Gulf of Alaska Pollock Model Updates

Cole Monnahan

Alaska Fisheries Science Center

September, 2022

Executive Summary

The GOA pollock assessment model structure has remained unchanged since 2019 when model 19.1 was adopted for management. The model fits the data well and has no major pressing issues. This year, a suite of updates based on based on PT and SSC suggestions were explored. Out of these, three minor alternative models related to small structural changes were deemed worthy of consideration and are put forth for consideration. They include (1) adding a consistent penalty for all recruitment deviations; (2) removing the prior on the NMFS BT catchability; and (3) Adding a more flexible form for fishery selectivity to remove persistent patterns in age 3 and 4 age composition residuals.

Table 1. Comparison of model alternatives for the 2021 assessment model.

Model name	Model Description	2022 ABC (t)	2022 OFL (t)	2022 SSB (t)	2021 SSB (t)	Rel Error 2022 ABC
19.1	2021 final (base)	133,081	154,983	186,481	195,758	0.0%
19.1a	Add σ_R=1.3	137,004	159,587	190,808	199,588	2.9%
19.1b	No BT Q prior	141,230	164,325	204,529	215,550	6.1%
19.1c	Update fishery selex	127,870	149,272	179,463	188,040	-3.9%

The authors recommend models 19.1a and 19.1c, and leave 19.b up to the discretion of the Plan Team.

Several research track models are presented as well, but are not ready to put forth for consideration in 2022.

Proposed models

Model 19.1a: Adding consistent recruitment deviation penalty

In December 2021 the SSC noted "… *that recruitment deviations in the GOA pollock assessment are unconstrained except for the terminal two years, and suggests that exploring a moderate constraint on recruitment deviations in all years, as is commonly applied in other assessments, may be warranted. At a minimum, this would allow an assessment of the sensitivity of results to only constraining the last two years.*"

Model 19.1 has a penalty of σ_R=1 on the first eight, and last two recruitment deviations. All other deviations are unpenalized and thus freely estimated. This generally does not cause any optimization convergence issues for maximum likelihood estimation, but does for integration via MCMC. The perennial question is what value to use for the penalty. Instead of continuing with the arbitrary value of 1, the WHAM assessment model (Stock and Miller 2021) was used to estimate this penalty (process error).

This model, using the Laplace approximation of the marginal likelihood, estimated that $\sigma_R \sim N(1.34, 0.14)$, significantly higher than the existing value of 1.0.

Consequently, a value of $\sigma_R=1.3$ was applied to all recruitment deviations in model 19.1a. This resulted in small differences except in the smallest deviations which were shrank toward the prior (hyperdistribution) as expected (Fig. 1). Interestingly, the smallest deviations (e.g., 2016) were well outside the expected range of the prior but resisted, indicating strong information in the prior that the cohort was extraordinarily small. Because these changes affect primarily only very small cohorts, there was a minor difference in recruitment and SSB (Fig. 2), and overall a 2.9% difference in 2022 ABC (Table 1).
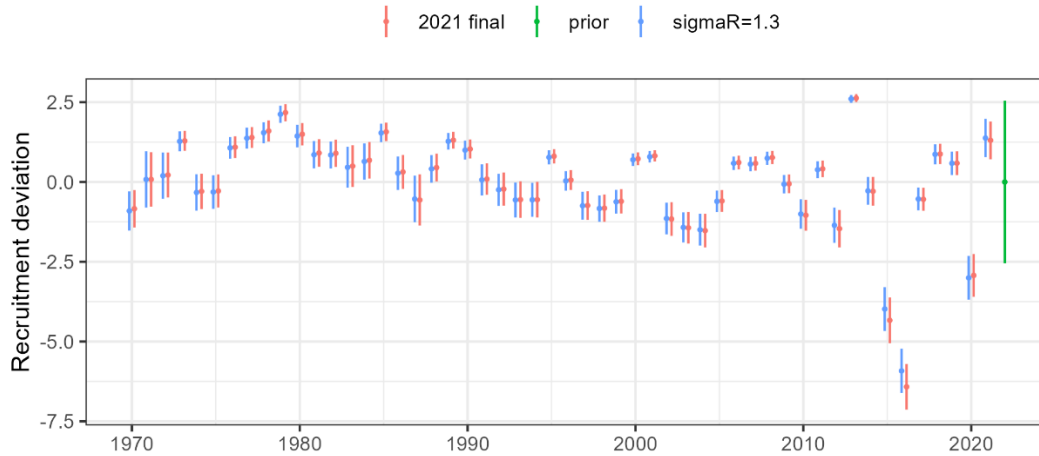

Figure 1. Estimated (log) recruitment deviations for the base and alternative models (colors), with points and bars showing the mean and 95% confidence interval. The prior for the alternative model ($\sigma_R=1.3$) is shown on the right in green.
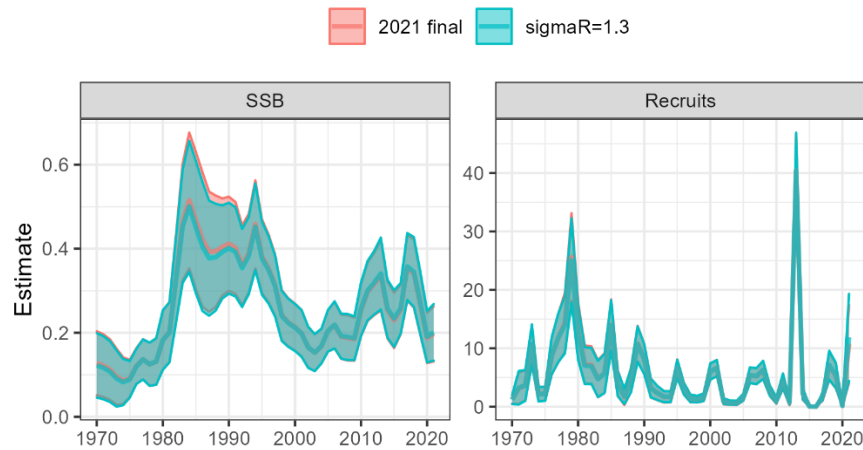

Figure 2. Annual spawning stock biomass and recruitment estimates (lines) and their uncertainty (ribbons).

Recommendation

Model 19.1a is very similar to the base model, only shrinking some of the smallest cohorts closer to the average, but provides several key advantages. First, it brings the model in line with what is done in most

other stock assessments. Second, it has improved statistical interpretation because the deviations in log space can be interpreted as process errors. This is advantageous when moving to a fully state-space approach where these will be integrated out and the process error estimated. Finally, it stabilizes Bayesian estimates when using MCMC because there are extremely long negative tails on the smallest cohorts and this can interfere with the integration algorithm. There appear to be no disadvantages to this change, and so the **authors recommend this update to the model for 2022.**

Model 19.1b: Removing the prior on bottom trawl catchability

In the 2021 SAFE (Monnahan et al. 2021), an analysis was done where successive surveys were dropped to explore their impact on the model. This identified that the NMFS bottom trawl (BT) survey had a large impact on the scale of the stock, and it was noted that this was constrained by a somewhat arbitrary, but informative, prior on the catchability for this survey. It was also noted at the Plan Team that there is a difference between dropping the whole survey vs dropping the prior, and so the Plan Team recommended "…*the author further research [prior on BT q], including conducting a prior sensitivity analysis and potentially looking at applying priors (if available) for other surveys in the assessment.*"

