Center for Independent Experts (CIE) Independent Peer Review Report

BSAI Rockfish – Pacific Ocean perch

Virtual Meeting. May 9-13, 2022

Dr. Noel Cadigan

Centre for Fisheries Ecosystems Research
Marine Institute of Memorial University of Newfoundland
St. John's, NL. Canada

Executive Summary

- The Bering Sea/Aleutian Islands Pacific Ocean perch (POP) stock assessment represents the best available science.
- The assessment model integrates relevant productivity information (surveys and fishery catches, and their length- and age-compositions) to estimate stock size and fishing mortality rates.
- There is a retrospective pattern in assessment model estimates of stock size that is related to recent survey biomass indices that were higher than the model could predict.
- Despite the retrospective pattern, the status evaluations are reliable.

Important research recommendations are:

- Consider using a stochastic initial equilibrium age distribution in the BSAI POP assessment model.
- Fit a model with a more flexible initial age-distribution (e.g., stochastic) and an asymptotic fishery age selectivity pattern to demonstrate the lack-of-fit that a historic domed selectivity pattern fixes.
- Further investigate change in M.
- Continue research to improve modelling of age and year variation in F.
- Provide a table of inputs for $F_{X\%}$ calculations. Describe what recruitment value R is used to derive $B_{40\%} = R*SPR(F_{40\%})$.
- Provide a retrospective analysis of current status evaluations.
- Consider fitting the BASI POP model to survey abundance rather than biomass indices.

Background

The stock assessment for Bering Sea/Aleutian Islands (BSAI) Pacific Ocean perch (POP) provides the scientific basis for the management advice considered and implemented by the North Pacific Fisheries Management Council. An independent review of these integrated stock assessments was requested by the Alaska Fisheries Science Center's (AFSC) Resource Ecology and Fisheries Management (REFM) Division. The goal of this review was to ensure that the stock assessments represent the best available science to date and that any deficiencies were identified and addressed.

The Review Panel (i.e., Panel) was composed of three independently appointed Center for Independent Experts (CIE) reviewers (Dr. N. Cadigan, Canada; Dr. G. Tingley, New Zealand; Dr. M. Cieri, US), and the chair was Dr. Pete Hulson from the AFSC. Main assessment documents and presentations were provided and given by Paul Spencer, and specific topics were presented by N. Laman, J. Thorson, R. Rameriz, and D. Anderl. The support of all these scientists and staff to the Review Panel process is gratefully acknowledged.

The CIE reviewers were required to have a working knowledge and recent experience in the application of statistical age-structured stock assessment methods in general and specifically application of ADMB in stock assessment. Additionally, the CIE reviewers were required to have:

- Expertise with measures of model fit, identification, uncertainty, forecasting, and biological reference points.
- Familiarity with federal fisheries science requirements under the Magnuson-Stevens Fishery Conservation and Management Act.
- Familiarity with groundfish fisheries and management.
- Working knowledge of trawl survey design and estimation of stock biomass
- Excellent oral and written communication skills to facilitate the discussion and communication of results.

The primary tasks for reviewers were:

1. Review the background materials and reports prior to the review meeting.
2. Attend and participate in the panel review meeting.
3. Provide and independent peer-review report.

Role of reviewer

On April 28, 2022, the Panel chairperson provided the link to a webpage where assessment documents and supporting materials would be provided. These documents are listed in **Appendix 1**. Some of the document links did not work and I eventually got the material on May 4 via a google drive. However, I had sufficient time to review all material before the panel meeting. I reviewed presentations and reports (see Appendix 1) and participated in the discussion of these documents, in accordance with the SoW and ToRs (see **Appendix 2**). This CIE report is structured according to my interpretation of the required format and content described in Appendix 2.1.

Summary of findings

I thank the POP assessment team for their hard work and willingness to respond to Panel requests.

ToR 1. Evaluation of the data used in the assessments, specifically trawl survey estimates of abundance, and recommendations for processing data before use as assessment inputs.

Aleutian Islands survey

There was a somewhat abrupt increase in the Aleutian Islands (AI) NMFS AFSC survey biomass estimates (Fig. 1) which seemed to be the cause of retrospective patterns in assessment model results. The average biomass during 2010-2018 was 968 845 kt, which was 1.82 times the average value during 1991-2006 (531 610 kt).

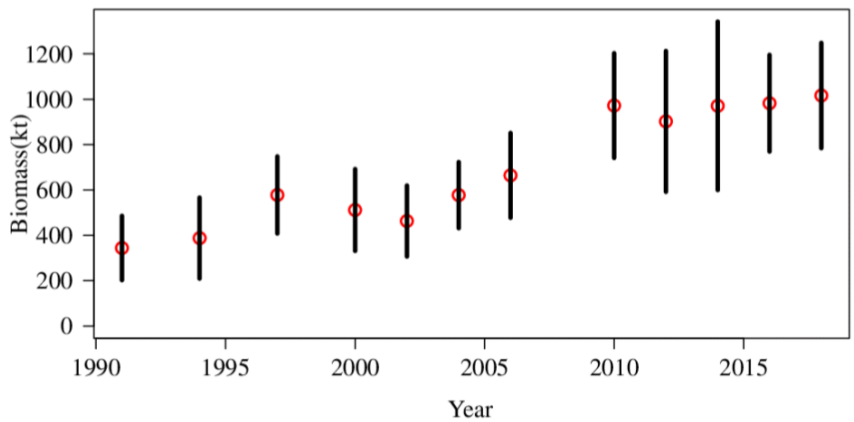

Figure 1. Aleutian Islands survey biomass time series.

The survey protocols were presented by N. Laman. The Panel asked if changes in survey protocols since 1990's warrant standardization? N. Laman responded that the one protocol change with the greatest potential impact was the switch from a tilt meter to an accelerometer in 2010. This change affected on/off bottom times but differences between the two approaches were not large. Hence, overall, N. Laman concluded that the need for standardization seemed low.

Issue 1: The replacement of the bottom contact sensor with an accelerometer happened at the same time (2010) when the survey biomass index shifted up.

Recommendation 1. The old bottom contact sensor and the newer accelerometer should have been used in all tows during the 2010 survey to get data on differences in estimates of time on bottom and tow distances from the two approaches. The survey biomass could have been calculated using the old survey protocols and the new ones, to see if this change in sensors made any difference. It will be useful to quantify the potential impact of this change in survey protocols.

As part of the presentation on assessment data inputs, P Spencer provided a table of AI survey biomass estimates by depth zones. The fractions of total biomass in each depth zone (Fig. 2) demonstrate that the overall depth distribution of biomass was almost identical in 2006 and 2010, but since then the biomass has shifted somewhat to the 100-200 m depths, and in 2018 to the 300-500 m depths. However, I see no reason why these changes in depths could be responsible for the increase in AI survey biomass in 2010. The POP assessment scientists were confident that the AI survey covered the POP distribution in this region. Little POP biomass was expected to be outside of the survey region.

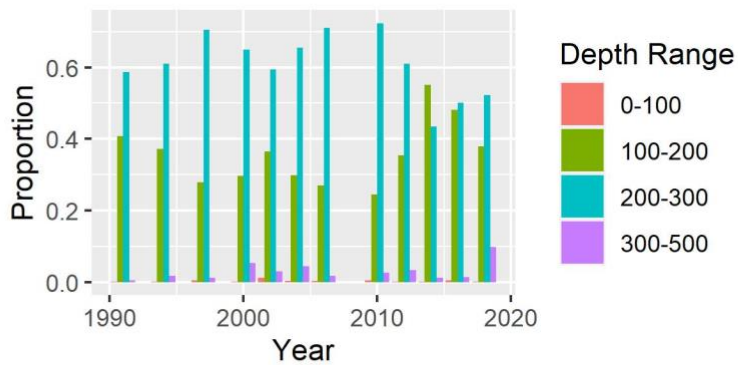

Figure 2. Annual proportion of Aleutian Islands survey biomass in 4 depth bins.

Spatial changes in survey biomass (Fig. 12.5 in Spencer and Ianelli. 2020) are also provided in Figure 3. Although this figure is difficult to interpret, it does not suggest that the increase in survey biomass in 2010 was associated with a small number of strata. The draft POP assessment document provided additional information: "Of the 45 AI survey strata, 79% of the 2018 population estimate was contained in 10 strata, with at least one of these ten strata occurring in each of the 4 major strata regions (i.e., 5 in the WAI, 1 in the CAI, and 2 each in the EAI and SBS). In 9 of these strata, the average population estimate from the 2010-2018 surveys exceeded population estimates from the 1991-2006 surveys. The average value for this ratio of abundances was 1.82 in the top 4 strata for 2018 population abundance"

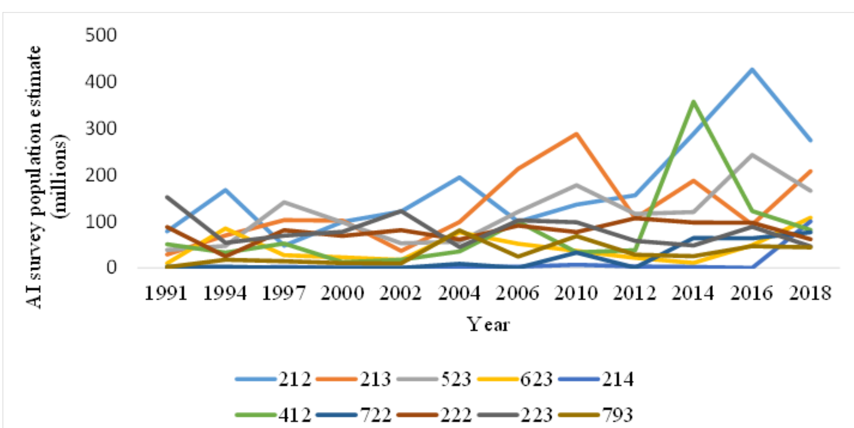

Figure 3. Ten strata with the largest abundances in the 2018 Aleutian Islands survey. In 9 of the 10 strata the abundance increased.

The 2010 shift in survey biomass could be due to a change in stock abundance and/or a change in average individual weight of POP. The stock assessment model used constant weights-at-age (see Fig. 9), but the assessment document indicted that "observed body condition has been lower than the survey mean since 2012". The document did not specify if the change in condition was based on whole weights or gutted weights. Nonetheless, this statement does not indicate that the increase in AI survey biomass in 2010 was associated with an increase in individual body weights. P Spencer also provided a comparison plot of AI survey biomass and abundance (Fig. 4). The biomass and abundance trends were similar overall, but there are some subtle differences. The biomass in 2010 was 1.46 times the 2006 value, whereas the 2010 abundance was only 1.3 times the 2006 value, based on my interpolations from Fig. 4. I consider this further for ToR 5.

