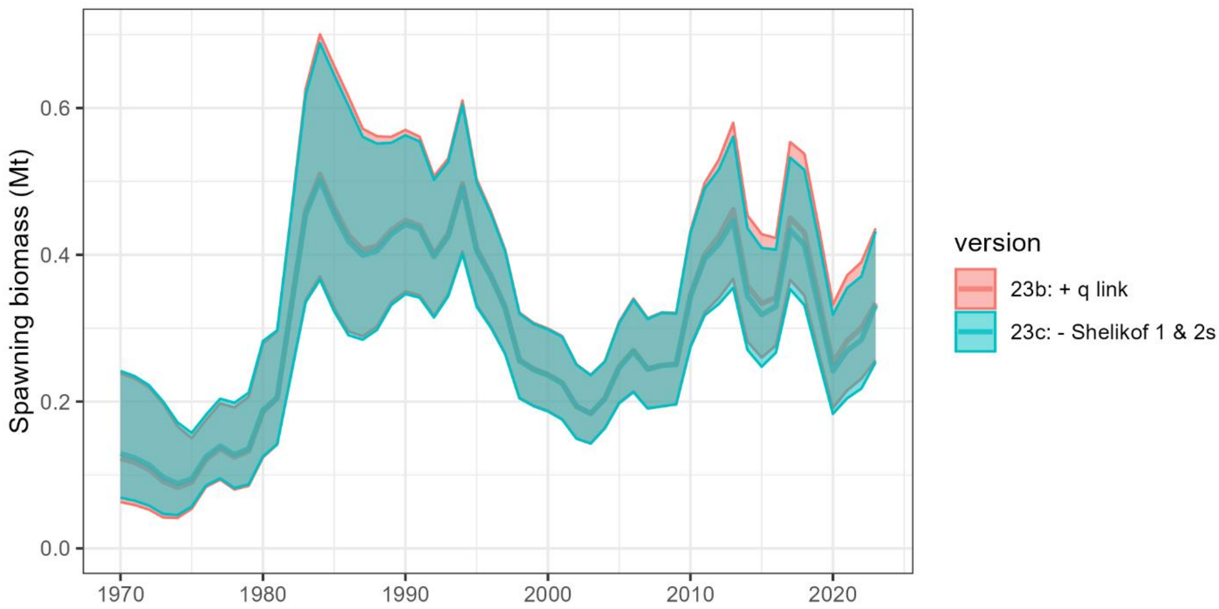

Figure 11. Spawning biomass comparison with and without the age 1 and 2 indices.

Model 23d: Switching to the Dirichlet-multinomial likelihood

Another major concern raised by the CIE reviewers was that the age composition effective sample sizes (ESS) were unreasonably small, being down weighted so much that important signals were missed. For example, the Shelikof straight age 3+ composition have estimated ESS of 5.7.

The Dirichlet-multinomial (D-M) is an alternative age composition likelihood that estimates the ESS jointly with the model (Thorson et al. 2017), replacing the need to iteratively tune via the Francis approach. It often will estimate higher ESS values and thus served as a reasonable alternative. The D-M was implemented in model 23c, and named 23d, and estimated higher ESS as expected (Fig. 12). It had no impact on the trend but did scale the population up slightly (Fig. 13).

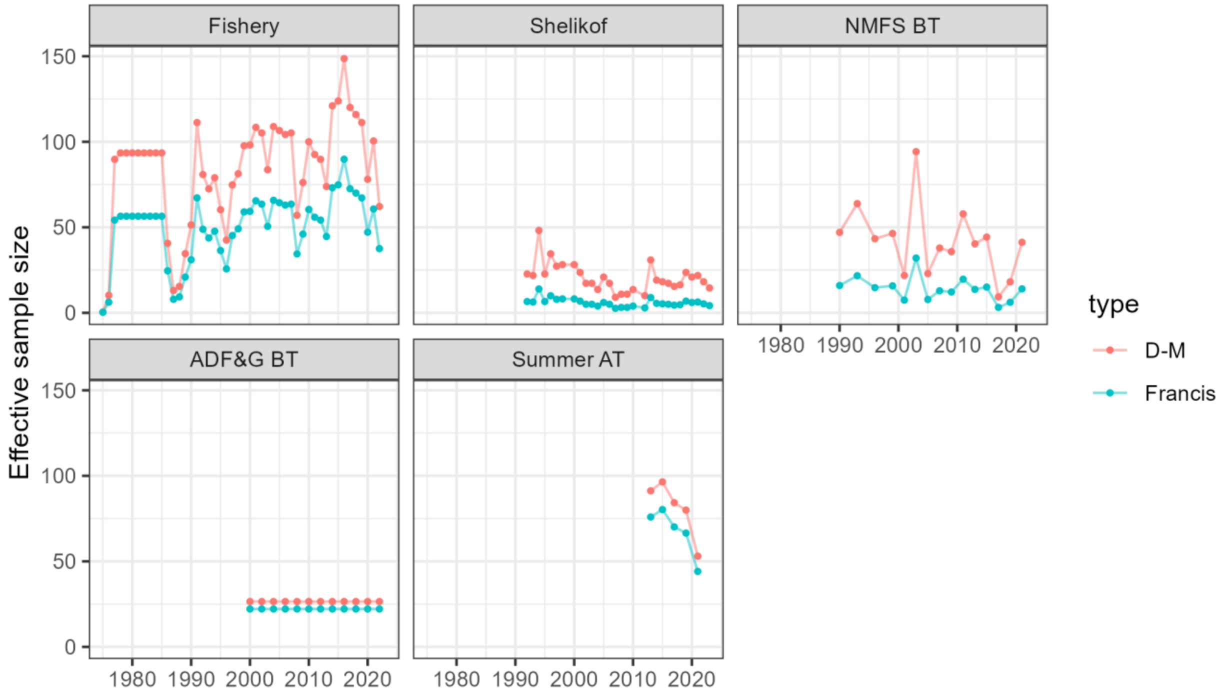

Figure 12. Estimated age composition effective sample sizes from the Francis-tuned multinomial (23c labeled here "Francis") and estimated via the Dirichlet-multinomial (23d, "D-M"). The D-M ESS is calculated from an estimated parameter and the input sample size.

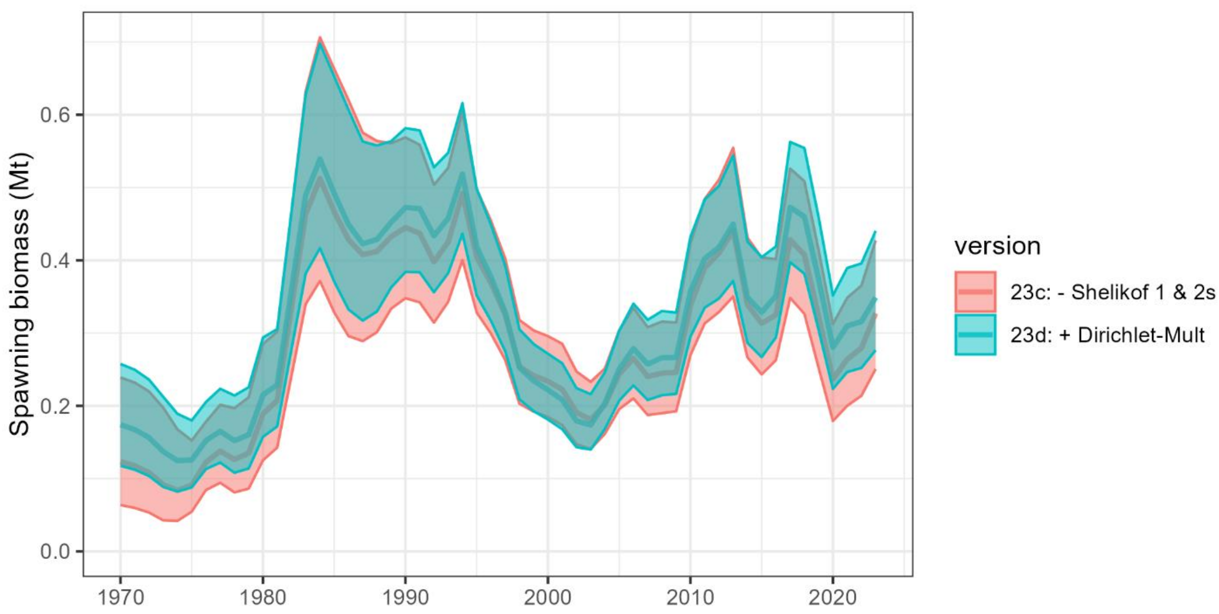

Figure 13. Spawning biomass comparison by adding the Dirichlet-multinomial distribution.

Model results

Each model built upon changes from the previous one and so incremental changes can be compared by looking at sequential models. No model changed the trend in a meaningful way, but the D-M reduced the CV of SSB by about 25% and estimated a larger initial population (Fig. 14).

Figure 14. Changes to SSB and the coefficient of variation (CV) for SSB (top row) and relative changes from model 23 (bottom row) across proposed models. Note that each model is cumulative with the previous.

I consider models 23c and 23d to be good candidates and put them forward for consideration this year. Note that 23c includes the cumulative changes of updated input CVs and ISS as well as the catchability link. Further, 23d is 23c with the Dirichlet-multinomial likelihood for all age composition data sets.

Selectivity was estimated similarly, except for the summer AT survey, which is estimated to select older fish more, particularly for model 23d (Fig. 15).