The informative prior is implemented as $\log q_{BT} \sim N(\log(.85), 0.1)$ and was originally developed in collaboration with GAP staff (M. Dorn pers. comm.). Comparing the resulting estimates of the catchability demonstrates that the prior dominates the estimate, and without the prior there is vague information with a mean catchability of 0.75 (Fig. 3).
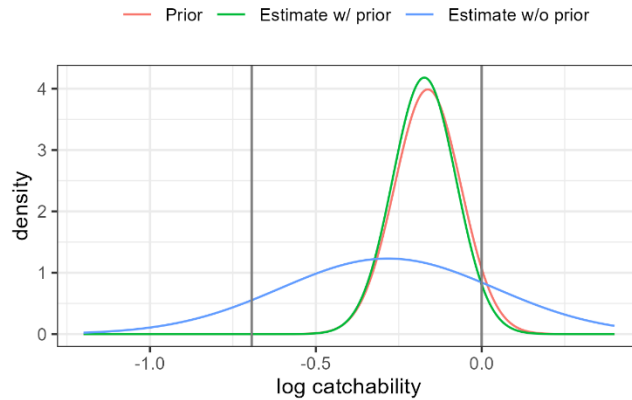

Figure 3. Asymptotic normal distributions estimated for the log of BT catchability vs the prior (mean 0.85). The estimate with without it suggests a lower catchability (mean 0.75).

The resulting estimated SSB have the same trend, but much higher without the BT data, and much closer to the base model when the prior is dropped (Fig. 4).
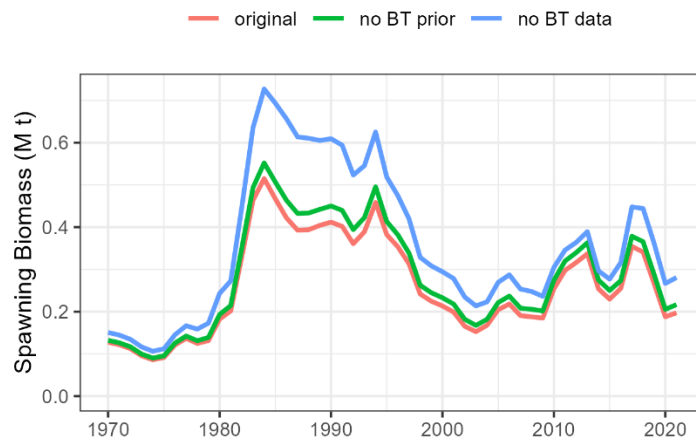
Figure 4. Spawning biomass estimates for models without the BT data or prior compared to the base model.

The more notable difference is in the uncertainty estimates, which are substantially higher (Fig. 5). These differences highlight the impact of the prior to set the scale of the population.
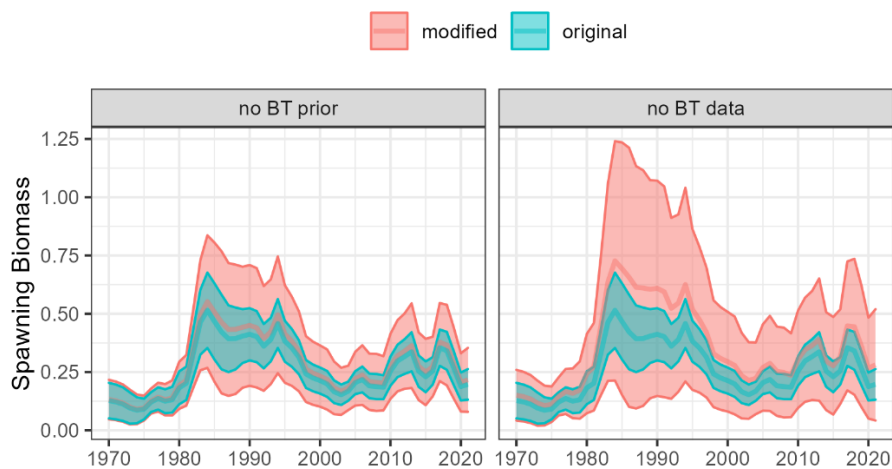

Figure 5. Spawning biomass estimates (lines) and confidence intervals (ribbons) for two modified versions (panels) against the base model (colors).

The median CV for SSB in the base model is 0.144, for no BT prior (proposed model 19.1b) it is 0.276 (a 92% increase), and without any BT data it is 0.363.

Model 19.1b proposes to remove the prior and estimate it freely. The previous analyses demonstrate this for the 2021 model, but the estimate of catchability may not be stable over time, especially as it interacts with time-varying catchabilities for the Shelikof acoustic and ADF&G bottom trawl surveys. This could be a concern if the estimate varies widely annually, as it could drive big changes in scale and potentially management advice. A retrospective analysis was thus done for 19.1b (10 peels), and estimates of catchability compared across years. The results were fairly stable from 2011 to 2015 at around 0.6, but then increased markedly, peaking in 2017 at 1.54 and slowly decreasing to 0.75 in 2021 (Fig. 6).
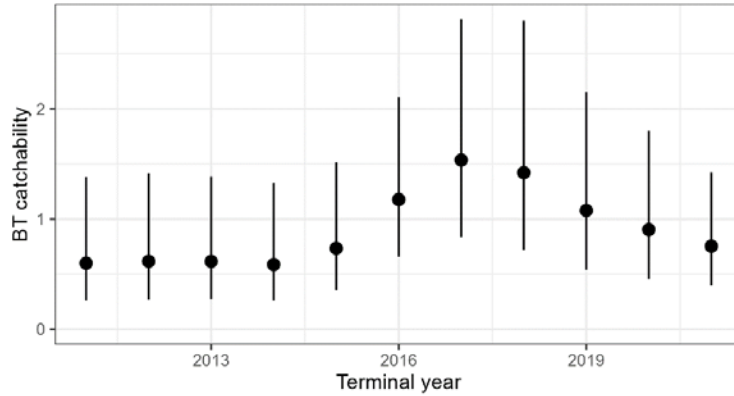
Figure 6. Estimated bottom trawl catchabilities from a retrospective analysis where successive years of data were peeled off to change to terminal year of the assessment. Estimates and 95% confidence intervals are shown.

The 2016-2018 was the period with pronounced divergent trends between the BT and acoustic indices, which likely drives the change in estimates. This change in scale is also apparent in SSB for the retrospective analysis (Fig. 7). Both values for Mohn's rho calculated with 10 peels are relatively small for the base model (0.056) and 19.1b (0.033), although the sensitivity of rho is more sensitive to the number of peels used in model 19.1b (results not shown).
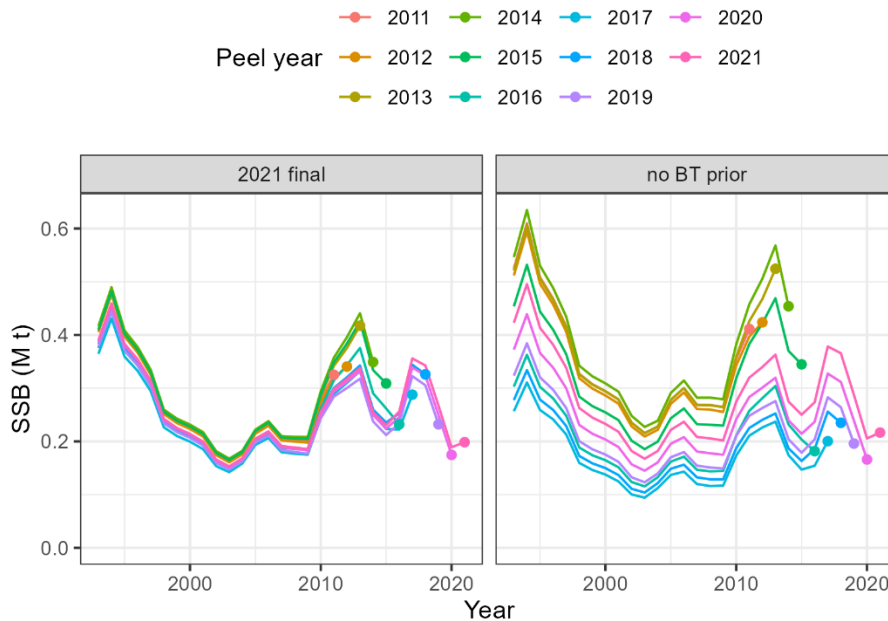