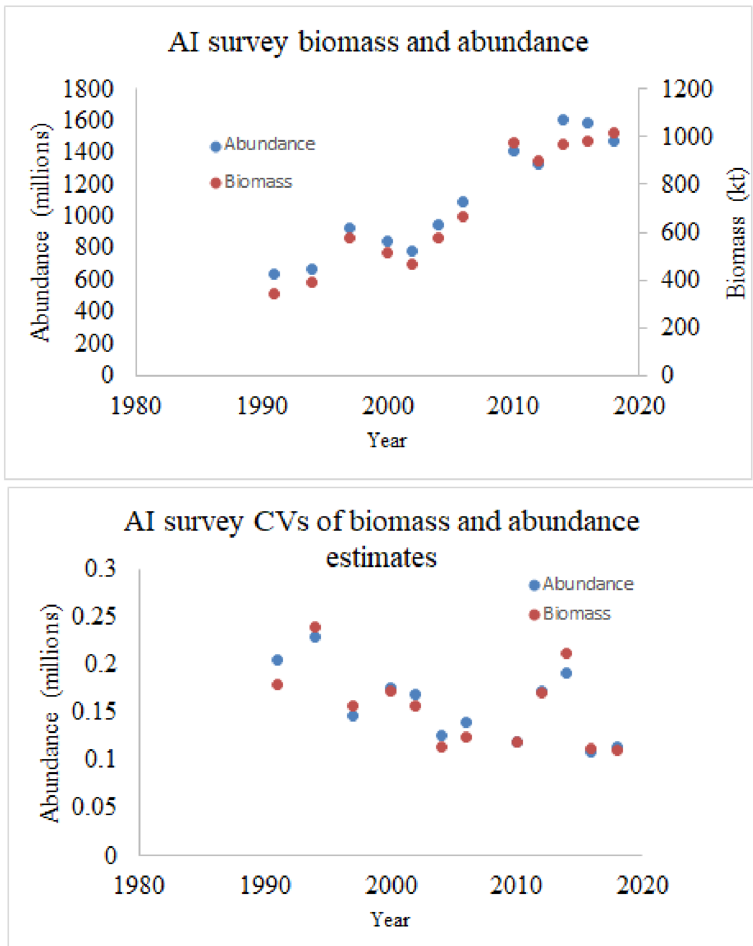

Figure 4. Aleutian Islands survey abundance and biomass estimates (top panel) and CVs (bottom panel).

Declines in the proportion of tows with no catch in 4 subareas (WAI, CAI, EAI and SBS) and an increase in the area associated with trawl tows contributing 95% (D95%) of abundance estimate (Fig. 12.4 in Spencer and Ianelli. 2020) also indicate an increase in POP size in the AI survey region.

The 2010 increase in the AI survey biomass could be caused by one or a few strong cohorts, but the issue with this assessment is that the composition data does not indicate this. Hence, there may be a conflict between the AI survey biomass indices and the composition data.

I find standardized proportions at age (SPAY) plots (e.g., Kumar, 2020) useful for summarize cohort information from age composition time-series. These plots are useful for checking which cohorts are relatively strong; however, they cannot be used to compare the size of recent cohorts relative to earlier years. The SPAY statistic, $p_{a|y}^{std}$, is:

$$p_{a|y}^{std} = \frac{p_{a|y} - \bar{p}_a}{n^{-1}\sum_y (p_{a|y} - \bar{p}_a)^2}, \bar{p}_a = n^{-1}\sum_y p_{a|y},$$

where $p_{a|y}$ is the proportion at age *a* in year *y* and there are data for *n* years.

A SPAY plot for the AI survey age compositions is shown in Fig. 5. Note that a small bubble means near-average. The 1984-1989 cohorts consistently tracked through the age compositions and their proportions-at-age were above average in most survey years during 1991-2018. However, these are not the cohorts that would have contributed to the 2010 increase in survey biomass because the A50 for the survey is estimated to be 6.14 years and ages 10 and older are almost fully selected by the AI survey; hence, these cohorts were fully selected in the 2000 survey. The 2010 increase would have been caused by cohorts born since around 1998. The 1996-2000 cohorts, and perhaps the 1995 and 2001 cohorts, were usually above average but not consistently so. Hence, these data do not provide evidence of a few strong year classes that could have produced the 2010 increase in survey biomass. The 2004 and 2005 cohorts were below average in the 2010 and 2012 surveys but have been a little above average in the 2014, 2016, and 2018 surveys. Recent cohorts (since 2010) seem weak.

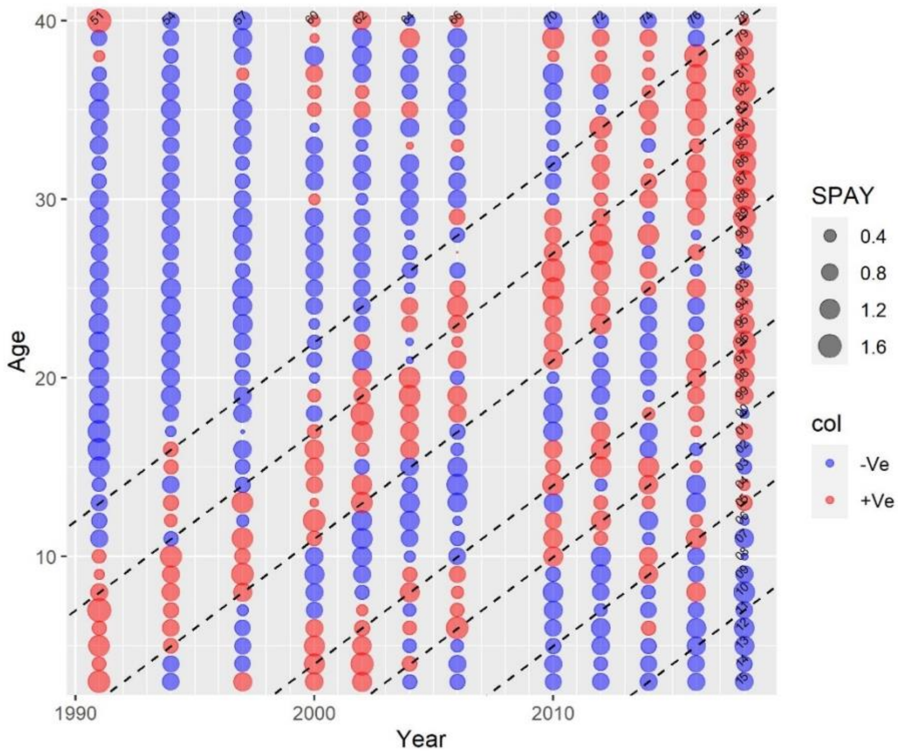

Figure 5. SPAY plot for the Aleutian Islands survey age compositions.

Issues 2. The Panel was not provided information on the length compositions of the AI survey catches. This could be useful to corroborate the patterns in Fig. 5. I realize the age compositions are derived from survey catch-at-length and an age-length key.

Recommendation 2. Provide time-series of the length compositions of the AI survey.

Eastern Bering Sea slope survey

This survey was conducted in 6 years (2002, 2004, 2008, 2010, 2012, and 2016; Fig. 6). Survey biomass increased, especially in 2016, although the CV for the 2016 was high. The age compositions for this survey (Fig. 7) did not track cohorts as well as the AI survey.

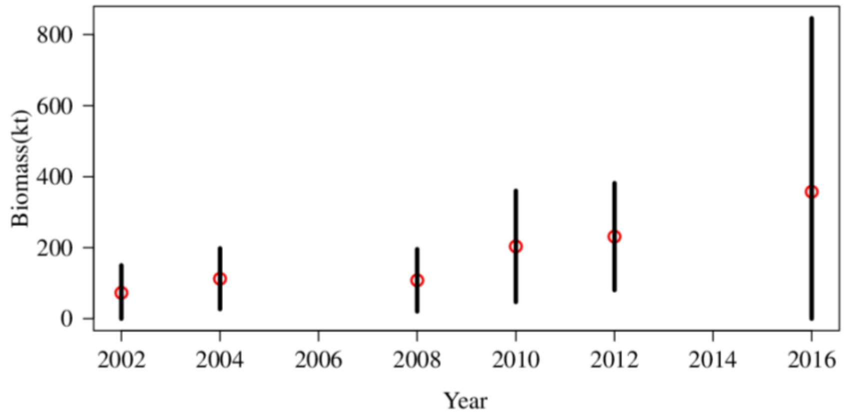

Figure 6. Eastern Bering Sea slope survey biomass time series.

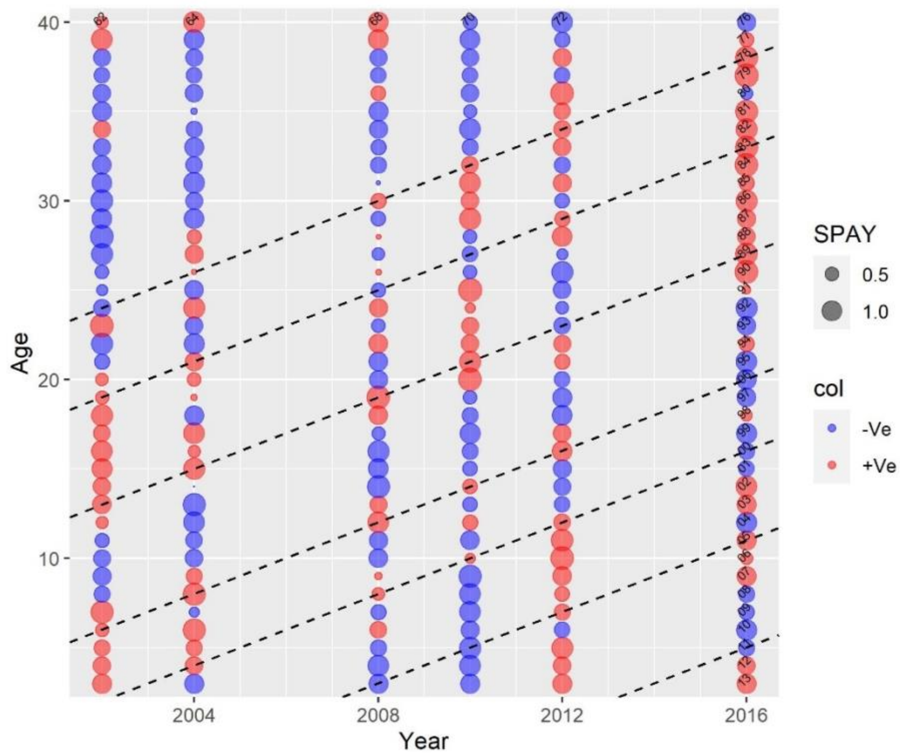

Figure 7. SPAY plot for the Eastern Bering Sea slope survey age compositions.