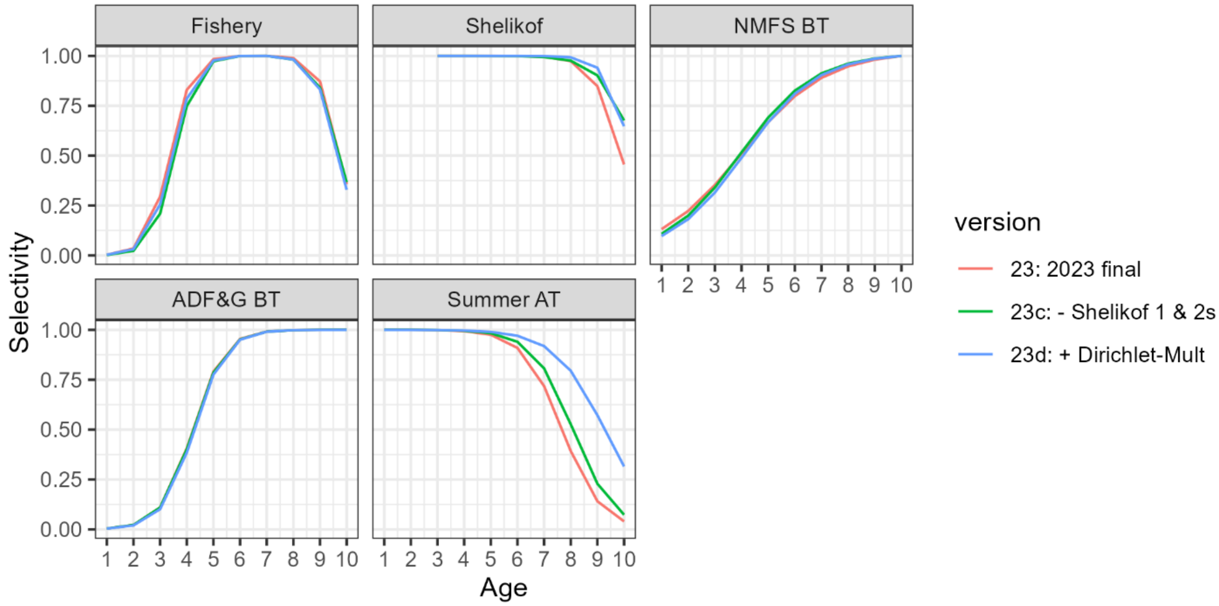

Figure 15. Selectivity estimates between candidate models.

As expected and discussed above, dropping the Shelikof age 1 and 2 indices had a large impact on recruitment, especially the smallest recent estimates (Fig. 16).

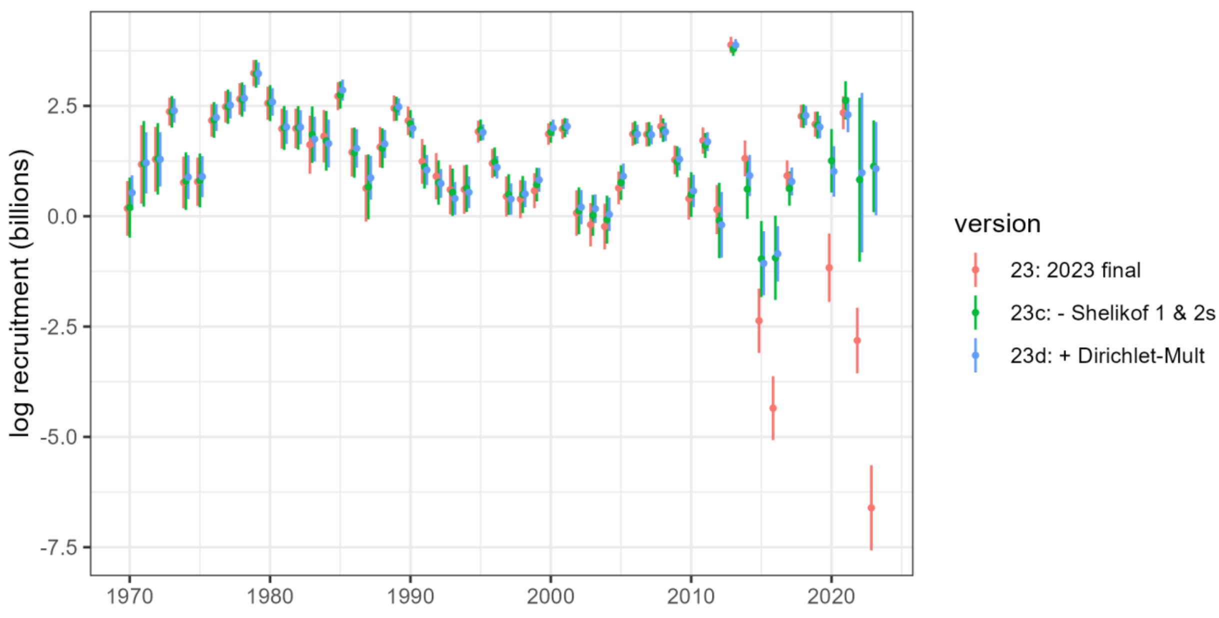

Figure 16. Comparison of log recruitment.

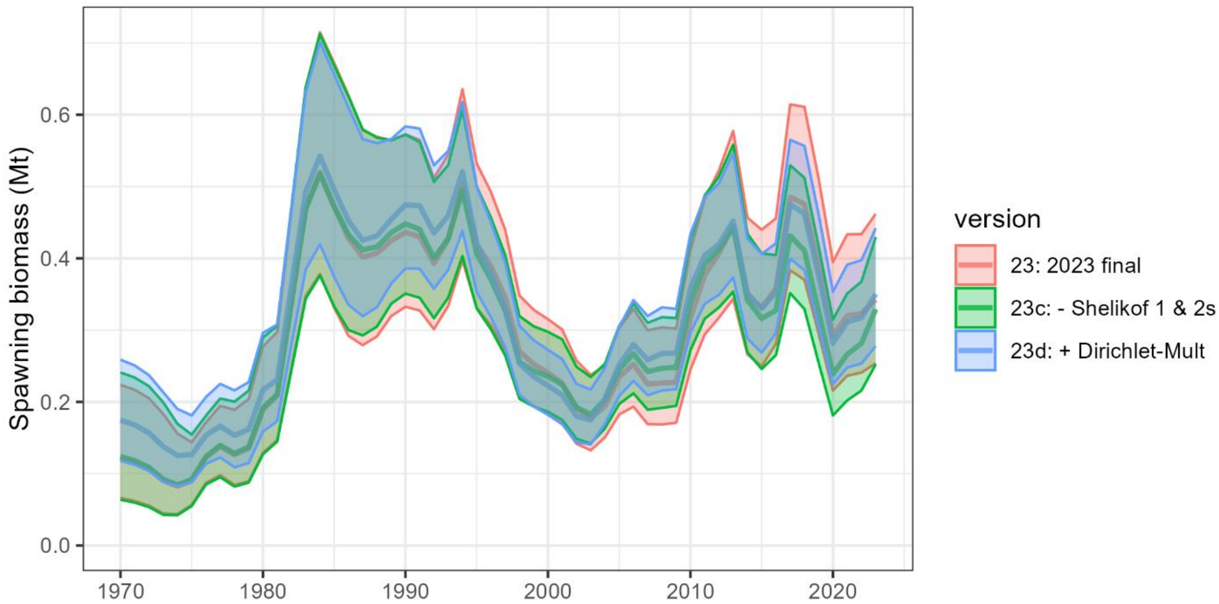

Figure 17. Spawning biomass comparison.

Table 3. Parameter estimates for model 23d. Random effects (deviations) are excluded.

Parameter	Estimate	SE
mean_log_recruit	1.393	0.148
log_slp2_fsh_mean	0.763	0.302
inf2_fsh_mean	9.672	0.122
log_slp2_srv1	0.765	0.602
inf2_srv1	10.288	0.350
log_slp1_srv2	-0.325	0.086
inf1_srv2	4.108	0.269
log_slp1_srv3	0.542	0.093
inf1_srv3	4.262	0.211
log_slp2_srv6	0.064	0.487
inf2_srv6	9.368	0.711
log_q2_mean	-0.246	0.092
log_q6	-0.488	0.132
transf_rho	0.476	0.185
log_Ecov_sd	0.011	0.141
Ecov_beta	0.332	0.036
log_DM_pars	-1.633	0.105
log_DM_pars	-0.655	0.156
log_DM_pars	-0.867	0.173

log_DM_pars	-1.984	0.165
log_DM_pars	-2.110	0.337
log_q6	-0.488	0.132

Model evaluation

Convergence, estimability, and stability

All models converged by standard diagnostics. Specifically, Newton steps using the inverse Hessian were successful and reduced the maximum absolute gradients to less than 1E-7. The Hessian matrices were also all invertible. A jitter analysis was performed on the original model and the two proposed for 2024, 23c and 23d. For each jitter analysis new initial values are generated by multiplying by fixed effects by random draws from a U(.9, 1.1) distribution, while leaving random effects initialized at zero. This was repeated 100 times, and for model 23 only 2 in 100 led to a different result that was <1 NLL unit worse, but also a similar SSB (Figs. 18 and 19). The two new proposed models all found the same mode and thus are considered stable. This is largely an effect of adding priors on selectivity. Overall the new proposed models converge well and are stable with respect to initial values.