Figure 7. Changes in spawning biomass as a function of peels in a retrospective analysis for the base model and proposed 19.1b alternative. Mohn's rho is 0.056 and 0.033, respectively, with the alternative model having a lower rho despite the larger fluctuations in scale.

Recommendation

The largest impacts from this change are increased uncertainty in the scale of the population, although the mean is fairly similar. This uncertainty likely more accurately reflects the information content in the data, and that is likely the only advantage. The disadvantage will be that the scale will be more sensitive in

future assessments, especially when there are conflicts in scale (catchability) with other surveys. The BT prior serves to stabilize the scale of the model, and thus management advice, which could be seen as an advantage. **The authors defer to the Plan Team for recommendation.**

Model 19.1c: Adding flexibility to fishery selectivity

The GOA Plan Team in its November 2019 minutes recommended the author examine fishery selectivity, as persistent patterns in the catch-at-age residuals may represent artifacts of the selectivity functional form used.

The base fisheries selectivity is double-logistic with time-varying ascending slope and inflection point. Plan Team discussions highlighted the persistent patterns predominantly in the age 4 fish, where long runs of negative and positive Pearson residuals were observed. Monnahan et al. (2021) explored increasing the process error for the ascending inflection point, but found it had no meaningful impact. This suggests that there is something fundamentally limiting by the ascending portion of the selectivity curve. It seems unlikely another parametric form, like the double-normal, would avoid this limitation.

Instead, an unusual but parsimonious approach was taken instead. A single fixed effect representing a constant offset (in logit space) from the parametric curve for age 4 fish was estimated. This allows the time-varying parametric curve more freedom to fit to ages 3 and 5, while only using a single additional parameter. Specifically, if we let $sel(a)$ be the parametric selectivity curve at age a determined by the parameters, then this value is updated given the new unbounded parameter θ_4 as follows:

$$sel(4) = \text{inv.logit}(\text{logit}(sel(4)) + \theta_4) \tag{0.1}$$

Where the function $y = \text{logit}(x) = \log(x) - \log(1-x)$ and inv.logit is the inverse of this function, namely $x = 1/(1 + \exp(-y))$. This series of transformations keeps the selectivity within the range of (0,1), no matter the value for the offset. Also note that a value of $\theta_4 = 0$ reverts to the parametric form.

The model estimates $\theta_4 = -0.985$ (CI: -1.47– -0.5), resulting in a lower value of age 4 selectivity across the time series, which causes a slight increase for age 3 (Fig. 8).
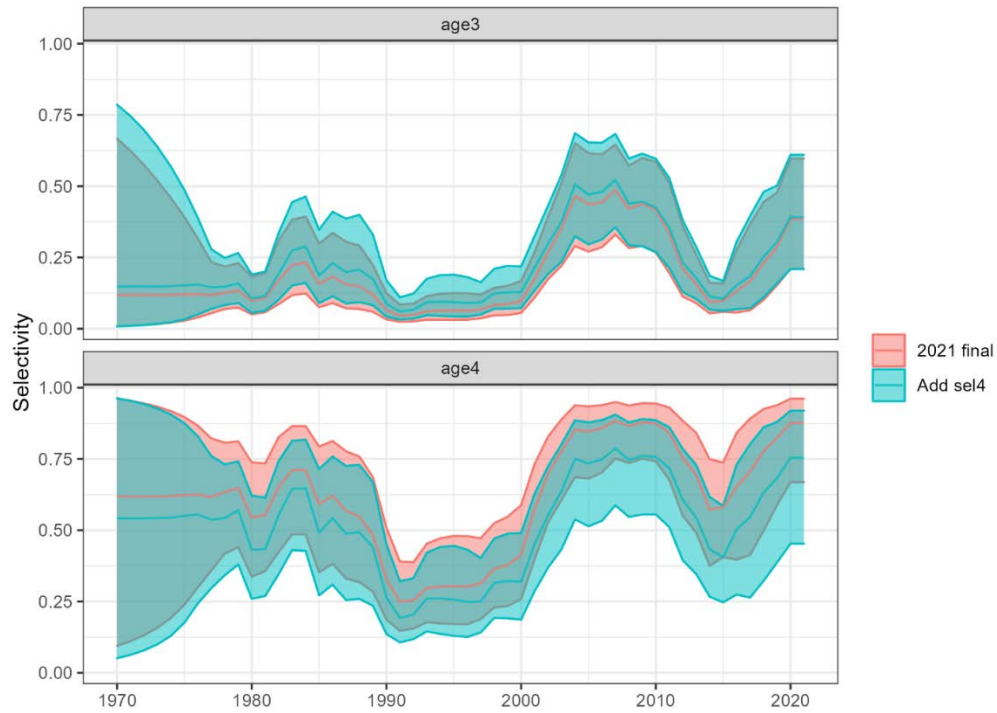
Figure 8. Estimates and confidence intervals (lines and ribbons) for annual fishery selectivity for two ages (panels) and two model versions (colors).

However, due to the time-varying nature of the ascending logistic, the new estimates can be higher or lower than the base model (Fig. 9).
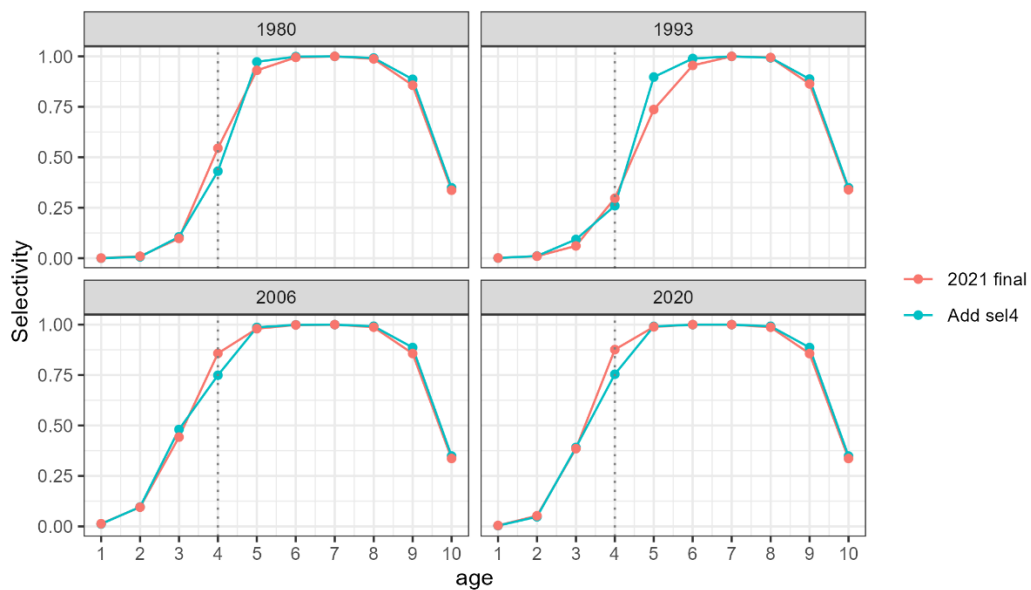

Figure 9. Example selectivity curves for four arbitrary years (panels) for the two model alternatives (colors).