Length at age

The main assessment document described that "Aleutian Islands survey data from 1991 through 2018 were used to estimate growth curves. The resulting von Bertalanffy growth parameters were Linf = 41.51 cm, k = 0.14, and t0 = -1.311. Growth information from the Aleutian Islands was used to convert estimated numbers-at-age within the model to estimated numbers-at-length." Weight-at-age was also derived from VonB estimates of length-at-age and an estimated weight-at-length model.

We did not review the details of this, but the growth model was estimated using mean length-at-age, presumably to account for the length-stratified age sampling prior to 2016. There are better approaches available (e.g., Perreault et. Al., 2020) but I suspect they will not produce much different estimates of the VonB parameters because Linf is very well identified for POP. Likewise, ageing errors will produce some bias in VonB parameter estimates (i.e., over-estimate k and under-estimate Linf) but the ageing errors seem fairly small for POP less than 20-year-old, and errors for older ages will not matter much in terms of estimating VonB parameters because at these older ages the fish have almost reached Linf.

If the surveys are length-selective rather than age-selective then this will also introduce bias in VonB parameter estimates because at young ages the survey will catch the faster growing fish and you end up with a biased sample of length at young ages. Although it is possible to account for this using individual age and length observations (e.g., Cope and Punt, 2007; Candy et al., 2007; Zheng et al., 2020; Frater and Stefansson, 2020), the application of these methods is not straightforward. An age- and length-structured stock assessment model like SS3 can simultaneously estimate VonB model parameters and account for length-selectivity but application of such a model is also not straightforward.

A plot of mean length versus year, at ages 5, 10, 20, and 30, suggested very little annual variation in size at age. Plots of annual estimates of the VonB k parameter indicate a declining trend but this may not be statistically significant. Estimates of k are often negatively corelated with estimates of Linf. Spatial variation also did not seem large (Table 1) and differences in estimates may not be statistically significant.

Recommendation 3. Conduct a spatiotemporal analysis of variation in size-at-age. Von Bertalanffy models with random year and area effects are an option, but more flexible models (e.g., Cadigan, 2020, incl. TMB code) could be useful.

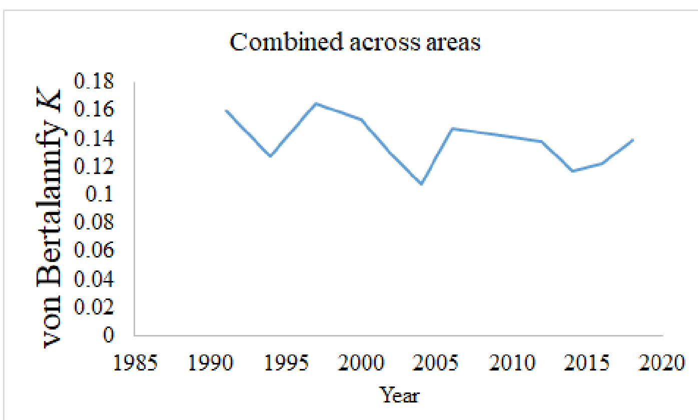

Figure 8. Annual estimates of the Von Bertalanffy k parameter**.**

Table 1. Von Bertalanffy parameter estimates for 4 AI survey subareas and the EBS slope survey.

Area	n	Linf	K	tzero
SBS	784	41.95	0.15	-1.19
EAI	793	42.27	0.14	-0.99
CAI	786	42.06	0.15	-1.06
WAI	784	41.77	0.12	-2.02
EBS slope	2424	43.21	0.13	-0.68

An age-length transition matrix was calculated externally and used to convert model predicted numbers-at-age to number-at-length. It was not clear to me how this transition matrix was calculated.

Issues 3. Some of the variation in size-at-age will be related to age measurement errors, which is also included in the assessment model. This source of uncertainty seems to be included twice.

Recommendation 4. A more thorough analysis of variation in size-at-age is needed, including separating the effects of individual variation in growth and age measurement errors.

Weight at age

It was not completely clear how weights-at-age were derived. Spencer and Ianelli (2020) only described that this was based on estimation of length-at-age and weight-at-length. The simplest way to do this is $W(age) = W\big(L(age)\big) = \alpha L^{\beta}(age)$ and if $L(age) = L_{\infty}[1 - \exp(-k(age - t_o)\}]$ then:

$$W(age) = W_{\infty}[1 - \exp(-k(age - t_o)\}]^{\beta}, \quad (1)$$

where $W_{\infty} = \alpha L_{\infty}^{\beta}$. However, this simple approach does not account for variation in length-at-age and will under-estimate mean weight-at-age because $\beta \gg 1$. This is illustrated in Fig. 9 and described as follows. The assessment document indicated that variation in length-at-age had a constant coefficient of variation (CV) that was a polynomial function of age. However, for illustration purposes I used a simple linear approximation of the CV that decreased from 0.15 at age 3 to 0.07 at age 15 and remained at 0.07 for ages 16-40. The assessment document was not clear about what the distribution of length-at-age was, but I think it was Normal.

If the distribution of length-at-age is lognormal so that the standard deviation of the distribution of log-length-at-age is approximately equal to the CV then the expected value of weight-at-age can be derived using the moment generating function of a Normal random variable, $M_X(t) = E(e^{tX}) = \exp\,(\mu t + \sigma^2 t^2/2)$. In this case:

$$W(age) = W_{\infty}[1 - \exp(-k(age - t_o)\}]^{\beta} \exp\left(\frac{CV^2\beta^2}{2}\right). \quad (2)$$

The bias term $\exp\,(CV^2\beta^2/2)$ is 1.112 at age 3 and decreases to 1.023 at age 40. If the distribution of length-at-age is Normal then, since $\beta \cong 3$,

$$W(age) \cong W_\infty[1 - \exp(-k(age - t_o)\}]^\beta \; (1 + 3CV^2). \quad (3)$$

Using the approximation $\exp\left(\frac{CV^2\beta^2}{2}\right) \cong 1 + \frac{CV^2\beta^2}{2}$ the lognormal bias term is larger than the Normal distribution bias term, $1 + 3CV^2$ (see Fig. 9). However, unless the age-distribution of the stock changes substantially then the effect (on stock biomass and SSB estimates) of using mean weight-at-age compared to the median (i.e., Eqn. 1) will be approximately constant over time.

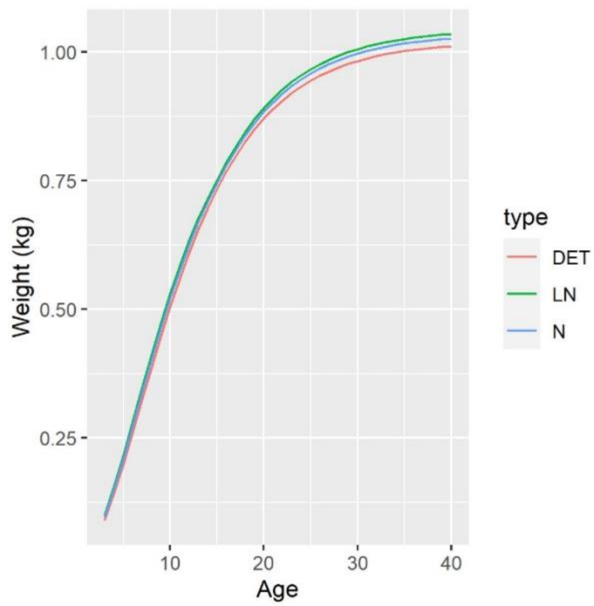

Figure 9. A comparison of weights-at-age for BSAI POP derived from a composite function of length-at-age and weight-at-length (deterministic, DET) and using statistical expectations based on Lognormal (LN) or Normal (N) distributions of length-at-age.

I was initially surprised at how similar mean size-at-age was for the fishery and AI survey. However, at younger ages (i.e., age <= 10) both the survey and fishery seem to have similar age-selectivity and presumably length-selectivity, and this may be why their sizes-at-age are similar.

Maturity at age

Two datasets were available to estimate maturity, and this was done within the assessment model so that uncertainty about maturity is included in SSB confidence intervals. However, the sample sizes were low compared to many other fisheries. I recognize that there are difficulties in obtaining samples at the time of POP spawning; however, more data will be useful.

Recommendation 5. Consider new sampling programs to collect information on POP maturity.

Fishery Sampling

The accuracy of commercial length- and age-compositions for population values is often uncertain in stock assessment. The POP assessment explored the spatial sampling of catches for length in good detail. I did not have concerns with this sampling. Similarly, the age sampling seemed good. My only concern was that I was uncertain how much POP catch came from hauls where POP would not be sampled. My impression was that this was not an important problem; however, it would be useful if this could be quantified somehow.

The reliability of fishery sampling can also be assessed by investigating how well cohorts are tracked through a time-series, especially for stocks with large variations in cohort sizes. SPAY plots are also useful for this; however, they are more difficult to interpret when there are changes in the fishery age-selectivity patterns, which is the case for POP. For example, cohorts at ages 3-7 in 1981 and 1982 were relatively strong (Fig. 10), but much less so 20 years later in 2001 and 2002. There was fairly good consistency between the strong cohorts in the AI survey age compositions and the catch age compositions (compare Figs 5 and 11), but there were some differences as well, especially at ages 3-6 in the last few assessment years.