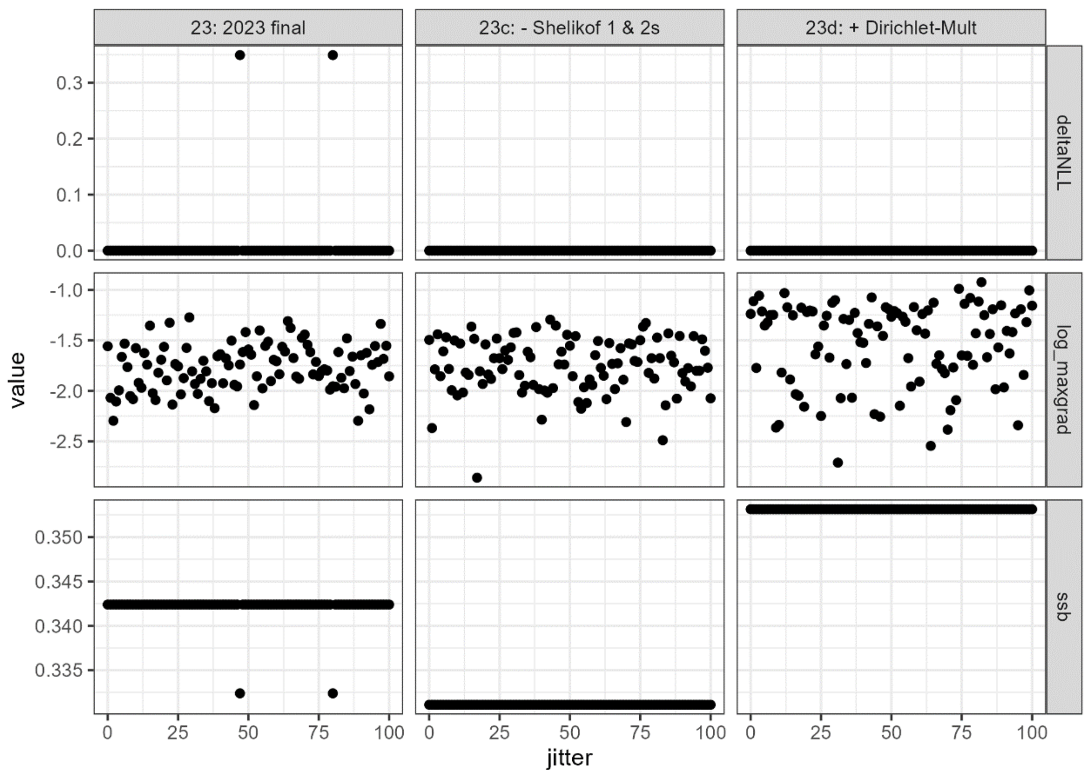

Figure 18. Jitter results for the change in NLL (deltaNLL), maximum gradient, and spawning biomass (ssb). Models 23c and 23d had no issues, while 1 in 100 jitters for model 23 was an issue.

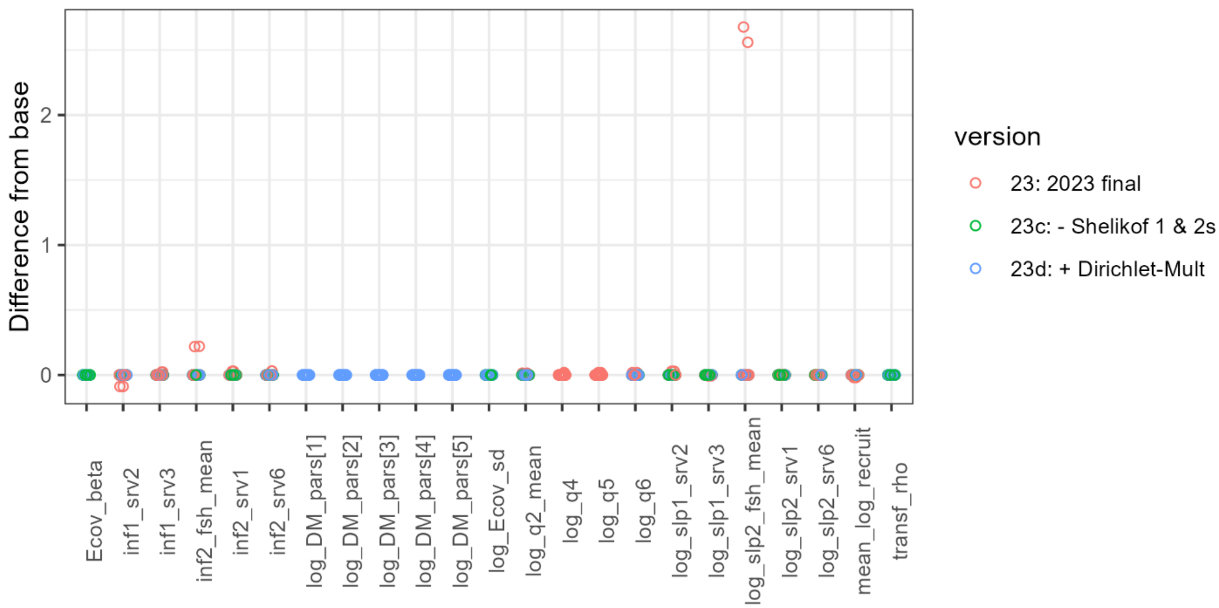

Figure 19. Jitter results for fixed effects. Model 23 had some selectivity parameters that found a different mode.

Posterior sampling was done via the adnuts R package (Monnahan and Kristensen 2018, Monnahan 2021), including a new algorithm in development which utilizes the joint precision matrix to decorrelate prior to sampling and which makes sampling much more efficient and works when process errors are estimated. The original model had severe convergence issues (Fig. 20), and the updated model much better (Fig. 21). This is primarily due to the regularization priors on selectivity.

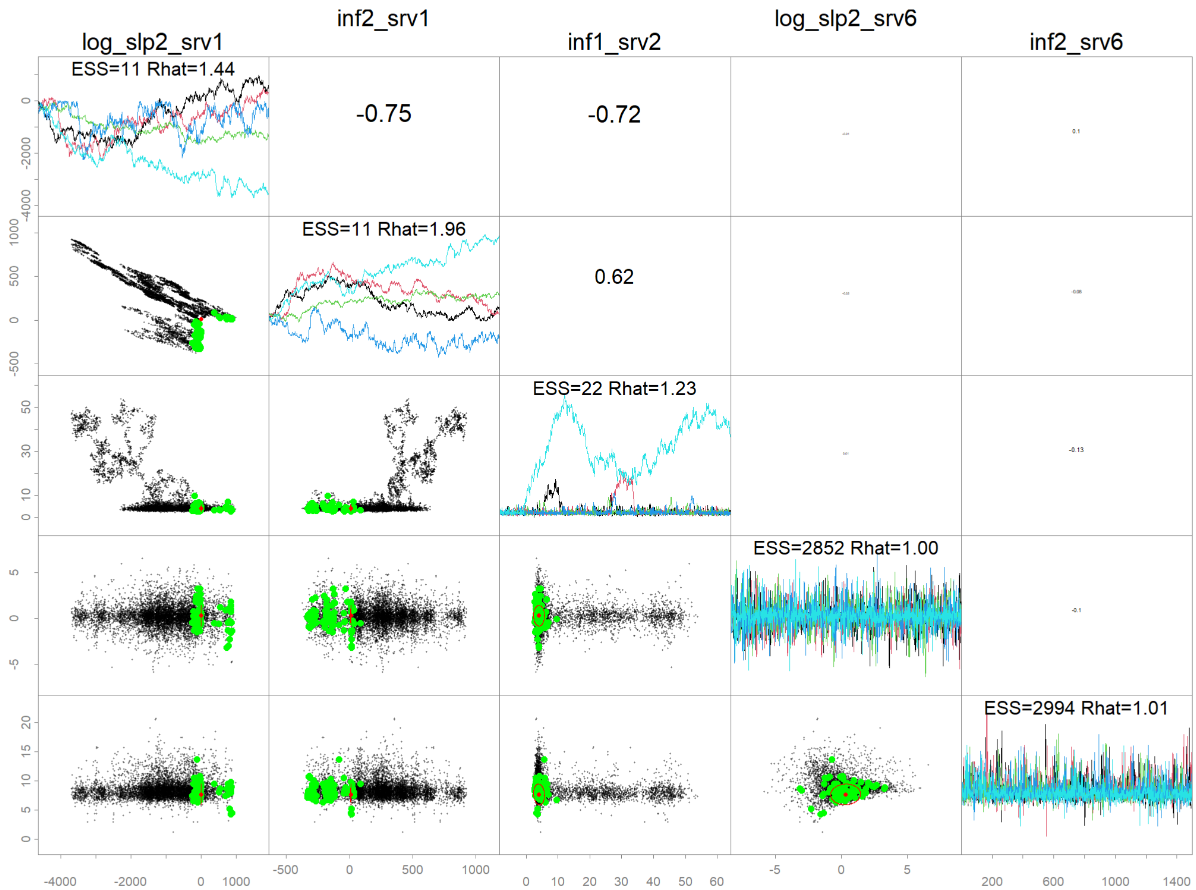

Figure 20. MCMC analysis of pars that mixed poorly for model 23. Diagonal panels show the trace plots, while lower triangle panels are pair-wise posterior samples (black points), the MMLE covariance (red point and ellipses) and green points show no-U-turn divergences. Note the poor convergence as noted by small MCMC effective sample size (ESS) and large Rhat values given on the diagonal panels.