This added flexibility reduces the patterns in age 3 and age 4 residuals (Fig. 10).
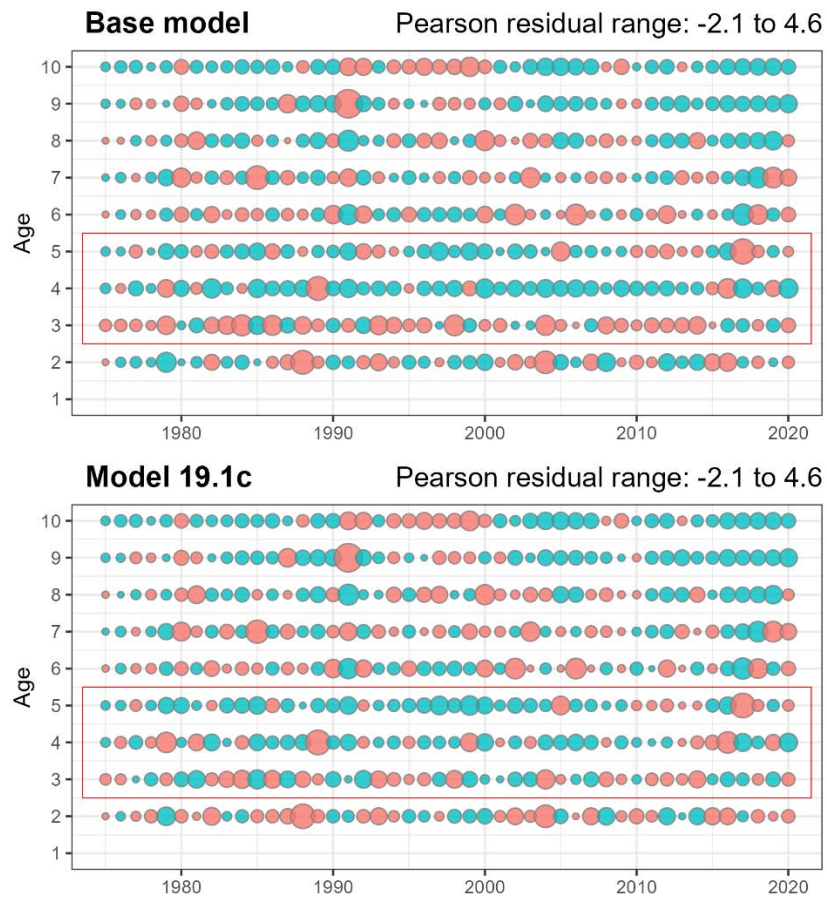
Figure 10. Pearson residuals for the base and alternative models (panels). Blue (negative) and red (positive) represent the sign of the residual while the size reflects the magnitude (range given in upper corner). The red boxes highlight the ages expected to be improved.

This approach adds a single fixed effect (the offset), but reduces the total negative log-likelihood by 8 units, suggesting a substantial improvement to the fits to data. The resulting SSB estimates are very similar (Fig. 11).
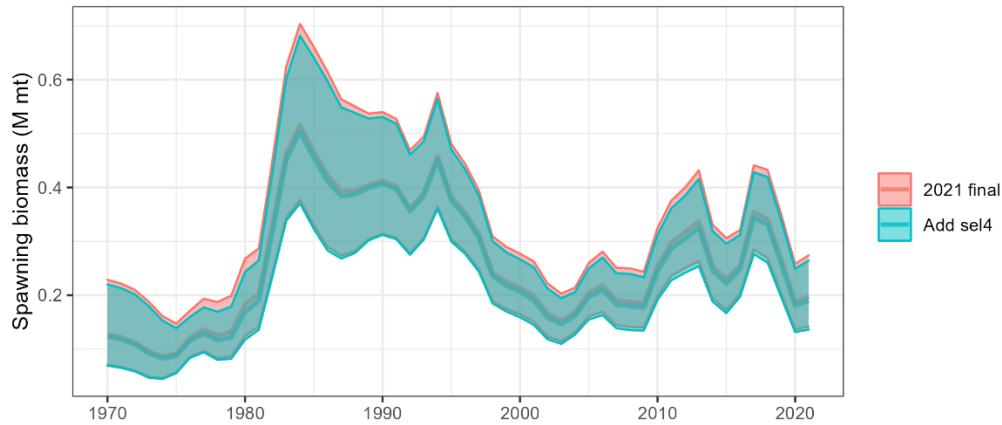
Figure 11. Estimates and uncertainties in SSB (lines and ribbons) between the base model and alternative 19.1c which has additional flexibility in ages 3-5 (colors).

Ages >8 also exhibit patterns of runs that are unlikely due to chance. Several approaches were explored to determine what kind of additional flexibility would be needed to eliminate those patterns. These include an offset on age 9 and adding time-variation to the descending portion of the parametric curve. Neither of these had a meaningful impact on the estimates (results not shown), but did demonstrate that to fit the data well the model needs to have close to constant selectivity for age 9, but very large variation in age 10. Thus, another alternative was explored that kept the parametric form identical for ages <= 9, but added a random-walk in logit space for age 10. The assumed process error was increased until the residual pattern visually improved.