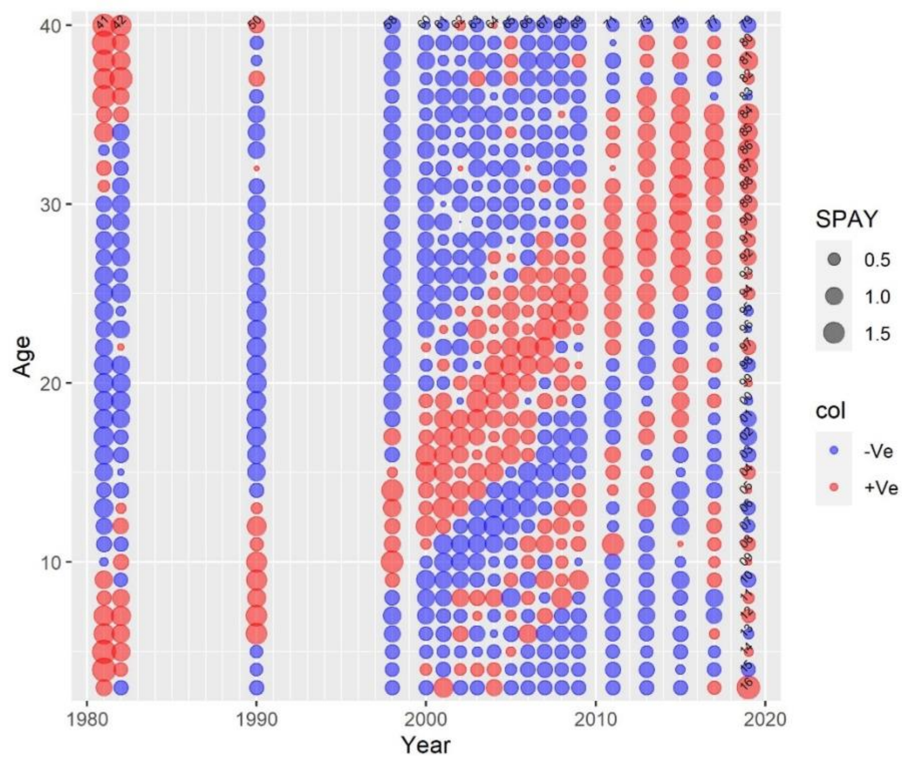

Figure 10. SPAY plot for the fishery age compositions during 1981-2019.

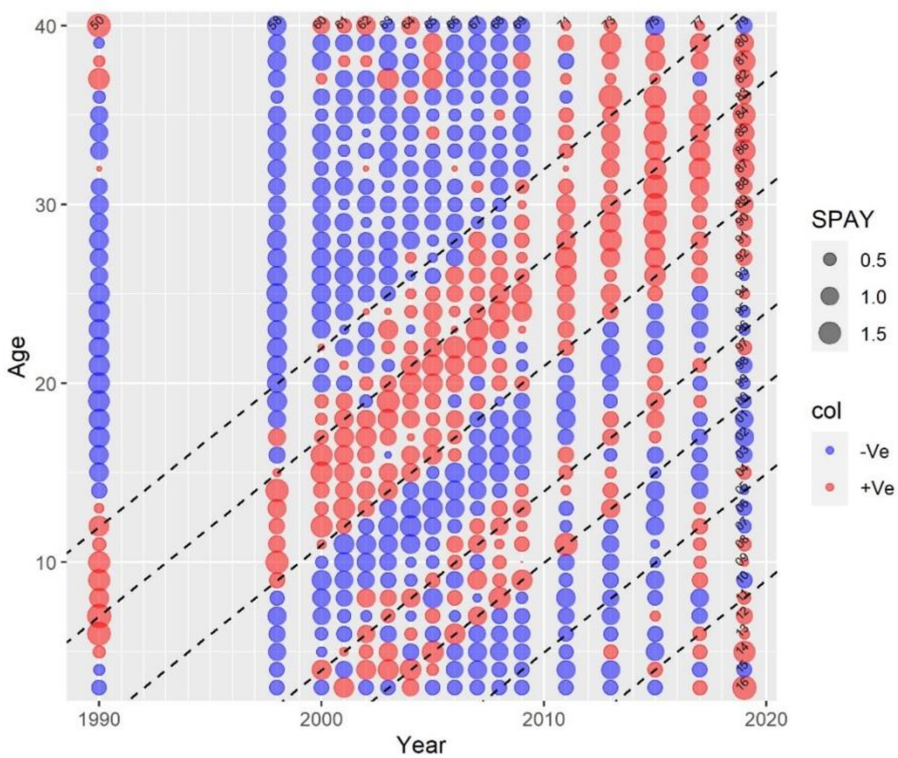

Figure 11. Recent SPAY plot for the fishery age compositions during 1990-2019.

Length composition sampling information from POP fisheries for years with no ageing is also used for fitting the assessment model. Standardized proportions-at-length (SPLY) statistics can also be computed to check how well cohorts track through length compositions. However, in addition to complications caused by changes in fishery size selectivity, tracking cohorts through length compositions also requires knowledge of growth rates. The SPLY statistics for the length compositions used for model fitting (Fig. 12) are difficult to interpret and suggest there may have been some abrupt changes in size selectivity between 1988 and 1989. Since 1990 the relatively strong 1984-1989 years classes that tracked through the age compositions (Fig. 11) also tracked through the length compositions (Fig. 13). However, the 1996 and 2000 cohorts which tracked through the age compositions fairly well are less evident in the length compositions. Part of the reason may be that there is too much pooling of lengths greater than 39 cm (see Research Recommendation 14 in ToR 5 section).

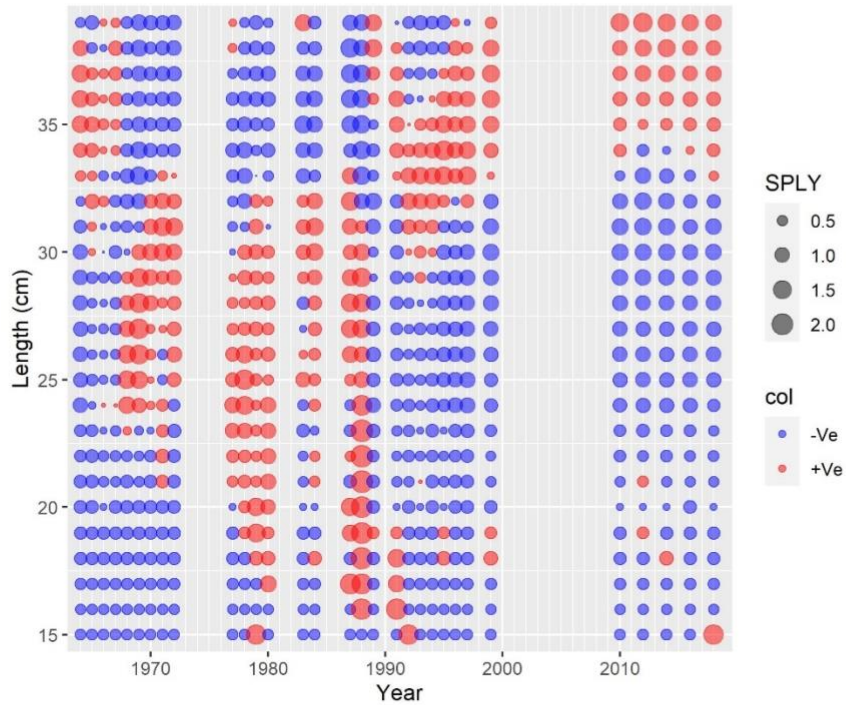

Figure 12. Standardized proportion-at-length (SPLY) plot for the fishery length compositions during 1964-2018.

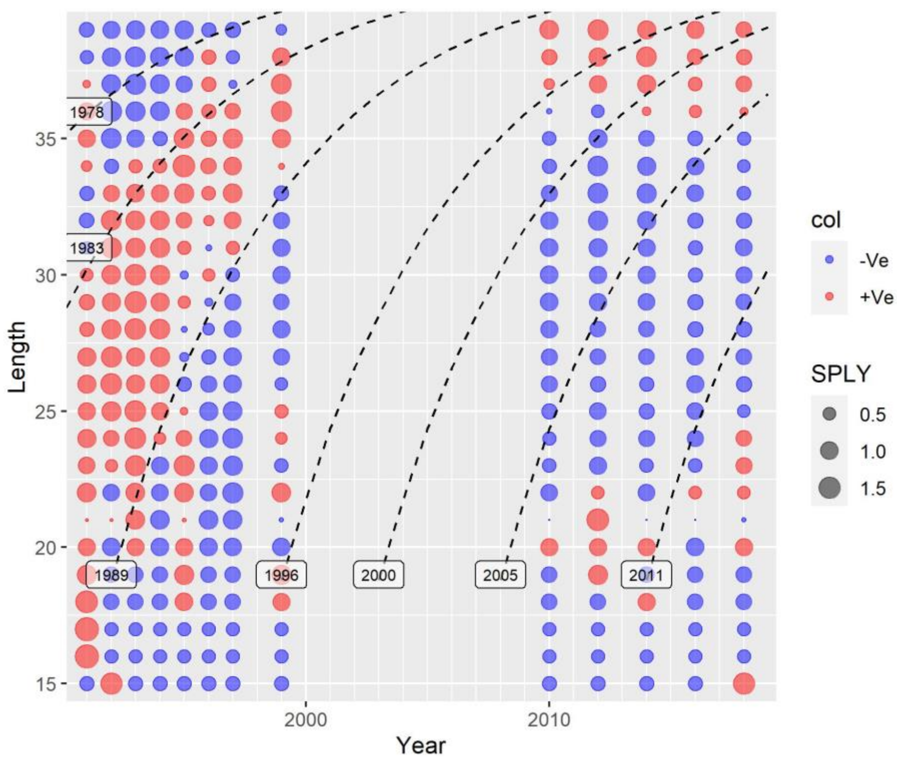

Figure 13. Recent SPLY plot for the fishery length compositions during 1990-2018. Dashed lines are Von Bertalanffy predicted lengths (based on assessment parameter estimates) at ages 3-40 for selected cohorts which are labelled.

Survey "availability"

Issue 4: The external estimation of the fraction (a) of the stock biomass available to the AI survey was not valid. This fraction should be a parameter and part of the assessment model.