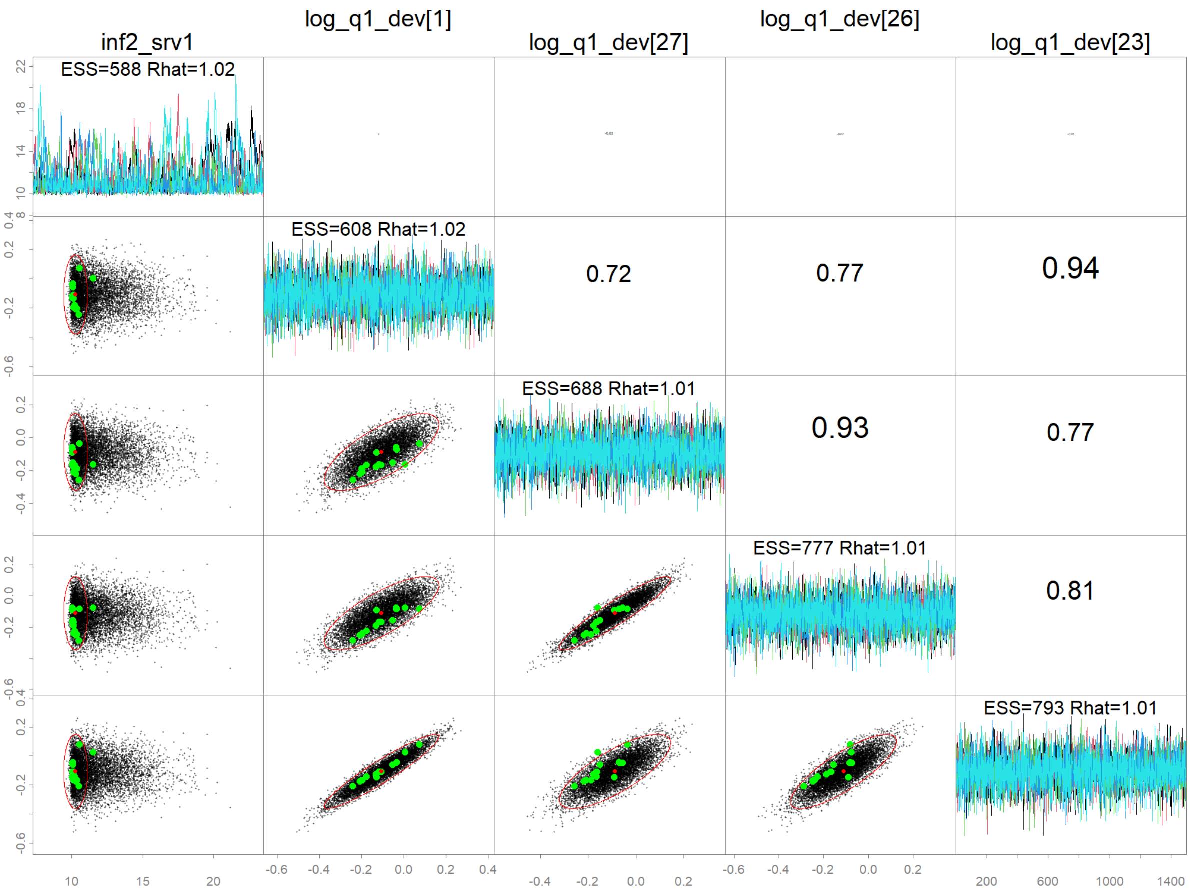

Figure 21. Same as the previous plot, but for model 23d. Note the much better convergence.

Posterior samples are not directly used for management but are used in the calculation of probabilities of depletion. It is clear that the new models will be effective at this.

Finally, a self-test simulation experimented was conducted for model 23c only. Model 23 did not have convenient capabilities to do this, and model 23d which uses a Dirichlet-multinomial has lingering issues that I was not able to resolve this year. This test involved simulating 100 data sets from the fitted model and refitting the matching model, comparing the estimated outputs (SSB and recruits) against the original which is taken as the truth. For this year I did not resample the process errors but that would be an improvement in future years.

For model 23c both recruitment and SSB showed no signs of bias (Fig. 22). As expected the variance in relative error for recent recruits and early SSB is higher due to limited information in the data.

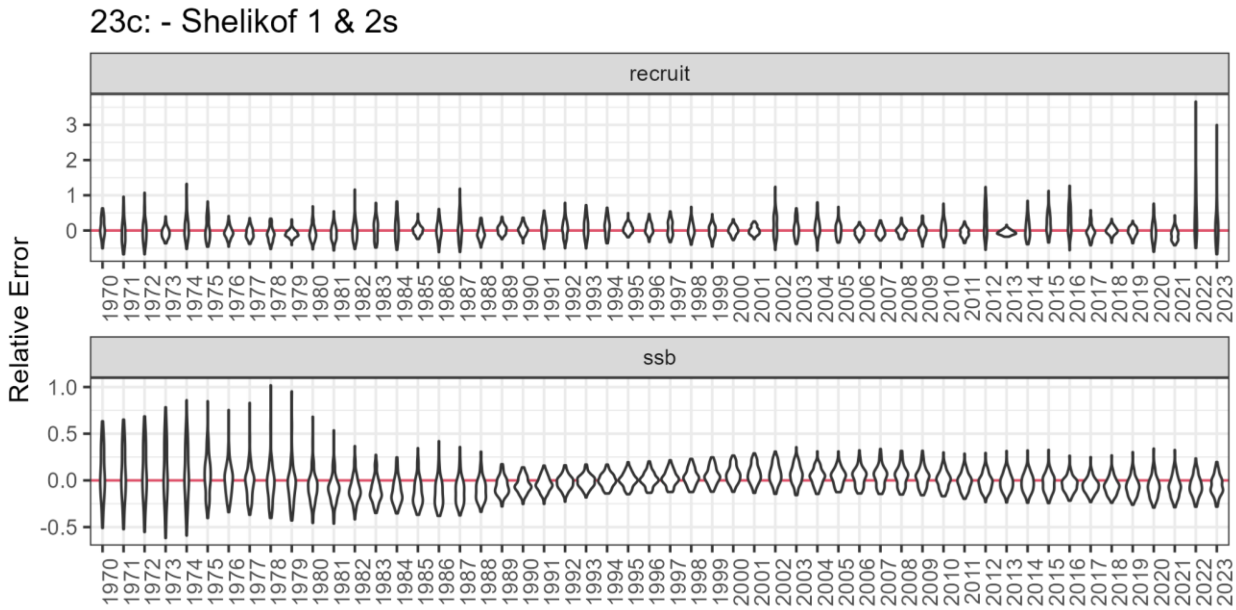

Figure 22. Results of self-test simulation for the only proposed model where it was feasible this year. Relative error is calculated with the original fitted model as the reference. The model exhibits no obvious signs of bias.

Model validation

Fits to the Shelikof index are much better with models 23c and 23d due to the inclusion of the catchability link (Fig. 23), and overall the SDNR values are improved and show no concerns (Fig. 24).

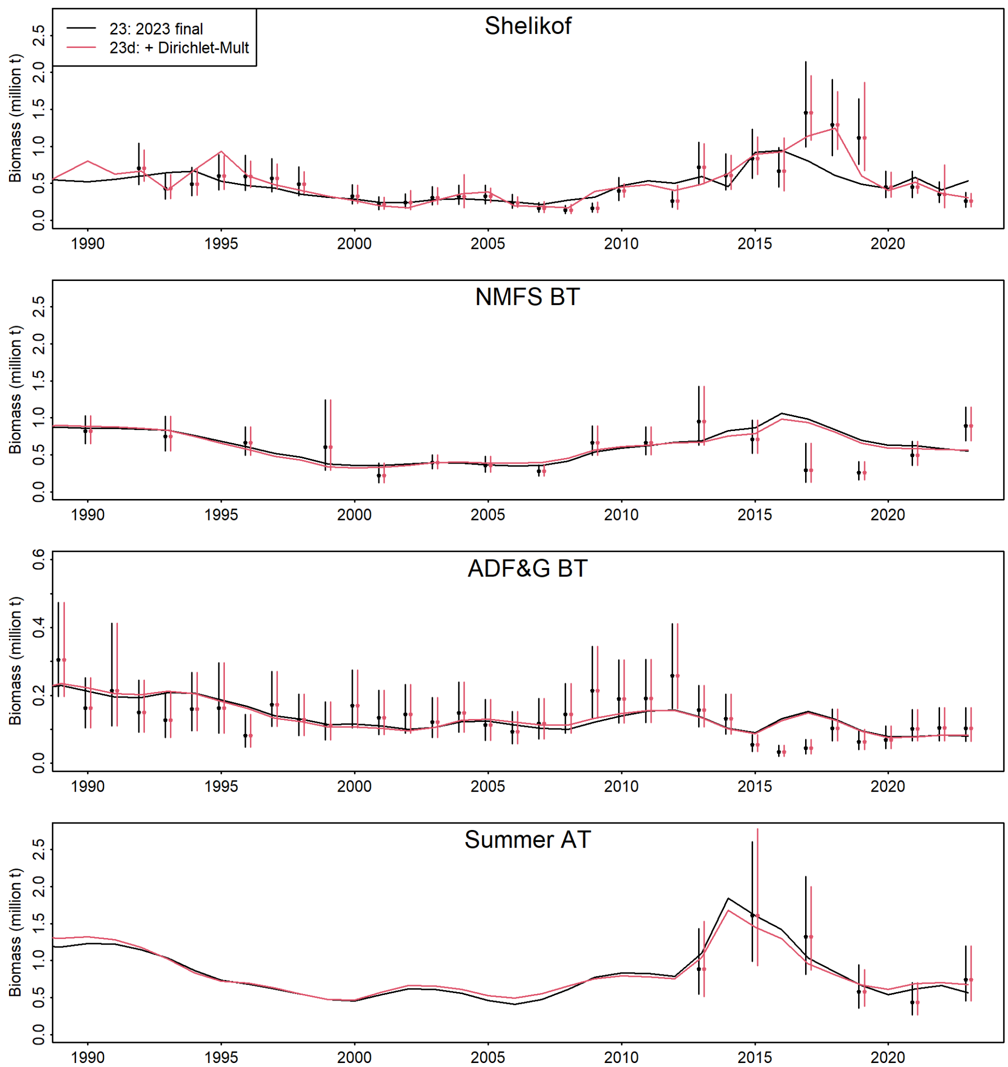

Figure 23. Model fits (lines) to index data (points with 95% confidence intervals) for models 23 and 23d. Model 23c is left off for visual clarity.