This parameterization was able to improve the residual patterns (Fig. 12), however the resulting estimates for age 9 and 10 selectivity were highly variable (Fig. 13).
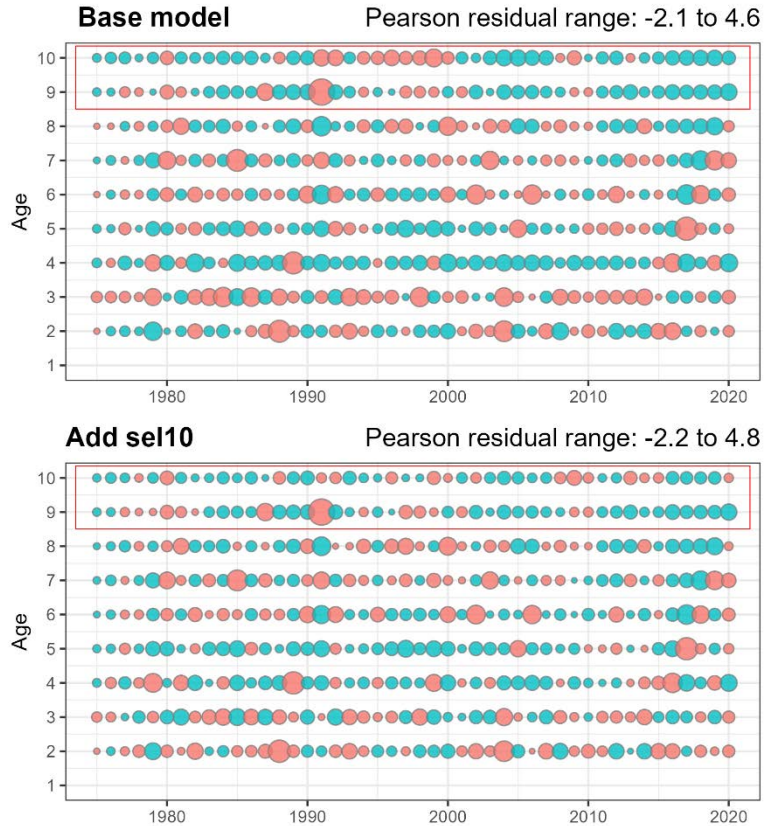
Figure 12. Pearson residuals for the base and alternative models (panels). Blue (negative) and red (positive) represent the sign of the residual while the size reflects the magnitude (range given in upper corner). The red boxes highlight the ages expected to be improved. The alternative model here is an experiment with a random-walk on selectivity at age 10 directly.
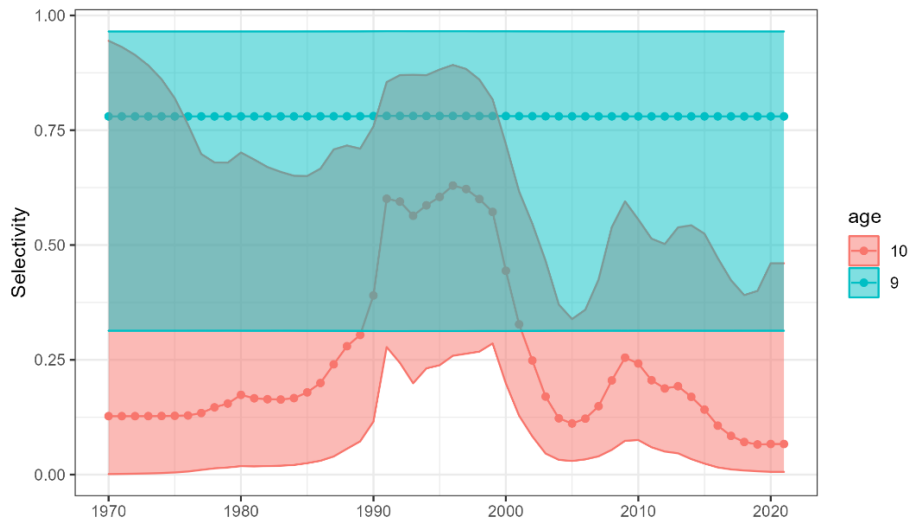

Figure 13. Estimates and uncertainties (points and ribbons) of age 9 and 10 selectivity when a random-walk process was used for age 10 instead of the parametric curve.

This alternative had little effect on SSB, except for a much wider confidence interval for the early years when the random-walk process was highly uncertain due to limited data to inform it (Fig. 14).
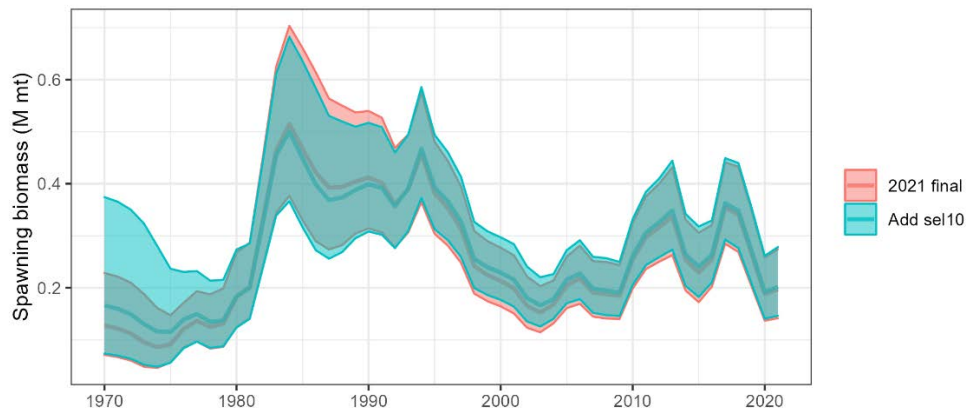

Figure 14. Estimates and uncertainties in SSB (lines and ribbons) between the base model and alternative 19.1c which has additional flexibility in ages 3-5 (colors), plus a non-parametric form for age 10.

Recommendation

It seems clear that the parametric form of the double-logistic is not sufficiently flexible for this stock, as there is too much fluctuation between some adjacent ages (3 vs 4, and 9 vs 10) over time to be captured even with very flexible time-varying parameterizations. Consequently, other parametric forms also seem unlikely to be able to capture this variation. The offset for age 4 approach was parsimonious (a single parameter), visually improved the residuals, and decreased the total NLL by 8 units. It thus seems a worthwhile addition to the model. In contrast, the complexity needed to reduce the patterns in age 9 and 10 fish were not parsimonious and resulted in highly uncertain estimates of selectivity for those ages. **The authors thus define model 19.1c as one that includes an offset to age 4, but no other changes, and recommend that for 2022.**

Updating the area apportionment calculations

The December 2021 the "*SSC suggests simplifying the computations in the Appendix to reflect the new season structure to the extent possible, without changing the underlying methodology. For example, it appears that seasons B1 & B2 (formerly C & D) could be combined as they use the same apportionment scheme.*"

The apportionment table will be simplified where possible to reflect the new seasonal structure. Note that this new structure was not supposed to change apportionment, and that motivates the current table which calculates by the previous four seasons and then sums them together into the new seasons.