The external estimation based on smoothed biomass estimates is biased. If we ignore the smoothing, the external estimate for year *t* is:

$$\hat{a}_t = \frac{\hat{B}_{AI,t}}{\hat{B}_{AI,t} + \hat{B}_{EBS,t}}. \quad (4)$$

The statistical expectation of $\hat{a}_t$ is:

$$E(\hat{a}_t) \approx \frac{a_t q_{AI} \sum_a S_{AI,a} B_{at}}{a_t q_{AI} \sum_a S_{AI,a} B_{at} + (1 - a_t) q_{EBS} \sum_a S_{EBS,a} B_{at}} \neq a_t. \quad (5)$$

Even if $S_{AI,a} = S_{EBS,a}$ for each age then:

$$E(\hat{a}_t) = \frac{a_t}{a_t + (1 - a_t) q_{EBS}/q_{AI}} \neq a_t. \quad (6)$$

Only if $S_{AI,a} = S_{EBS,a}$ and $q_{EBS} = q_{AI}$ does $E(\hat{a}_t) = a_t$; however, it is not reasonable to assume equal catchability for the two surveys because their footgears are different, among other reasons. A research recommendation for this issue is provided in Section 5 (#20). A simple solution is to include a_t as a model parameter. In a modern state-space stock assessment we would treat $logit(a_t)$ as a random walk or a Gaussian AR(1) stochastic process. In the context of the BSAI model, I suggest that $logit(a_t)$ could be modelled using a spline smoother.

ToR 2. Evaluation of analytical methods used in assessments, particularly in regard to selectivity, modeling of natural mortality, and data weighting assumptions.

BSAI Pacific Ocean perch model in general

The model is a deterministic (i.e., not state-space) age-structured cohort-dynamics model. The model is estimated using estimates of total fishery landings, a survey biomass index, age compositions from surveys, and age and length compositions from the fishery. Those length compositions are only used when age compositions are not available. Various priors are also used, and some of the likelihood components are weighted somewhat subjectively.

Literature evidence was presented that the stock structure assumption underlying the assessment model seem appropriate. I conclude that the model framework is appropriate but there are some ad hoc aspects. Specific details of the implementation and estimation of the POP assessment model could benefit from additional research, refinement, and testing, which I describe as follows:

Issue 7. Initial age distribution. This assumed an equilibrium age distribution for ages 4-40+ based on the assessment value for M. Typically, an initial value for Z is used or estimated but M is used for POP because it is thought that F was very low prior to the first model year. This may be true, but there was some F prior to 1961. It was not clear why age 3 in the first year was not modelled via the equilibrium age distribution. Also, the equilibrium age distribution could be improved as follows.

The equation used in the assessment model was

$$N_{a^+,1960} = R_{init} e^{-(a^+-3)M} / (1 - e^{-M}).$$

This assumes that POP live forever, which they don't. If A_{max} is the maximum age, then the equation should be

$$N_{a^+,1960} = R_{init} e^{-(a^+-3)Z_{init}} \left(\frac{1 - e^{-(A_{max}-a^+ +1)Z_{init}}}{1 - e^{-Z_{init}}} \right). \qquad (7)$$

If $Z_{init} = M$ then the choice of A_{max} will not affect the equilibrium age distribution much (Fig. 14) for a range of realistic values for A_{max}. If $Z_{init} > M$ then the choice matters even less. Of much greater importance is the value of Z_{init} used. A higher value may be appropriate if there was some fishing prior to the first year, or a density-dependent component to M, etc. I used the value $Z_{init} = 0.056 + 0.010$ in this figure only for illustration purposes.

There is some evidence of lack-of-fit to the initial fishery lengths compositions that could be consistent with an initial equilibrium age-structure based on Z a little greater than M. This may not have much impact on current stock status, but I conclude this should be further investigated (also see Research Recommendation 7 in Section 5).

Recommendation 6. Conduct a sensitivity analysis to realistic choices for Z_{init} in Equation (7). A likelihood profile for Z_{init} would also be informative.

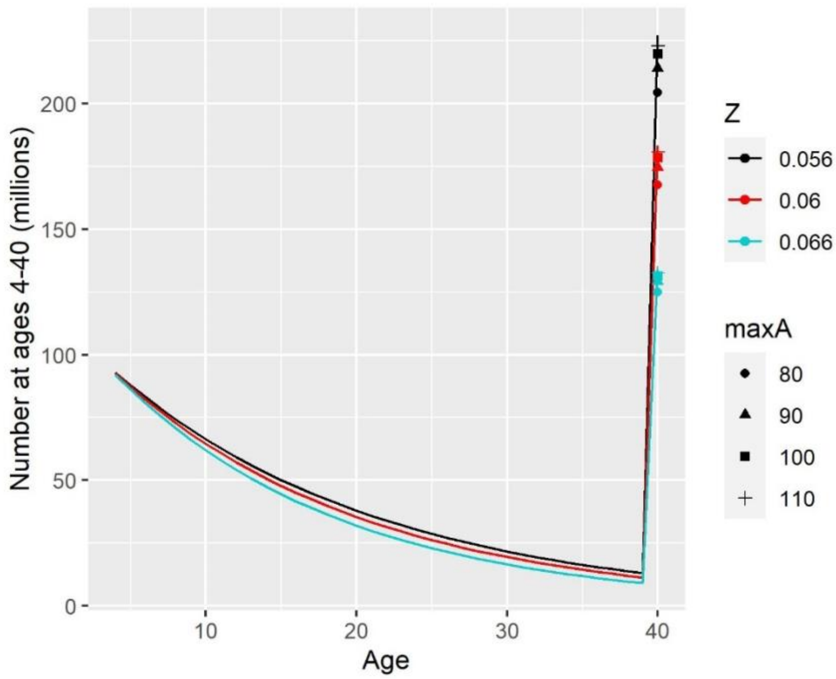

Figure 14. Equilibrium age distributions for three choices of Z, four choices for maximum age (maxA), and Rinit = 98.25 million.

A stochastic initial equilibrium age-distribution is more realistic. Afterall, if recruitment variability happened after 1960 why couldn't it happen before? Such a distribution could use the same types of recruitment deviations that the assessment model estimates; that is,

$$N_{a,1960} = R_{init} e^{-(a^+-3)Z_{init}} \exp{(\delta_a)},$$

where $\delta_a \sim N(0, \sigma_r^2)$. For the plus age group,

$$N_{a^+,1960} = R_{init} e^{-(a^+-3)Z_{init}} \left(\frac{1 - e^{-(A_{max}-a^+ + 1)Z_{init}}}{1 - e^{-Z_{init}}} \right) \exp(\delta_a), \tag{8}$$

but the value of σ_r should be smaller. This is because the plus group abundance depends on many recruitment deviations,

$$N_{a^+,1960} = R_{init} e^{-(a^+-3)Z_{init}} \sum_{k=0}^{A_{max}-a^+} e^{-kZ_{init}} \exp{(\delta_k)}. \tag{9}$$

If the $\delta_k{}'s$ are independent $N(0, \sigma_r^2)$, then:

$$
\begin{aligned}
Var\left(N_{a^+,1960}\right) &= R_{init}^2 e^{-2(a^+-3)Z_{init}} \sum_{k=0}^{A_{max}-a^+} e^{-2kZ_{init}} \exp(\sigma_r^2)\{\exp(\sigma_r^2)-1\} \\
&= R_{init}^2 e^{-2(a^+-3)Z_{init}} \exp(\sigma_r^2)\{\exp(\sigma_r^2)-1\}\left\{\frac{1-e^{-2(A_{max}-a^+ +1)Z_{in}}}{1-e^{-2Z_{init}}}\right.
\end{aligned} \tag{10}
$$

Therefore, the CV of $N_{a^+,1960}$ is:

$$
CV(N_{a^+,1960}) = \left\{\exp(\sigma_r^2)\{\exp(\sigma_r^2)-1\}\right\}^{1/2} \frac{\left\{\dfrac{1-e^{-2(A_{max}-a^+ +1)Z_{init}}}{1-e^{-2Z_{init}}}\right\}^{1/2}}{\left\{\dfrac{1-e^{-(A_{max}-a^+ +1)Z_{init}}}{1-e^{-Z_{init}}}\right\}}. \tag{11}
$$

If $A_{max} = 100$, $\sigma_r = 0.75$ and Z_{init}=*0.056* then the CV for $N_{a^+,1960}$ (i.e., standard deviation of δ_{a^+} in Eq. 8) should be 0.2. I verified the variance equation with monte carlo expectations based on 10 000 simulations of Eq. 9.

Recommendation 7. Consider using a stochastic initial equilibrium age distribution.

It is possible that σ_r could be lower prior to the 1960's if stock sizes were higher and recruitment was more stable (i.e., the portfolio effect). For example, if $\sigma_r = 0.25$ then the standard deviation of δ_{a^+} should be 0.045.

Selectivity

Survey selectivity was assumed to be a logistic function of age, and the age at 50% selection for the AI survey was $A_{AI,50}$ = 6.14 and by age 10 fish were almost fully selected. I usually look at residual patterns versus age, aggregated over years, to check if the selectivity model is appropriate. A plot of observed and model-predicted sampled numbers at age, aggregated over years, is useful for this purpose and commonly reported in many stock assessments. This information was not provided; however, while there were some patterns in the residuals (mostly negative for ages 15-20 and mostly positive for ages 33-37), the model fit to the age compositions seemed OK overall, but some cohorts have been under-estimated, and I am not sure if age measurement error is the reason for this (see ToR 4).

The EBS survey selectivity was also modelled as a logistic function. While the fit to this shorter index series was not as good as the AI survey fit, there was no evidence of selectivity model misspecification.

Recommendation 8. Provide observed and model-predicted age compositions, aggregated over years.

The fishery selectivity was time-varying via a bicubic spline model approximation. Fully selected F was also time-varying, with independent annual deviations. The equation used was

$F_{a,y} = s^f_{a,y} F_y = s^f_{a,y} e^{\mu_f + \varepsilon^f_y}$. There is confounding between the maximum value of $s^f_{a,y}$ and F_y and I was not sure how this was addressed. Otherwise, I liked the concept, but the practice could be improved. In state-space stock assessment models we use similar modelling approaches but $\log(F_y)$ will typically be assumed to follow a random walk or some type of correlated stochastic process. Fig. 15 suggests that a Gaussian AR(1) process assumption is appropriate, apart from the slightly significant negative partial autocorrelation at lag 4. The estimated F_y's for POP had strong autocorrelation and the model fitting should be improved by including this correlation in the likelihood for the F deviations. Using independent deviations will "penalize" F's towards an average value, but because the assessment model is designed to fit the total catches almost exactly, I don't expect that there will be a smoothing bias issue as long as the reported catches are accurate. However, this could affect likelihood profiles. For the bicubic spline, it is better to have more nodes at ages for which selectivity changes substantially. Consider moving the knot at age 30; for example, have knots at: 3, 9, 15, 21, and 40. These suggestions are included in a general recommendation to modernize the stock assessment (see #21 in Section 5).