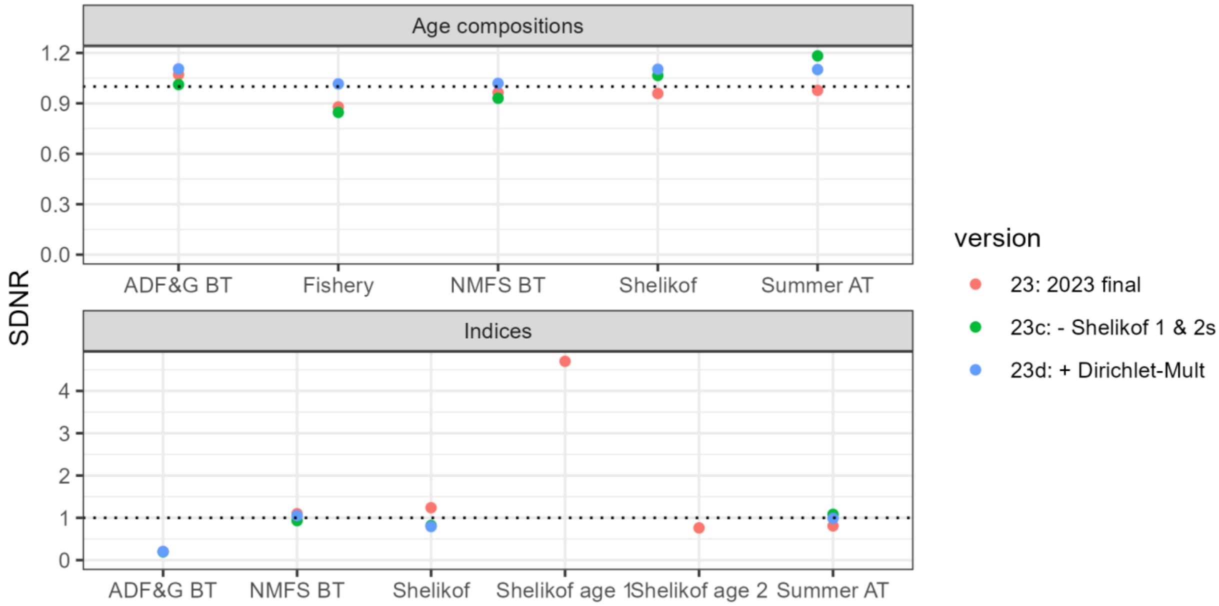

Figure 24. Standard deviation of normalized residuals (SDNR) for age compositions and indices.

The age compositions fit well aggregated across years (Fig. 25), although the NMFS BT shows some potential misfit.

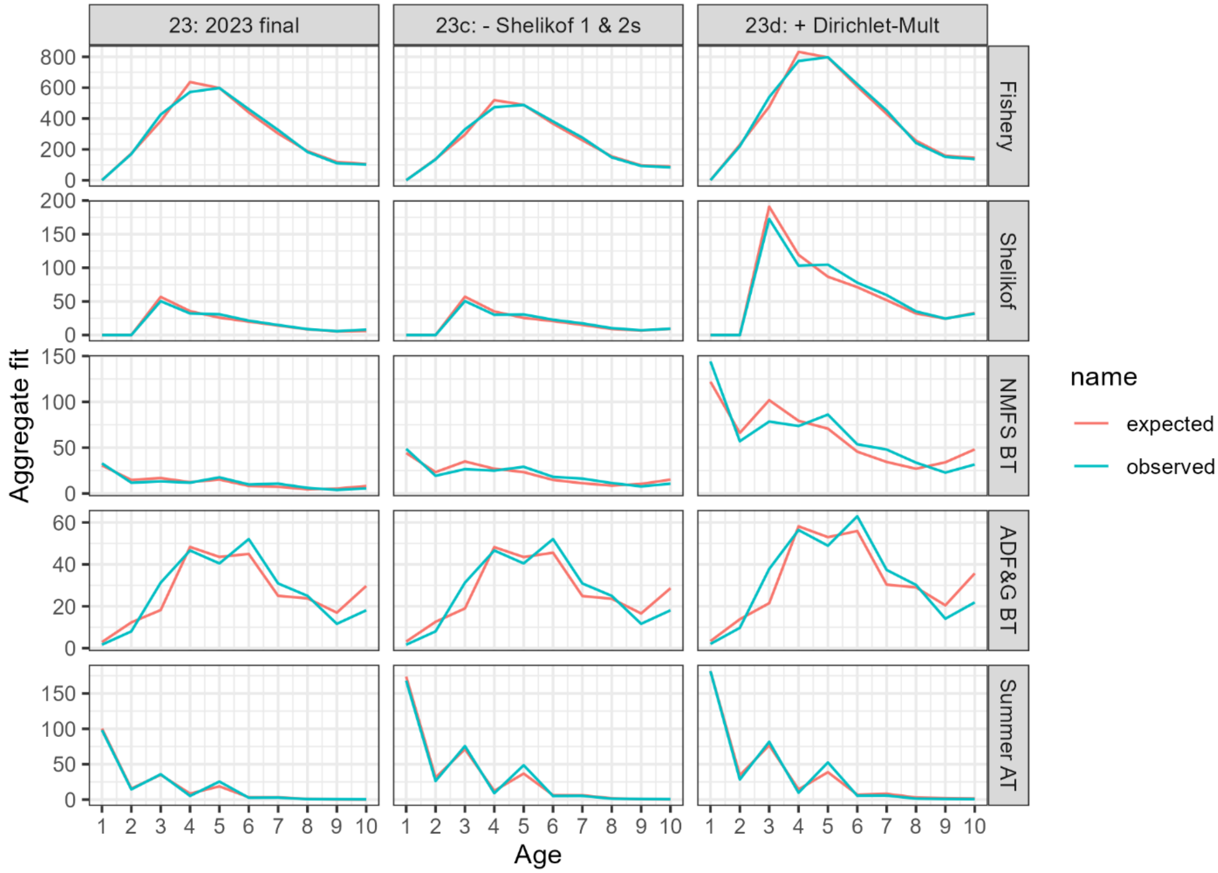

Figure 25. Aggregated age composition fits (summed across years) for the three models. The higher y-axis values for model 23d are due to the Dirichlet-multinomial putting higher weight on these data (higher ESS). Note that Shelikof age 1 and 2 fish are not in the age compositions, and hence are zero.

One-step-ahead (OSA) residuals were calculated externally via the compResidual R package (Trijoulet et al. 2023). These age composition residuals are expected to be independent and standard normal under a correctly specified model, and deviations from this indicate lack of fit (Thygesen et al. 2017). QQ plots show that there may be a few outliers for the fishery and model 23d, but overall the fits are acceptable and show no major concern (Fig. 26).

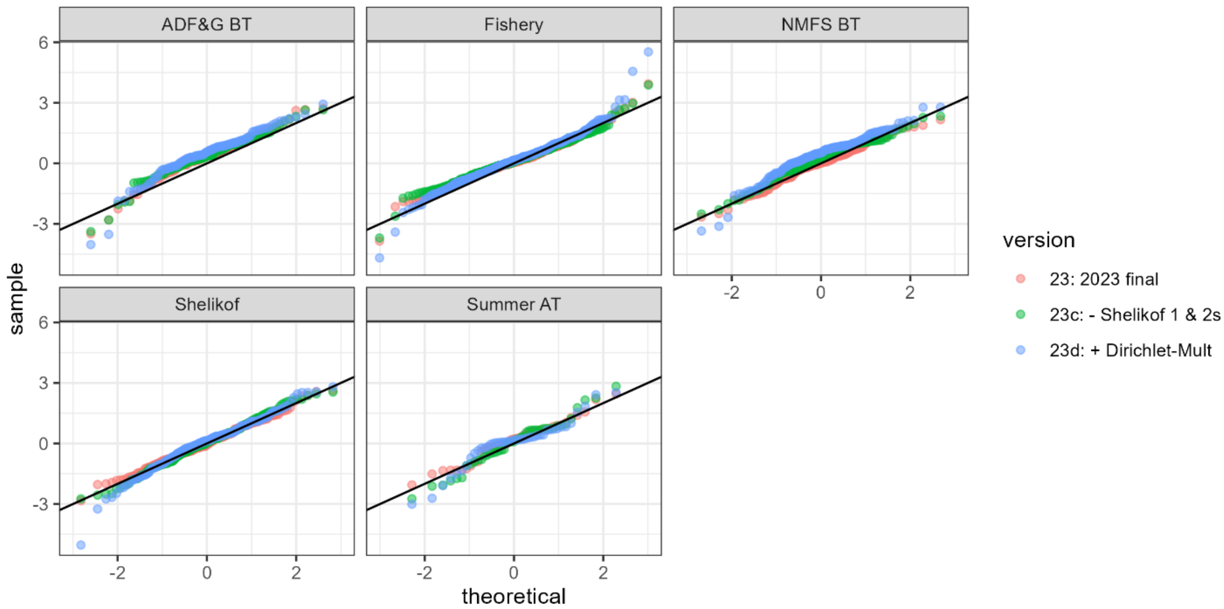

Figure 26. QQ plots of OSA residuals.

Bubble plots indicate some lack of fit in all data sets for all models (Fig. 27). One important observation is that the age compositions in model 23d are weighted much higher and thus are expected to be more difficult to pass validation tests. It appears that temporal patterns of misfit in the other two models are exacerbated in model 23d, for instance age 3 for Shelikof in recent years, or the apparent misfit to the cohort starting around 2000. In other cases, the misfit appears unique in 23d, such as the NMFS BT for ages 7 and 8.