A separate issue with winter apportionment is the lack of consistent acoustic surveys in areas outside of Shelikof. The average of the last 3-4 surveys are used for the current year. In some areas the surveys use averages that include data as far back as 2013 (Chirikof and Mozhovoi). An alternative way of getting an estimate of the biomass in each area is to fit a time-series model to the data and use the estimate from this. As proof of concept, an AR(1) time-series model was coded in Template Model Builder (Kristensen et al. 2016). This model fits to data from the 7 areas jointly, estimating a shared correlation parameter. Each area is assumed to start at its individual mean (Fig. 15).
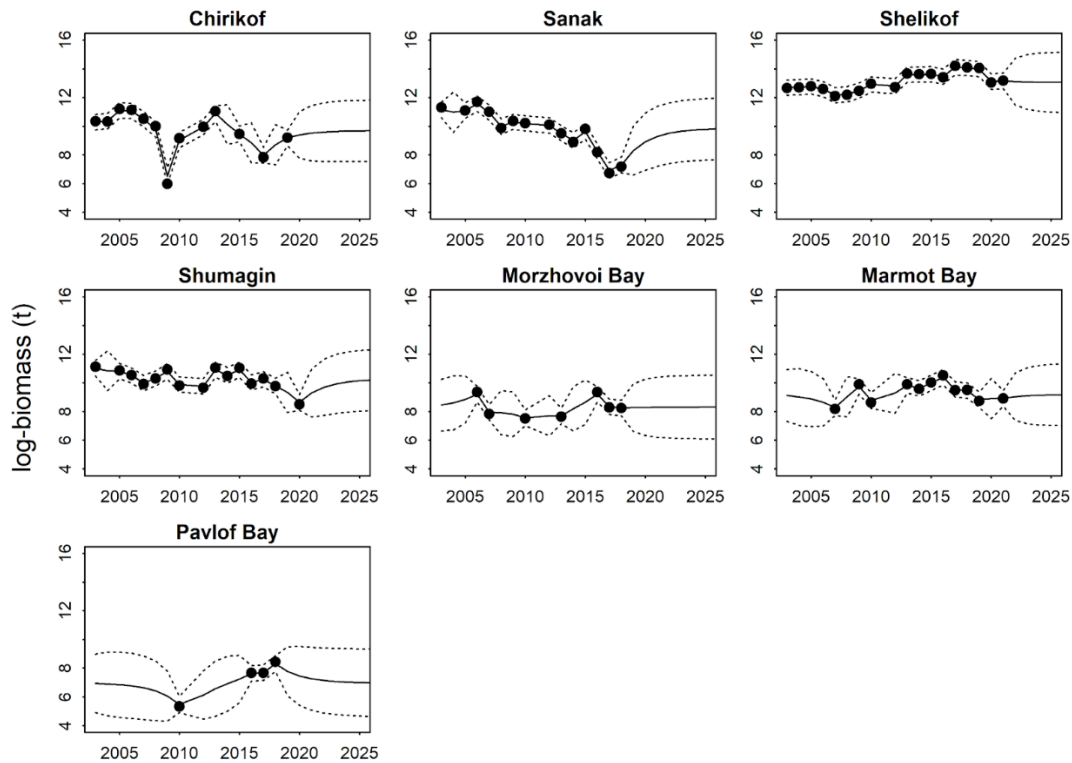
Figure 15. AR(1) time-series estimates (solid line) with 95% confidence intervals (dashed lines) to data (points) from the winter spawning areas surveyed by MACE (panels). Time and logistical constraints preclude consistent sampling, and the model estimates are proposed as an alternative for the winter apportionment calculations. The time-series are projected into the future to demonstrate the behavior in the absence of data (reversion to mean).

There are several advantages worth highlighting. First, uncertainty in the area apportionment can be calculated (via the delta method) and this could be used to help prioritize which areas to survey in subsequent years to minimize this uncertainty. This will likely depend on complex, interacting factors like absolute biomasses, recent trends, and the number of recent surveys per area. Second, as years without data accumulate, the model will naturally revert to the estimated mean of the time-series, which arguably is a better estimate to use than an average of old data. Finally, this model assumes stationarity which may not be true in the face of climate change. But the current approach also does not account for that, as old data will not necessary reflect the current biomass.

During the pandemic most of these areas were not surveyed due to logistical challenges, and future reductions in survey effort may continue this trend. **The authors therefore pose this time-series approach as a more robust alternative to winter apportionment in the face of limited data.**

The 2022 SAFE will also include a second appendix table showing apportionment for the subsequent year as well. This was requested by Obren Davis.

Models under development

Re-evaluating the SE stock distinction

Finally, the SSC encourages the authors and GOA GPT to re-evaluate whether assessing Southeast Alaska walleye pollock as a separate stock is justified or whether the available data support a single, gulf-wide stock assessment. This evaluation may also benefit from considering recent studies on the genetic structure of walleye pollock across Alaska and the North Pacific

Recent genetic studies have not conclusively identified the stock substructure sufficiently to justify changing to a gulf-wide assessment. Samples from the SE GOA were recently taken and analyses are underway, but the AFSC genetics team needs more time to complete them (I. Spies pers. comm. Sep '22). This issue will be revisited once those analyses are completed.

Updating structure of catchability for the Shelikof acoustic survey

In December 2021 the SSC highlighted the need to examine how catchability for the winter Shelikof acoustic survey.

The SSC supports future research to identify the optimal level of constraint on among-year variation in Shelikof Survey catchability (q), including the potential to estimate the process error variance internally within the assessment model.

The SSC reiterates its recommendation from December 2020 to explore the use of covariates related to the timing of the survey to inform survey catchability in the Shelikof Strait survey. For example, the difference in timing between peak spawning and mean survey date or, alternatively, the proportion of mature fish in the survey, are likely to inform time-varying catchability in the survey.

Currently the winter Shelikof acoustic survey catchability is modeled as a random walk with assumed process error. The original logic was that some of the stock spawned outside of Shelikof Strait and thus were unavailable to the survey. Fish tended to spawn in other areas with some consistency, so a random walk on catchability was implemented to account for variation in spatial availability. Several overlapping efforts were done to explore alternative catchability structures. None of these are proposed for 2022, but are summarized here for Plan Team feedback.

First, the ADMB model is unable to estimate the magnitude of the process error used in the random-walk process using penalized maximum likelihood. Instead, this parameter was estimated in two ways: integration with Markov chain Monte Carlo (MCMC) and using marginal maximum likelihood via TMB in the WHAM model.

Estimating the process errors directly

MCMC has a long history of implementation in hierarchical (random effects) models, and in particular the Hamiltonian Monte Carlo family of algorithms is well-suited to such models (Betancourt and Girolami 2015, Monnahan et al. 2017). The no-U-turn sampler (NUTS) algorithm is implemented in ADMB via the 'adnuts' R package (Monnahan and Kristensen 2018) and thus the base model, and provides a promising path for full integration of random effects and their variances. The primary challenge is obtaining a good dense mass matrix (a tuning parameter of NUTS). Typically this is the estimated covariance (in unbounded space), but with process errors the model was not estimable and thus no covariance matrix available. Instead, the following procedure was used:

1. Put very tight normal priors at a relatively high value on the process errors (sigma) and estimate model and covariance matrix. This gets covariances for all fixed effects and approximate ones for the random effects (albeit restricted), while also containing elements for sigma.

2. Take the temporary priors off in the TPL and recompile, but do not re-estimate the model.

3. Read in the original covariance matrix from file, and modify the matrix rows/cols to have marginal elements of 1 and zeroes elsewhere.

4. Run the `sample_nuts` function with `metric=cov.new` as typically done. This will write the modified covariance matrix to file and use that for sampling.

This procedure is relatively straightforward to implement, and R code is available upon request. Before applying it to both time-varying catchabilities (the ADF&G survey as well), several small parameterization issues were solved (not shown). Then, 5 NUTS chains with 2000 iterations (250 warmup) were run in parallel, initiating from the mode. The chains took 13 minutes to run and passed all convergence checks with enough effective samples for inference.