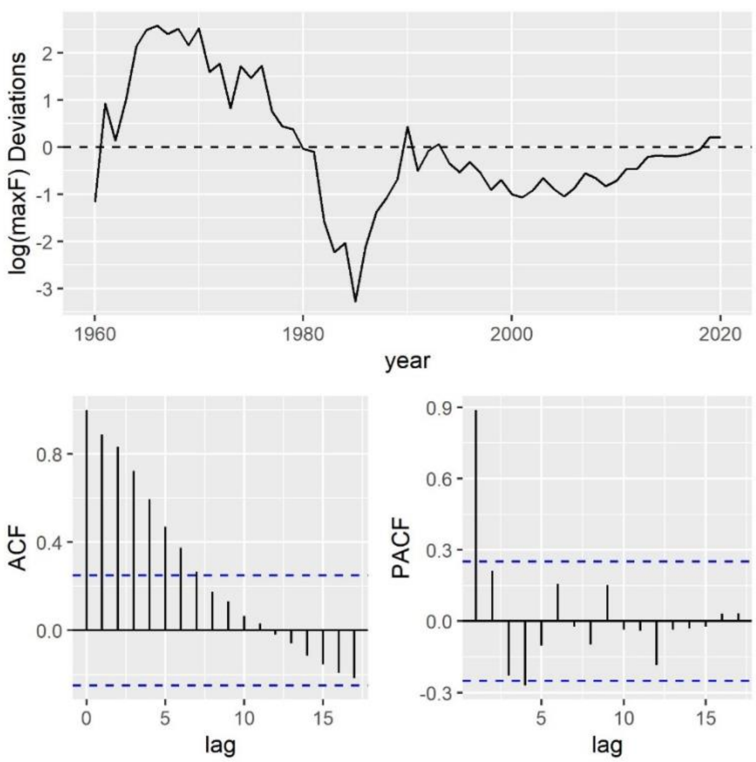

Figure 15. Top panel: Deviations in the log of the fully selected F estimated for BSAI POP. Bottom panels: Autocorrelation function (ACF) and partial autocorrelation function (PACF).

Issue 8: The assessment model estimated a very strong domed fishery selectivity pattern prior to the 1995 knot, and especially prior to the 1979 knot. This suggests that the assessment produces cryptic biomass, especially in the first part of the time-series. This is always a concern. However, the survey indices have asymptotic selectivity and therefore cryptic biomass is not a concern after 1990.

I am also concerned that the domed F pattern is linked to the assumption about the initial equilibrium age distribution (see **Issue 7**). Convincing rationale/evidence for the dome was not provided. I am uncertain what the evidence was for the relatively low F's the model estimates at older ages prior to 1979. I think the veracity of the estimates needs more investigation.

Recommendation 9. Fit a model with a more flexible initial age-distribution (e.g., stochastic) and an asymptotic fishery age selectivity pattern to demonstrate the lack-of-fit that a domed selectivity pattern fixes.

Natural Mortality

The assessment indicated that change in predation or BSAI POP condition were not issues, suggesting that change in M was not an important issue. However, the assessment text, "observed body condition has been lower than the survey mean since 2012", suggests the potential that starvation mortality (e.g., Björnsson et al., 2022; Regular et al., 2022) may have increased but I have no idea if the recent decrease in condition was to critical levels.

An informative prior on M was used, with a mean and coefficient of variation each set to 0.05. Three methods to approximate M based on Then et al. (2015), and three choices of maximum age (79, 104, and 129), were provided. The average of the 3x3=9 values based on Then et al. (2015) was 0.059, which was close to the estimated M from the 2020 assessment (0.056). A likelihood profile (Fig. 16) indicated that M was 0.06, although the two surveys and their age compositions favored much lower M's and the fishery age and length compositions favored higher M's. It is possible that the fishery length compositions were best fit with M≈0.07 because the model tends to over-estimate the proportion in the 39+ cm length bin and a higher M will improve this fit. It is also possible that this is connected to the assumption about the equilibrium age distribution.

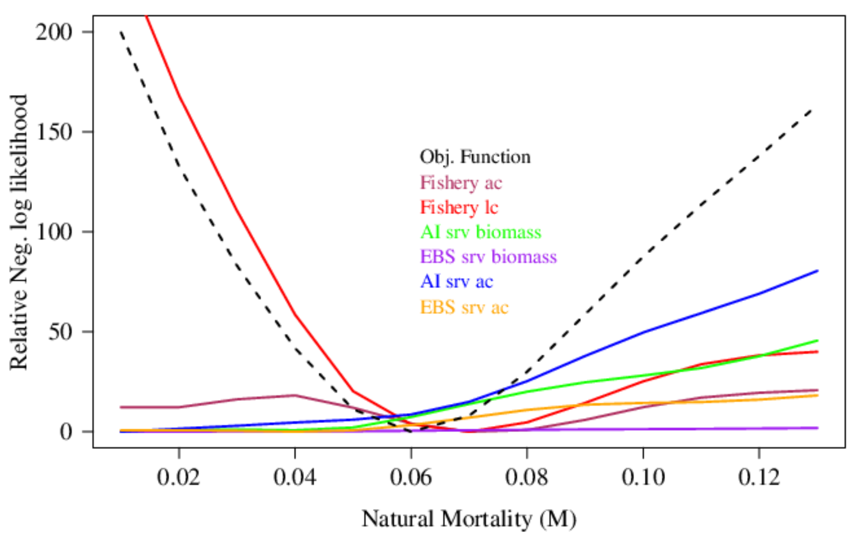

Figure 16. Likelihood profile for M.

To address change in M the assessors provided model runs with M estimated using four different options from the base case: 1) Remove prior; 2) Blocks every 4 years; 3) 2 blocks, and 4) 4-year blocks, no prior. The M estimates are shown in Fig. 17. We were not provided with model fit comparisons (e.g., AIC) so it was not clear if any of these options fit the assessment data better. Maybe model fit AIC comparisons are not possible because of the data-weighting procedures used. When M was estimated in 4-year blocks with no prior then the overall trend was a decrease until about 2010 and then an increase. The M in the first-time block was estimated to be substantially higher, and I suspect that this may be linked with possible mis-specification of the equilibrium initial age distribution assumption. This difference might not occur if a stochastic equilibrium initial age distribution was used. The increase since 2010 is consistent with the observed decrease in fish condition recently, although again I am not sure if any POP have been sampled in critical condition where short-term survival is doubtful.

The various M configurations resulted in some differences in biomass estimates but the fits to the AI survey indices were very similar, suggesting that q estimates also changed. For the 4-year blocks with no prior option, it seems q increased to close to 2 which may be unreasonable.

It is well known that the absolute value of M is confounded with the survey catchability age-pattern, and this is also something that should be considered when examining the M values. Time change in M is easier to estimate than absolute M values.

Recommendation 10. Investigate change in M using a model like $\log(M_y) = \log(M_o) + \delta_y$ where M_o is a fixed baseline assumption, and use a random walk for δ_y with a random walk standard deviation (σ_M) that is set small (i.e., 0.1). Possibly σ_M could be estimated if a profile likelihood and more detailed residual diagnostic analysis indicates the estimation is reasonable.

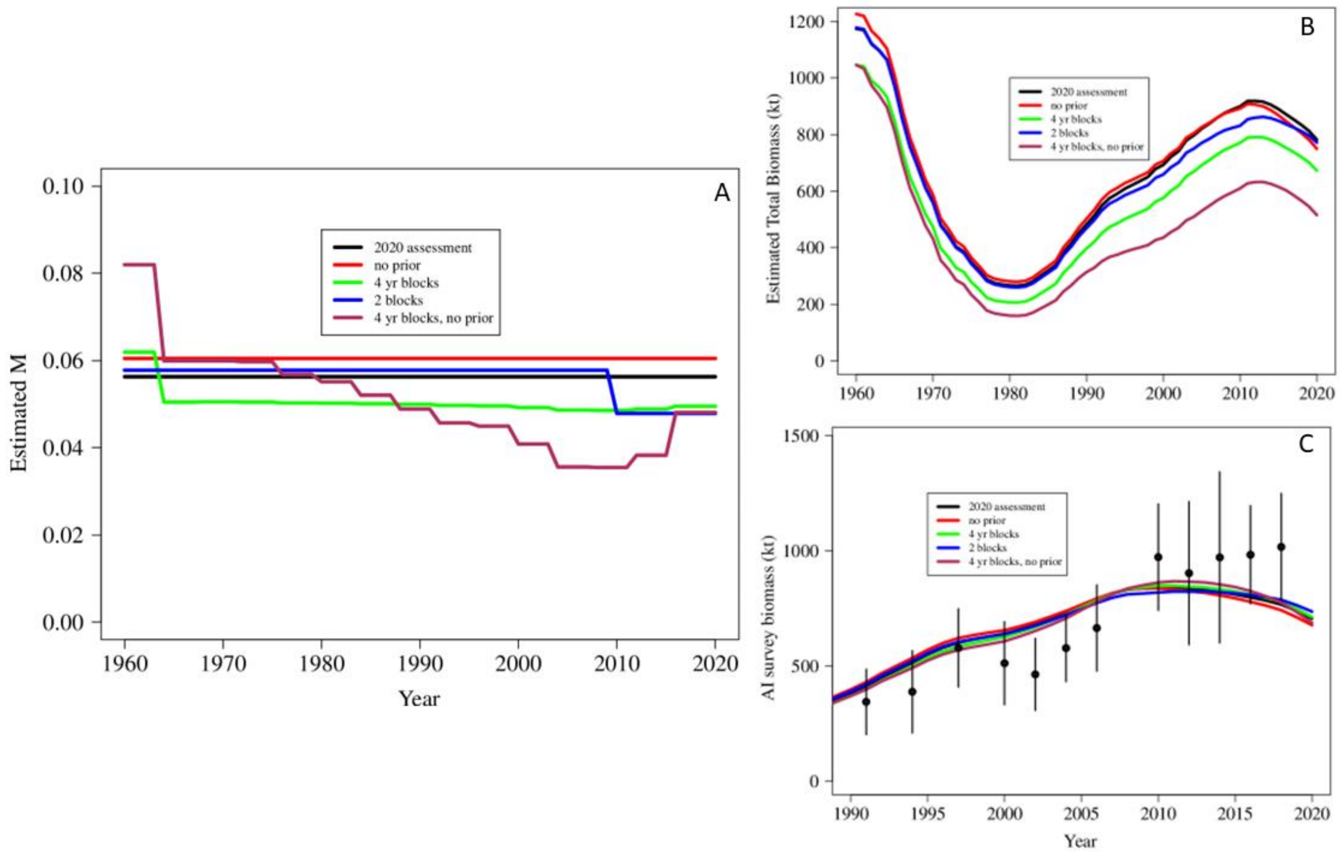

Figure 17. Panel A: Model results for different M-configurations. Panel A: M; Panel B: total biomass; Panel C: AI survey biomass.