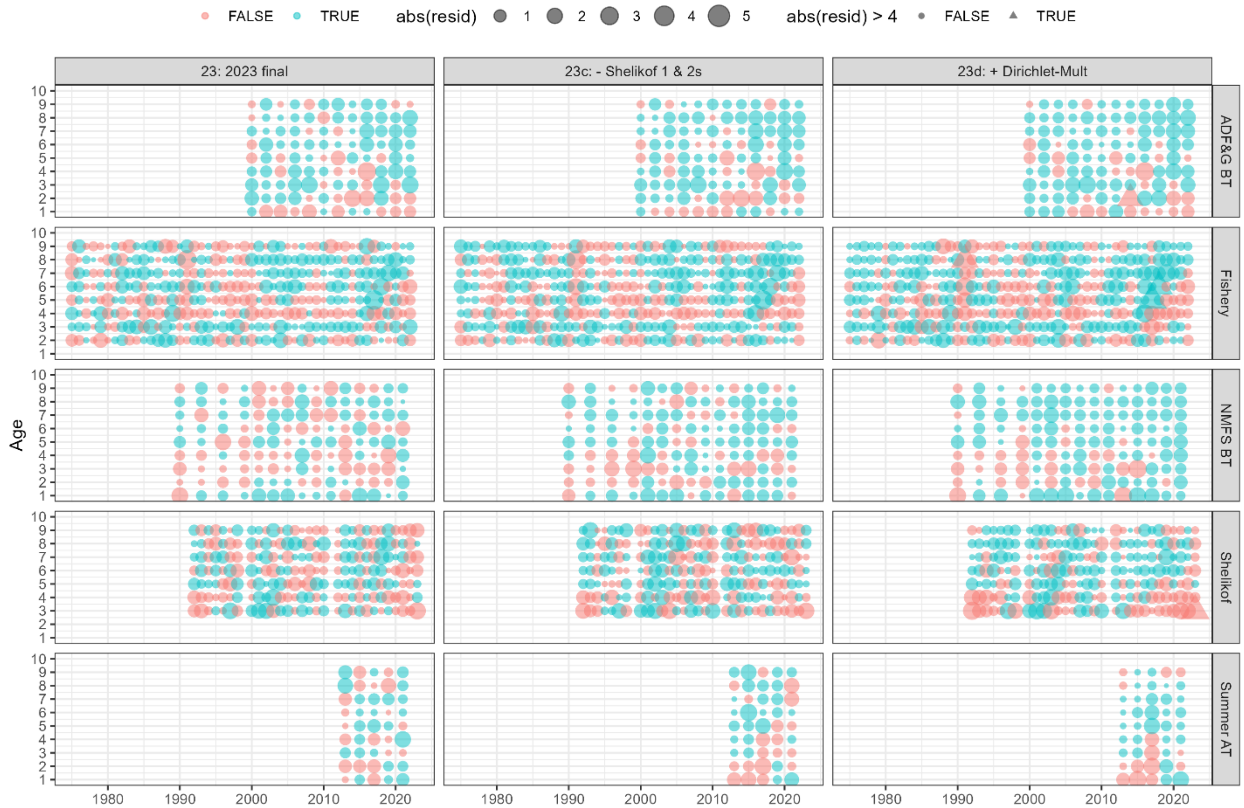

Figure 27. Bubble plots for OSA residuals for the age compositions.

Retrospective patterns

The proposed models do not improve on the retrospective patterns from the base model (Fig. 28). Estimation of cohort size was affected, primarily due to the exclusion of the age 1 and 2 indices starting in model 23c (Fig. 29).

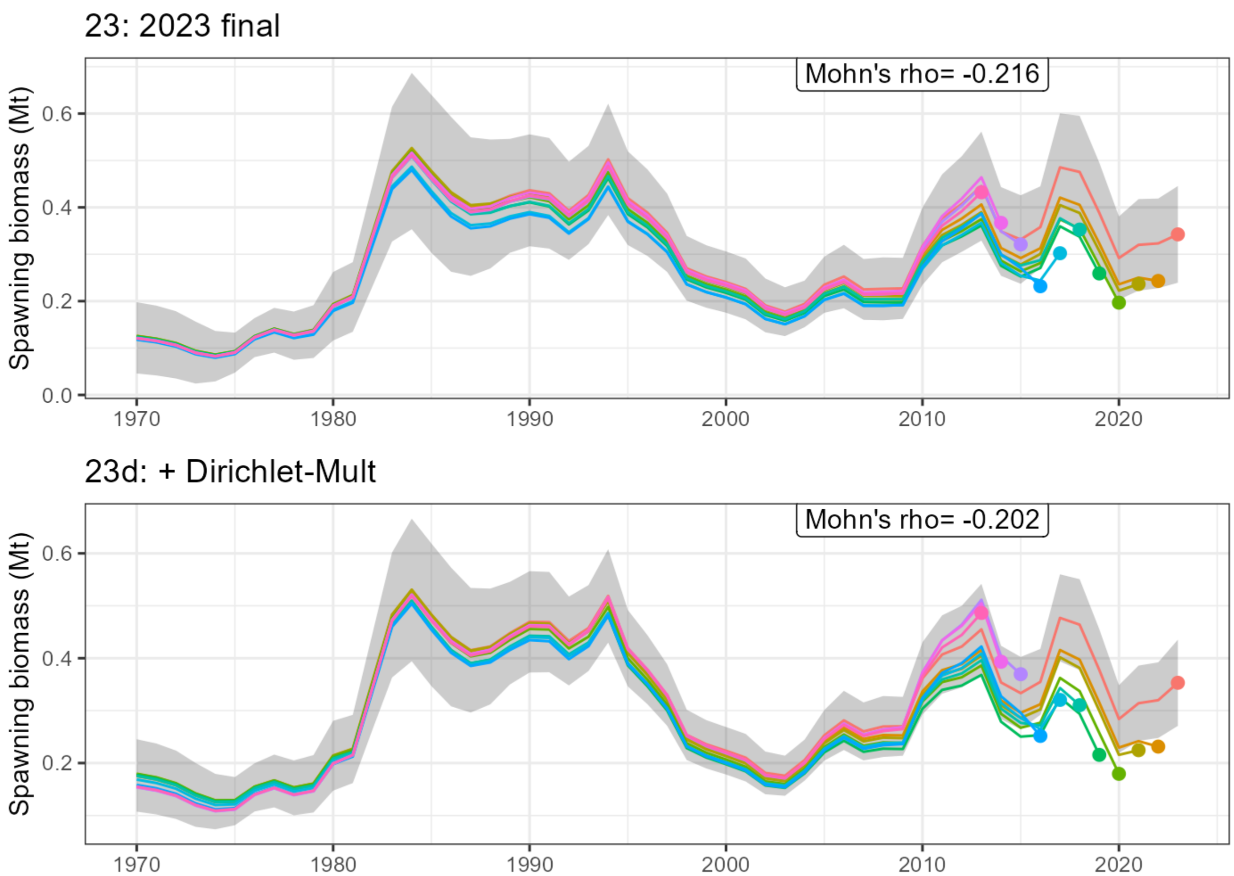

Figure 28. Retrospective fits for spawning biomass. The gray lines indicate the uncertainty from the base (peel 0).

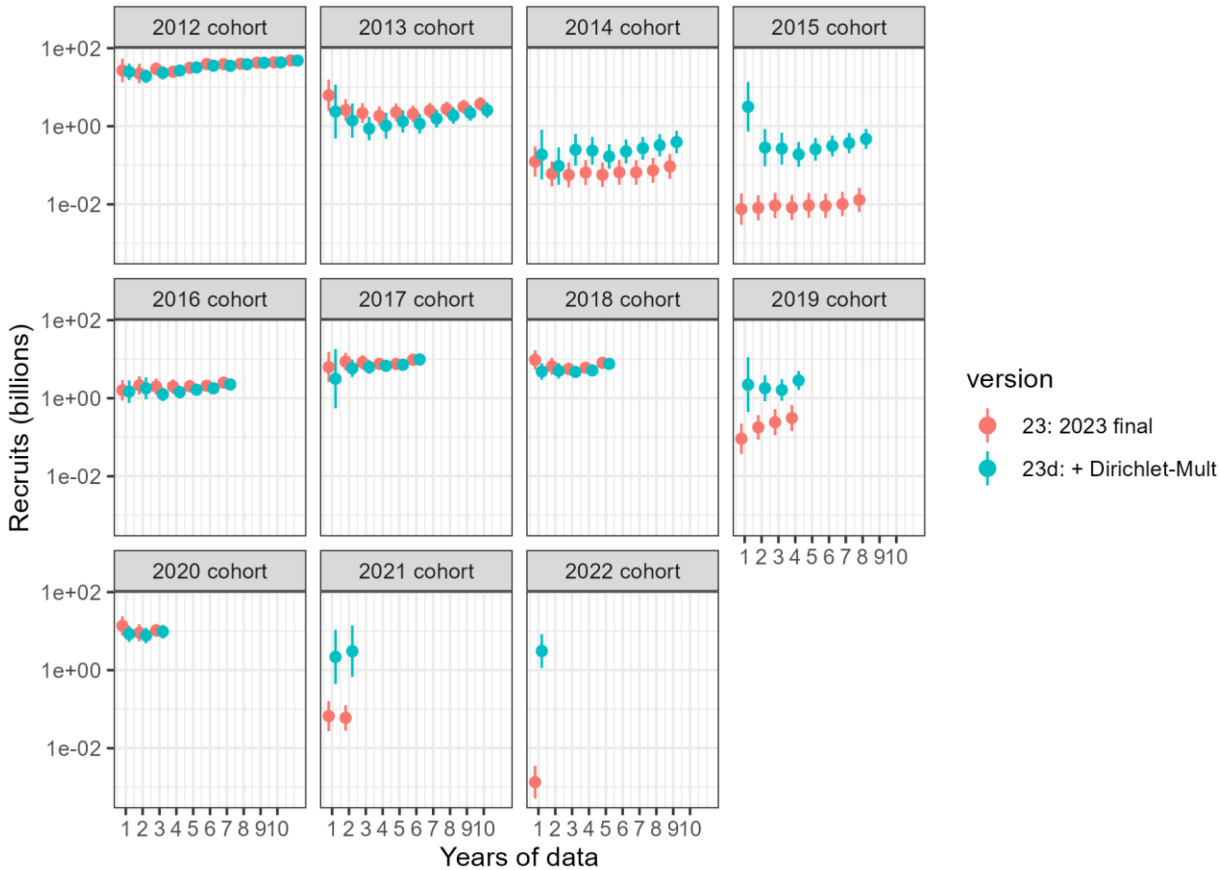

Figure 29. Recruitment estimates by cohort as data are added to the model, taken from retrospective peels.