These results were compared to a bridged version of the assessment in the WHAM modeling framework, the same one above used to estimate σ_R. The two ways of estimating these process errors were very similar (Fig. 16) and substantially larger than the assumed value in the base model. This demonstrates that process errors are estimable in ADMB models using MCMC.
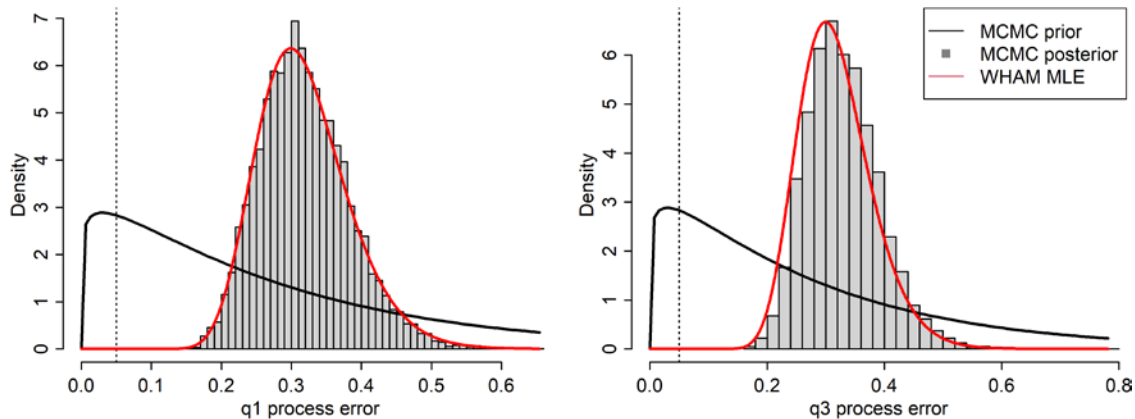

Figure 16. Estimates of the random-walk process errors in the model for the Shelikof survey (q1, left panel) and ADF&G (q3, right panel). Bayesian posteriors are compared to marginal maximum likelihood estimates from WHAM. The value assumed in the base model is shown as dashed vertical line.

Despite the much larger process errors (and flexibility in catchabilities; Fig. 17) the spawning stock biomass is relatively unchanged (Fig. 18).
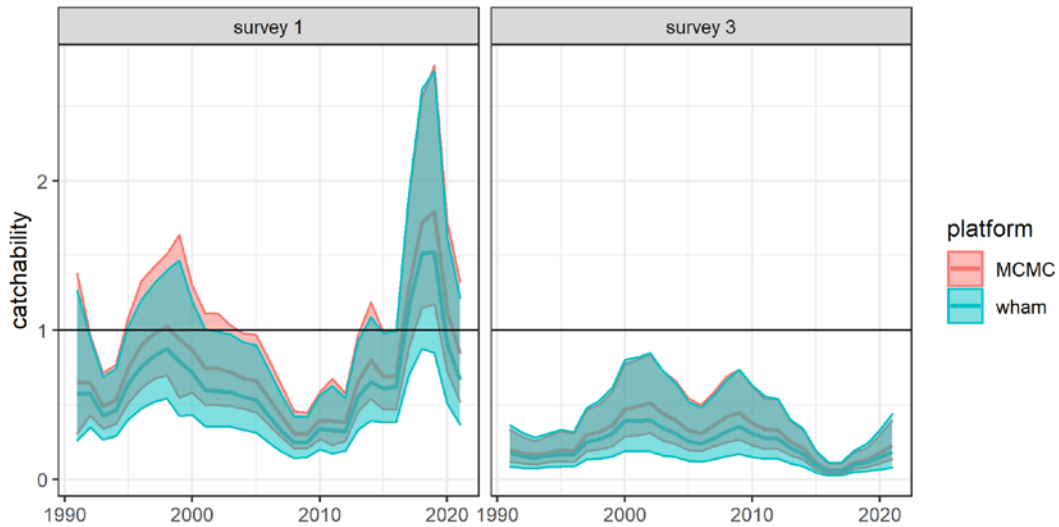
Figure 17. Estimates and uncertainties (lines and ribbons) for catchabilities using two approaches to estimating the process errors (colors).
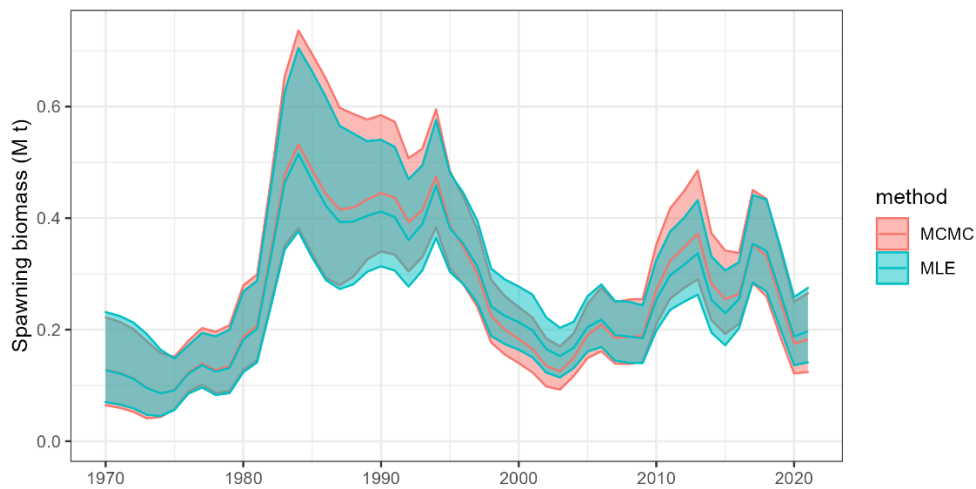

Figure 18. Estimates and uncertainties in SSB (lines and ribbons) between the base model and one with the process errors estimated via MCMC.

Incorporating timing covariates

The second approach investigated is whether the timing of the winter Shelikof survey relative to peak spawning can help explain some variation in catchability. The driving hypothesis is that if the survey is too early or too late then the spawning fish will not have arrived or have already left, thus reducing temporal availability. Incorporating these covariates measuring this timing mismatch would therefore be able to better fit the Shelikof index. Two timing metrics are used: (1) 'Fem30p' is the proportion of female fish over 30cm that were in a spawning or spent stage and (2) 'mismatch' uses the estimated spawning dates from larval surveys to calculate the relative temporal lag between the Shelikof survey and inferred timing of peak spawning. This work is lead, and has been presented previously by, Dr. Lauren Rogers. **The advances this year were to incorporate the spawn timing mismatch covariates into a WHAM model and explicitly test this hypothesis in a more robust statistical framework.**

WHAM allows fitting a time-series smoother to the covariates jointly with the model, and then uses the expected values (instead of the raw data) in the relationship with catchability (Stock and Miller 2021). This approach thus allows fitting noisy time-series or ones with missing values. Catchability was modeled as linear on the log scale:

$$log(q_y) = \mu + \tau_y + \beta_1 E_y + \ldots + \beta_n E^n_y$$

where μ is a global intercept, τ the annual random walk process, and β_i the estimated fixed effect coefficient for the ith term of a polynomial against the environmental covariate E^i_y. Here the focus was on each of the two timing covariates at a time and only explore a linear effect. The random-walk component, ostensibly attempting to control for spatial variance in availability, can also be left in or taken out. Thus, three versions of the model were fit to each covariate: just the random-walk (RW), just the covariate (Cov) and both components simultaneously (RW+Cov).

The models with just the covariates were also able to capture a large amount of the variation in catchability (Fig. 19). But the models that included both components were selected by AIC, and there was a 56.3% and 74.1% reduction in random-walk process variance w/ added covariate (Table 2). This was not true for white noise covariates (results not shown). Together, this suggests there is statistical evidence for a signal in the timing covariates and the random-walk component, and there is a clear improvement to the fit to the data (Fig. 20).

Table 2. Results for the two timing covariates. For the 'RW' models, the covariate time-series smoothers were fit internally but not linked so that AIC is comparable.