Overall, I was satisfied with the M value estimated by the assessment model.

Recruitment

Recruitment was also estimated as a constant effect plus independent deviations constrained by recruitment variability (σ_r) that was fixed at 0.75. I was not sure of the rationale for fixing σ_r; I usually find that σ_r is a parameter that is identified by fisheries composition data. However, this parameter was estimated as part of the Panel meeting and the estimate was (0.72) was similar to the fixed value.

The assumption of independent deviations is common in stock assessment and seemed appropriate for BSAI POP (Fig. 17).

Recommendation 11. Estimate σ_r.

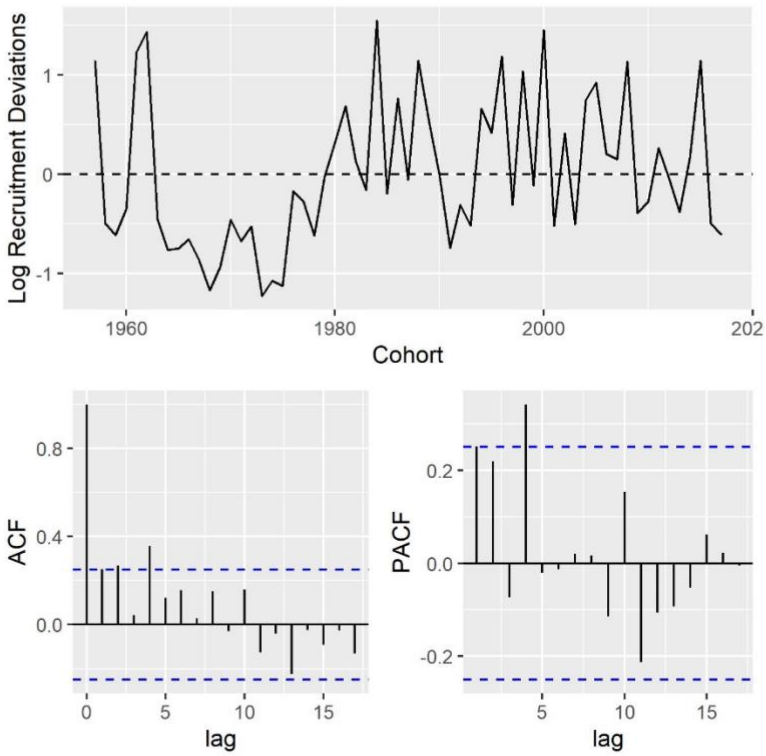

Figure 18. Top panel: Deviations in the log recruitment deviations estimated for BSAI POP. Bottom panels: Autocorrelation function (ACF) and partial autocorrelation function (PACF).

Data weighting

It is difficult to decide which of the data weighting procedures are preferable for composition data. Some of the procedures investigated are arcane. The Dirichlet-multinomial seems most objective to me and has some desirable properties in the stock assessment context, as described by Thorson et al., (2017). However, I usually find that Pearson and Deviance residuals indicate a different correlation structure than provided by the Dirichlet-multinomial or the Multinomial. Usually, residuals indicate positive correlations between age or length groups that are close together, and negative correlation for other age or length groups. There must be some negative correlation because the sum of composition counts is fixed. The Dirichlet-Multinomial or Multinomial gives negative correlation between all age or length groups. Hence, these choices of distributions seem incorrect. However, better modelling approaches for compositional data are ongoing but not yet published and available.

The assessment authors investigated different weighting options (Francis and Dirichlet-multinomial) and they did not seem to provide improvements compared to the base model approach. Francis (2011) recommended that fitting survey indices well was preferable to fitting compositional data well. However, for BASI POP the only way to fit the AI survey index well was to give low or no weight to all composition data.

Overall, I conclude that there were no good reasons to change the data weighting procedure from the base model settings.

ToR 3. Evaluation of the ability of the stock assessment model for BSAI Pacific ocean perch to provide parameter estimates to assess the current status of the stock.

The assessment model produces the stock quantities required to calculate $B_{40\%}$, $F_{35\%}$, and $F_{40\%}$, which are reference points used to evaluate stock status. I interpret 'ability' to be reliability. BASI POP stock status is based on $B/B_{40\%}$, $F/F_{35\%}$, and $F/F_{40\%}$. We did not review the details of the reference point calculations which is a shortcoming of the review. The reference points changed. For example, the assessment document indicted $B_{40\%}$ = 258,295 t last year but $B_{40\%}$ = 233,899 t this year. I am unsure what changed with the inputs to the calculations of $B_{40\%}$ to produce the change in the result.

There is a retrospective pattern in model estimates of SSB, but the pattern is increasing so that successive models produce higher estimates of SSB in recent years. These retrospective patterns will not affect conclusions regarding this ToR. The Panel should have been provided with retrospective status estimates; however, I do not think these would lead to different conclusions about current stock status. For example, $B_{2021}/B_{40\%}$ = 3.43 based on the 'last year' summary table results, and $B_{2021}/B_{40\%}$ = 3.23 based on 'this year' summary results. I am not sure if these results are retrospective evaluations of 2021 biomass status, but this is the type of result that would have been useful to evaluate this ToR.

Recommendation 12. Provide a table of inputs for $F_{X\%}$ calculations. Describe what recruitment value R is used to derive $B_{40\%} = R*SPR(F_{40\%})$.

Recommendation 13. Provide a retrospective analysis of current status evaluations. This will provide additional information on the reliability of the status evaluations.

ToR 4. Evaluation of the strengths and weaknesses in the stock assessment model for BSAI Pacific Ocean perch

Strengths

The age structured POP assessment model:

1. integrates much of the relevant productivity information (surveys and fishery catches, and their length- and age-compositions) to estimate stock size and fishing mortality rates.
2. The model can be estimated using both length- and age-compositions by utilizing a stochastic Von Bertalanffy growth model. This is very useful feature and avoids a major 'blight' for some purely age-based assessments that suffer from poor age sampling in some years.
3. Estimation utilizes some information about the precision of age-compositions.
4. Utilizes additional information about Q in the form of a data-based prior.
5. Accounts for uncertainty in the value of M.
6. Accounts for aging error.
7. Accounts for uncertainty in the maturation ogive in statistical inferences about SSB.

Weakness

1. There is a retrospective pattern in biomass, SSB, q, etc. Usually, retrospective patterns occur when there are time-patterns in residuals, which is the case for BASI POP.
2. The parametric logistic survey selectivity models and the spline model for fishery selectivity may be too restrictive. There is some evidence of lack of fit and age patterns in the POP model composition Pearson residuals.
3. There is a potential that the variation in the distribution of length-at-age is too large because it also includes measurement error in age.
4. The fishery age selectivity pattern is strongly domed prior to ~1990; hence, the model has a lot of historic cryptic biomass, and I was unsure of the evidence for this.
5. Cohort patterns in age compositions observed and predicted by the model are broadly similar overall, but there are some discrepancies which I did not understand (Fig. 19). I realize that some differences are to be expected because age measurement error will "blur" variation in cohort strength. However, the model seems to have under-estimated the strength of the 1997, 1999, and 2001 cohorts in both the survey (Fig. 20) and catch (Fig. 21) age compositions, more than the model over-estimated other adjacent cohorts.
6. The 2000 cohort declined rapidly in abundance after age 10, in both the AI survey and the fishery. The model could not fit this decline well. AI survey age compositions for other recent cohorts have also declined more rapidly than the model predicts after about age 10. This is consistent with a recent increase in M.
7. The survey length composition plus group should be extended to around 45+ cm. The present setting of 39cm+ may mask the distribution of older ages. This is probably a minor weakness.
8. There are some "missing fish" at lengths 15-17cm in the model because some of these fish will be age 2 which is not a model age. This may only be a minor issue because the fishery catches very few fish at these sizes. This is a minor weakness.