Likelihood profiles

Likelihood profiles were run for two parameters that affect the scale of the stock: natural morality (*M*) and catchability for the NMFS BT survey (q2), where the latter has a strong prior known to mitigate uncertainty in scale for this stock. In both cases the only random effects marginalized were associated with the catchability link for Shelikof (the smoother on the observed covariate). Natural mortality is assumed aged-based and known in the assessment. This age-based curve was scaled up by multiplying by a constant, and *M*=0.3 at age 6 is used as a reference point for the *x*-axis in Fig. 30.

Initially a profile on mean recruitment, R0, was done, but was deemed ineffective and misleading because without a sum-to-zero constraint the deviations can offset the change in R0 and keep recruitment equivalent with only the recruitment penalty changing.

There was a strong total signal on *M*, but also conflict among the data sources (Figs. 30, 31). Notably, the profile is narrower for model 23d due to the increased weight on the age composition data when using the Dirichlet-multinomial distribution.

Similarly, model 23d has a stronger data signal on q2, and there was some clear conflict between Shelikof ages and the NMFS BT ages (Fig. 32).

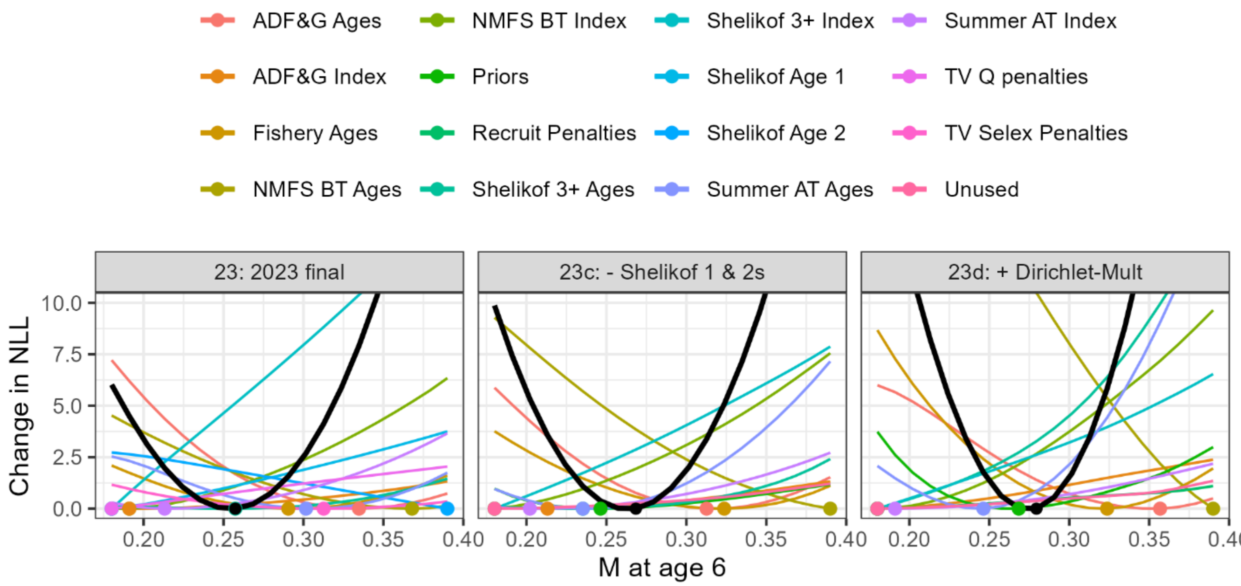

Figure 30. Likelihood profile across natural mortality (M), showing total (black line) and components (colors) which had non-negligible impacts with their minima shown as colored points. The model currently assumes age-based M, with age 6 assumed to be 0.3.

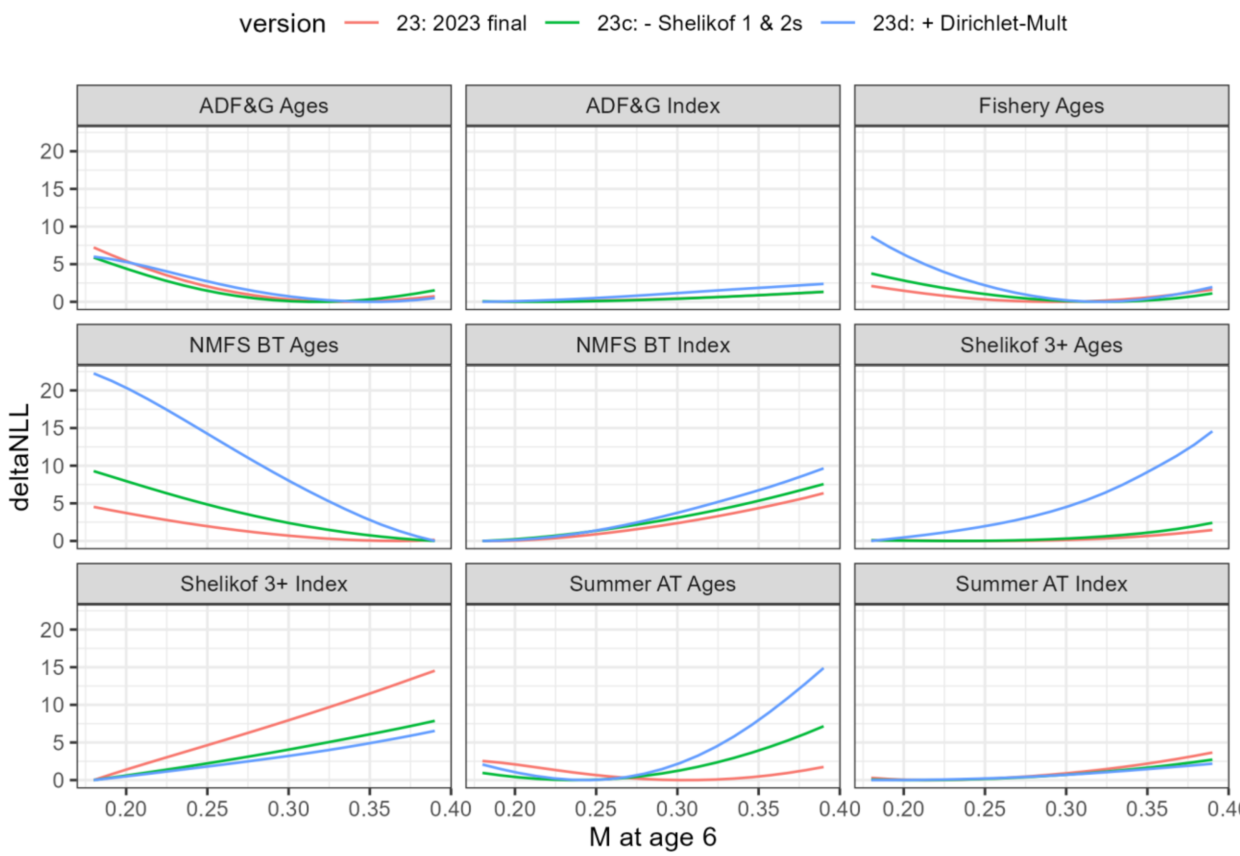

Figure 31. Profile likelihood results for natural mortality (M) plotted by data component. A value of 0.3 is used in the model.

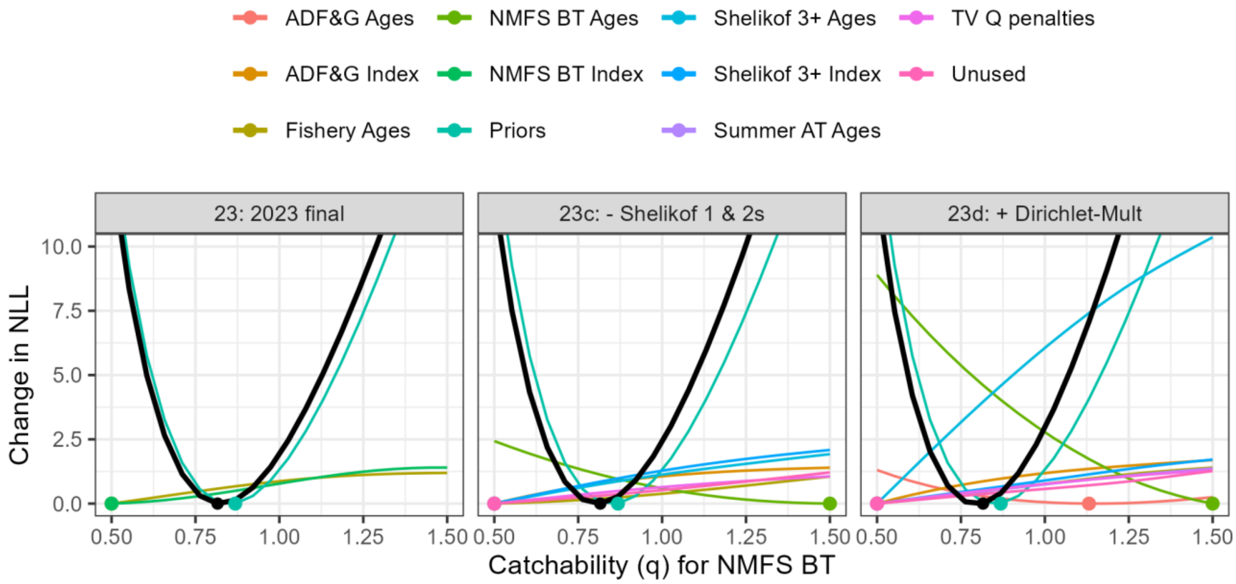

Figure 32. Likelihood profile across catchability for the NMFS bottom trawl (q2) showing total (black line) and components (colors) which had non-negligible impacts, with their minimal shown as colored points.