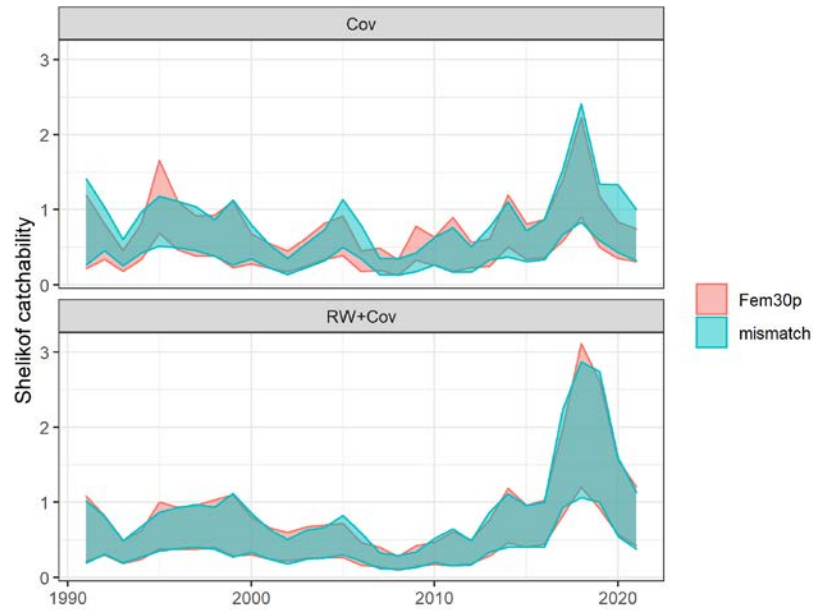|	Fem30p dAIC	mismatch dAIC	Fem30p RW sigma	mismatch RW sigma
RW	8.6	12.7	0.34	0.34
Cov	13.8	7.1	--	--
RW + Cov	0.0	0.0	0.22	0.17

Figure 19. Catchability esitmates (lines) and uncertainties (ribbons) for models with the covariate and with and without the random-walk process (panels) for the two timing covariates (colors).
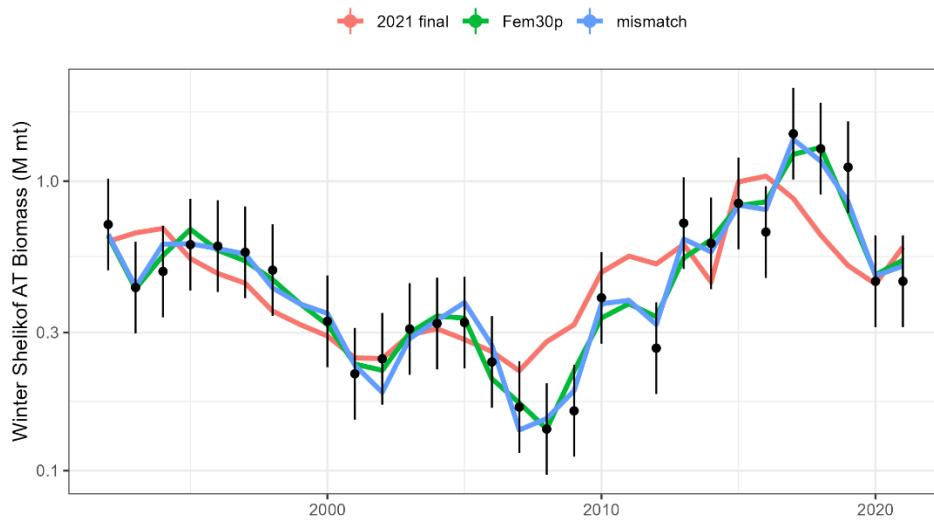

Figure 20. Expected Shelikof indices by model (color) and the data (points and vertical lines). The two timing covariate models are the best-fitting versions that also include the random-walk process.

Despite the differences, there was little change in the estimated SSB between the two covariates for the selected models (Fig. 21).
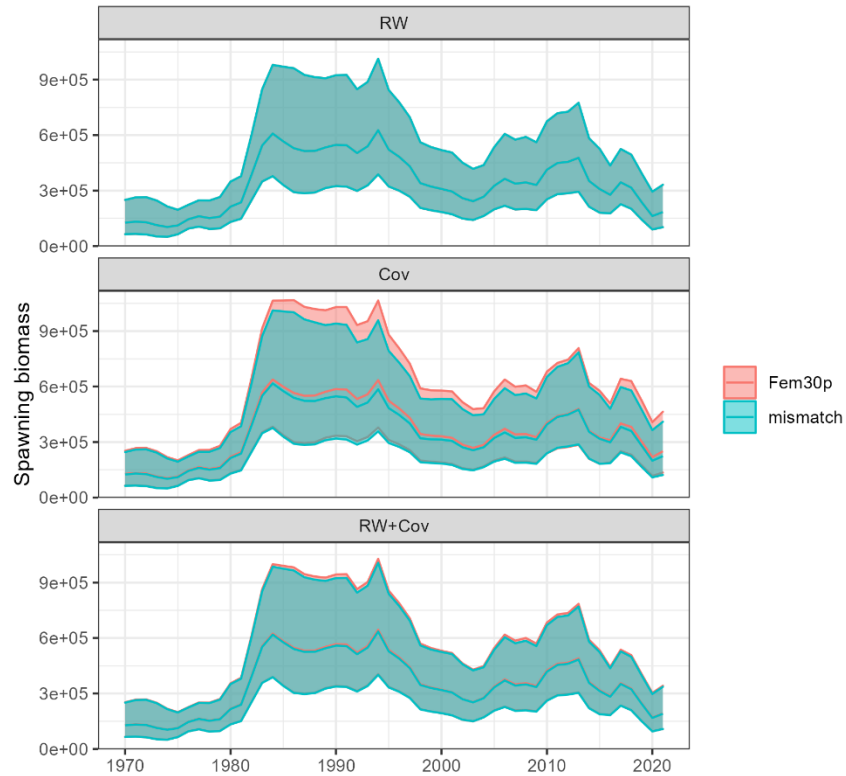
Figure 21. Estimates of spawning biomass (t) for three different catchability structures (panels) and two timing covariates (colors). These estimates differ from the base model because the WHAM version is missing important features (e.g., BT prior, ageing error matrices).

It is not possible to directly compare SSB estimate to the base model because some features used in ADMB are not possible in WHAM and so they are not comparable. **<u>Despite this, some preliminary conclusions can be made in order to stimulate discussion from the Plan Team.</u>**

- WHAM provides a powerful framework to explore covariate linkages to key parameters, and here it was clear there was a signal in both timing covariates.
- Both the smoother and covariate were significant, and the covariate reduced the process variance substantially. It is hypothesized that WHAM has thus partitioned the variability in catchability into spatial (RW) and temporal components (Cov).
- A random-walk only model can achieve the same fit, but requires a higher process error and thus uses more DF, and has no mechanism.
- This approach could be incorporated into the ADMB production assessment but would not be straightforward.

The authors consider this an exciting and potentially very valuable result for the assessment. But for now, this research is ongoing and is expected to be published in 2023.

References

Betancourt, M., and M. Girolami. 2015. Hamiltonian Monte Carlo for hierarchical models. Current Trends in Bayesian Methodology with Applications:79.

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., M. W. Dorn, A. L. Deary, B. E. Ferriss, S. K. Shotwell, A. Tyrell, and S. Zador. 2021. 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.

Monnahan, C. C., J. T. Thorson, and T. A. Branch. 2017. Faster estimation of Bayesian models in ecology using Hamiltonian Monte Carlo. Methods in Ecology and Evolution **8**:339-348.

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.