Sensitivity tests

Given the conflict in data sources identified above it is important to test the sensitivity of the model to the different surveys. I therefore reran the model leaving each survey out (indices, ages, and lengths, as appropriate). The model is insensitive to trends in SSB for all stocks, and fairly insensitive to scale when dropping surveys. The key exception is the NMFS BT where the uncertainty of SSB increases, but also the absolute scale increases notably without the survey in it for all versions of the model (Fig. 33). This is the survey with a fairly informative prior on *q*, and one that has recently fit poorly to the index (Fig. 23), as well as exhibiting conflict between its index and age composition (Figs. 30, 31). This sensitivity test was also presented at the 2021 September Plan Team (see page 13 of this report), and is a known issue with this stock. The difference this year is that the mean SSB has shifted up, whereas in 2021 the scale did not change as much and only uncertainty increased. As such, this presents a new issue with the stock that is likely related to the recent misfit to the index, but is independent of the model version used.

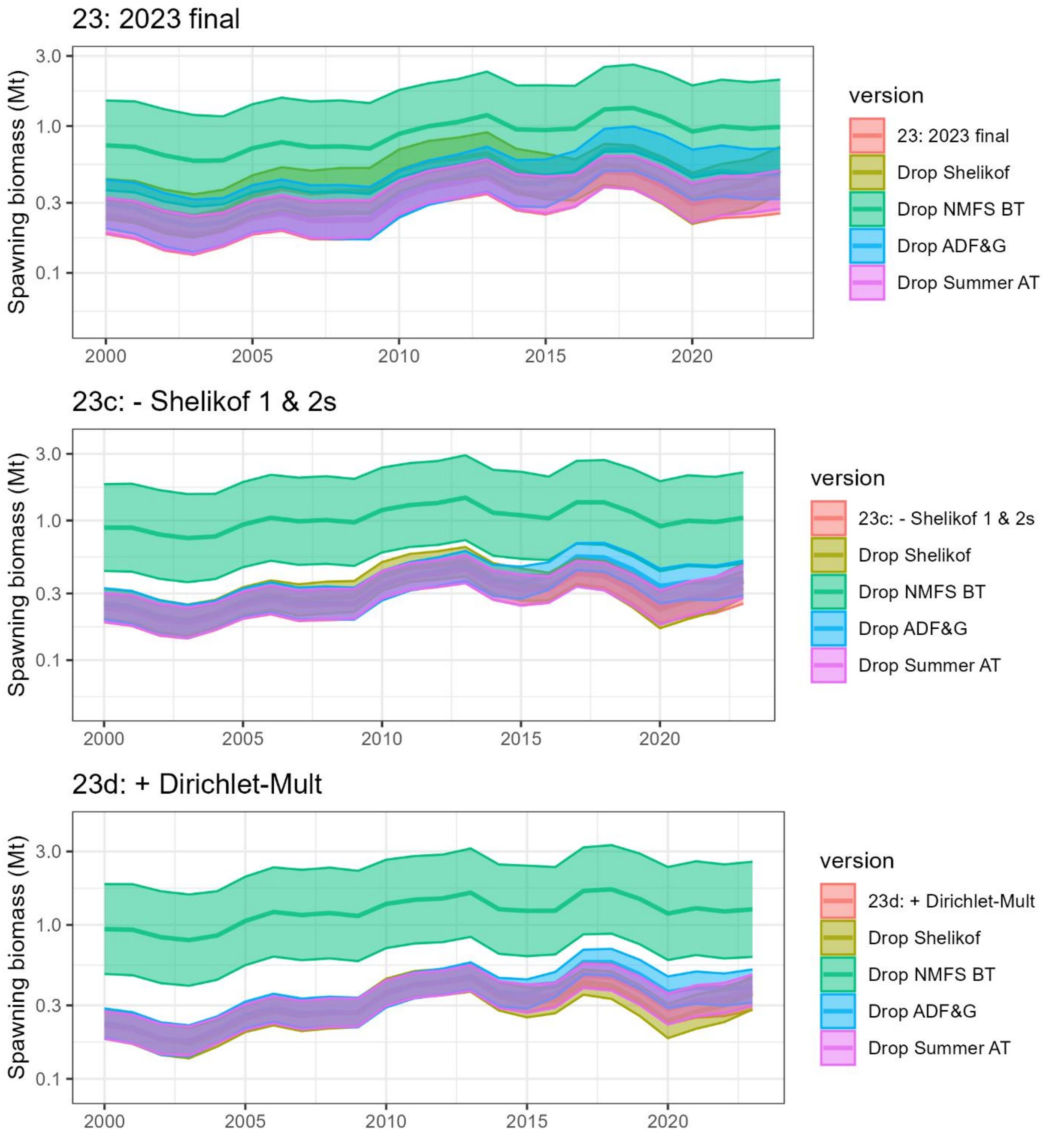

Figure 33. Results of sensitivity test of leaving each survey out (colors). Note the y-axis is in log scale to better highlight differences, and slightly different ranges among the 3 panels (model versions).

Recommendations

All proposed models are more stable, have more reliable estimation, and substantially improved MCMC performance. I believe that all changes through model 23c are reasonable improvements to the stock assessment. I therefore propose this for consideration in 2024. Additionally, adding the Dirichlet-multinomial to model 23c (23d) has some clear benefits, but also some new issues. I thus focus my recommendations on these two models.

Model 23d has increased weight on the age compositions, and as a consequence slightly worse age composition fits as measured by OSA residuals. However, I argue that the higher ESS values are more appropriate, and that the Francis approach was down weighting the age data too much. I believe one consequence of this down weighting is that it obscures misfit. So, in that sense, the added misfit with model 23d may be appropriate.

Removing the age 1 and 2 Shelikof indices from the model is unfortunate because it removes the best early signal of cohort strength available to the model. But these data simply do not fit well, even with very large CVs, and it is clear that in years with very small estimates is not a reliable signal. This is known to the MACE scientists, who designed and execute this pre-spawner survey. The extremely small cohorts estimated from them has no major impact on the population dynamics of the stock because in natural space their contribution is negligible in both cases. However, when left in, the vastly overstate the recent drop in productivity and leads to estimates of recruitment variability that are unreasonably large (see risk table discussion in the 2023 SAFE for more).

References

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

Hulson, P.-J. F., and B. C. Williams. 2024. Inclusion of ageing error and growth variability using a bootstrap estimation of age composition and conditional age-at-length input sample size for fisheries stock assessment models. Fisheries Research **270**:106894.

Kimura, D. 1989. Variability in estimating catch-in-numbers-at-age and its impact on cohort analysis. Effects on ocean variability on recruitment and an evaluation of parameters used in stock assessment models. Can. Spec. Publ. Fish. Aq. Sci **108**:57-66.

Kristensen, K., A. Nielsen, C. W. Berg, H. Skaug, and B. M. Bell. 2016. TMB: Automatic differentiation and Laplace approximation. Journal of Statistical Software **70**:21.

Monnahan, C. C. 2021. adnuts: No-U-turn sampling for ADMB and TMB models. R package version 1.1.2.

Monnahan, C. C. 2024. Toward good practices for Bayesian data-rich fisheries stock assessments using a modern statistical workflow. Fisheries Research **275**:107024.

Monnahan, C. C., G. D. Adams, B. E. Ferriss, S. K. Shotwell, D. R. McKelvey, and D. W. McGowan. 2023. Assessment of the walleye pollock stock in the Gulf of Alaska. In Stock Assessment and Fishery Evaluation Report for Groundfish Resources of the Gulf of Alaska. Prepared by the Gulf of Alaska Groundfish Plan Team, North Pacific Fishery Management Council, P.O. Box 103136, Anchorage, AK 99510. North Pacific Fisheries Management Council, Anchorage, AK.

Monnahan, C. C., and K. Kristensen. 2018. No-U-turn sampling for fast Bayesian inference in ADMB and TMB: Introducing the adnuts and tmbstan R packages. Plos One **13**:e0197954.

Rogers, L. A., C. C. Monnahan, K. Williams, D. T. Jones, and M. W. Dorn. 2024. Climate-driven changes in the timing of spawning and the availability of walleye pollock (*Gadus chalcogrammus*) to assessment surveys in the Gulf of Alaska. ICES Journal of Marine Science. 10.1093/icesjms/fsae005.

Stock, B. C., and T. J. Miller. 2021. The Woods Hole Assessment Model (WHAM): A general state-space assessment framework that incorporates time- and age-varying processes via random effects and links to environmental covariates. Fisheries Research **240**:105967.

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

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

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