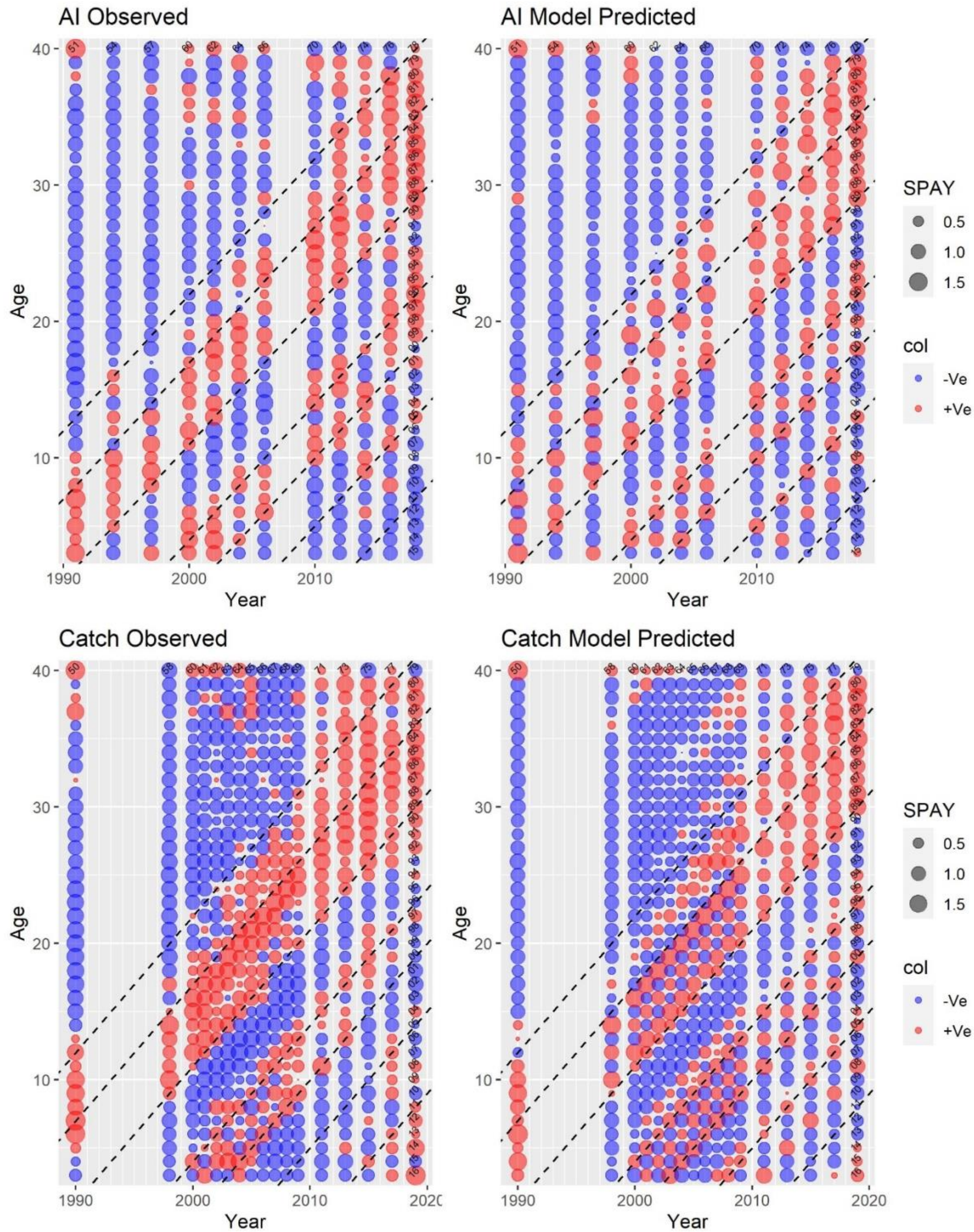

Figure 19. SPAY comparison of observed fishery survey age compositions and model predictions. The model predicted SPAY statistics are based only on the years with sampling for ages.

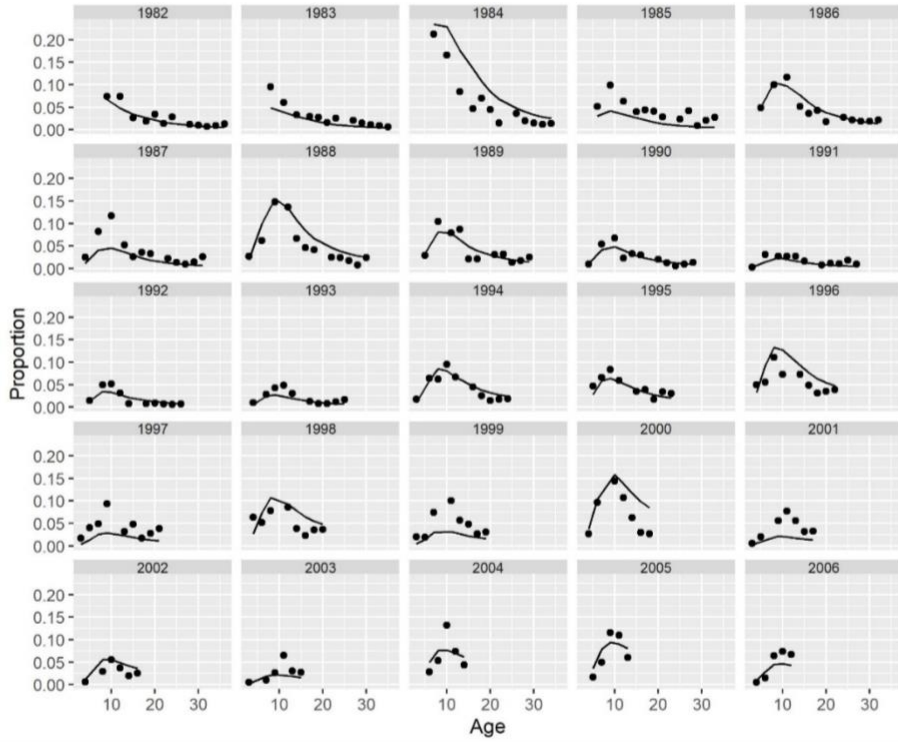

Figure 20. Observed (points) and model predicted (lines) AI survey age compositions. Each panel is for a cohort.

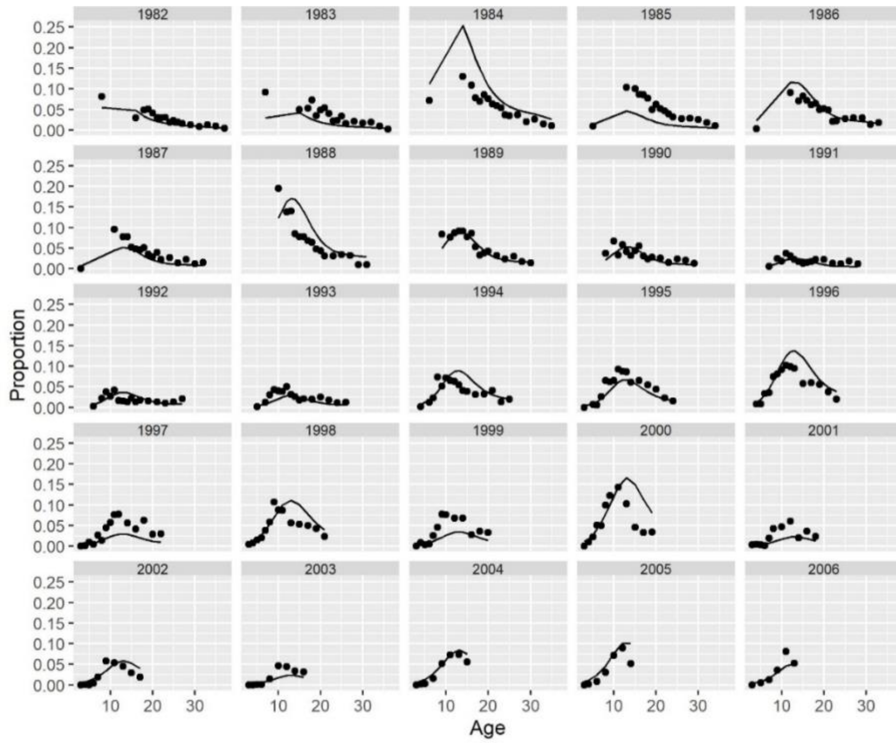

Figure 21. Observed (points) and model predicted (lines) fishery catch age compositions. Each panel is for a cohort.

ToR 5. Recommendations for improvements to the assessment models.

Research Recommendations were provided throughout this report and are summarized in this section. Some additional recommendations are also provided. I ranked recommendations as high (**), medium (*) and low (no asterix).

1. *The old bottom contact sensor and the newer accelerometer should have been used in all tows during the 2010 survey to get data on differences in estimates of time on bottom and tow distances from the two approaches. The survey biomass could have been calculated using the old survey protocols and the new ones, to see if this change in sensors made any difference. It will be useful to quantify the potential impact of this change in survey protocols.

2. Provide time-series of the length compositions of the AI survey.

3. *Conduct a spatiotemporal analysis of variation in size-at-age. Von Bertalanffy models with random year and area effects are an option, but more flexible models could be useful.

4. *A more thorough analysis of variation in size-at-age is needed, including separating the effects of individual variation in growth and age measurement error.

5. *Consider new sampling programs to collect information on POP maturity.

6. **Conduct a sensitivity analysis to realistic choices for Z_{init} in Equation (7). A likelihood profile for Z_{init} would also be informative.

7. **Consider using a stochastic initial equilibrium age distribution in the BSAI POP assessment model.

8. **Provide observed and model-predicted age compositions, aggregated over years.

9. **Fit a model with a more flexible initial age-distribution (e.g., stochastic) and an asymptotic fishery age selectivity pattern to demonstrate the lack-of-fit that a domed selectivity pattern fixes.

10. **Investigate change in M using a model like $\log(M_y) = \log(M_o) + \delta_y$ where M_o is a fixed baseline assumption, and use a random walk for δ_y with a random walk standard deviation (σ_M) that is set small (i.e., 0.1). Possibly σ_M could be estimated if a profile likelihood and more detailed residual diagnostic analysis indicates the estimation is reasonable.

11. *Estimate σ_r.

12. **Provide a table of inputs for $F_{X\%}$ calculations. Describe what recruitment value R is used to derive $B_{40\%}$ = R*SPR($F_{40\%}$).

13. **Provide a retrospective analysis of current status evaluations. This will provide additional information on the reliability of the status evaluations.

14. **Consider fitting the BASI POP model to survey abundance rather than survey biomass. The trends in survey abundance (Fig. 4) seem somewhat more consistent with model estimates, and fitting to survey abundance indices may reduce retrospective patterns. However, the choice of fitting to abundance or biomass seems subjective. If the model behaves somewhat differently when fit to abundance or biomass, then the assessment team will need to do research to better understand which of the survey abundance or biomass estimates provide more reliable indices of trends in stock size.

15. *The survey length composition plus group should be extended to around 45+ cm. The present setting of 39cm+ may mask the distribution of older ages.

16. *The model has an external age-length conversion matrix so the model could produce population numbers-at-length. Biomass could be derived using the sum (over length) of numbers-at-length times weight-at-length rather than inferring weight-at-age from length-at-age and weight-at-length. SSB could similarly be derived from maturity-at-length times biomass-at-length.

17. *Survey biomass could be modelled with length-based stock biomass and length-based selectivity. This may be an improvement if fish selectivity depends on length more than age, and if size-at-age changes over years.

18. Consider estimating the age-length conversion matrix internally. The external estimation has confounded variation in length-at-age with age measurement error.

19. There are some "missing fish" at lengths 15-17cm in the model because some of these fish will be age 2 which is not a model age. This may only be a minor issue because the fishery catches very few fish at these sizes.

20. *Estimate survey availability (a_t) as assessment model parameters. Time variation in $logit(a_t)$ could be modelled using a spline smoother.

21. **Use a Gaussian AR(1) distribution for annual fully selected F deviations. The bicubic spline model for $s_{a,y}^f$ may benefit from different placement of age and year knots.

Conclusions and Recommendations

The Bering Sea/Aleutian Islands Pacific Ocean perch (POP) stock assessment represents the best available science. The assessment model integrates relevant productivity information (surveys and fishery catches, and their length- and age-compositions) to estimate stock size and fishing mortality rates.

The model fit is good overall; however, there is a retrospective pattern in assessment model estimates of stock size that is related to recent survey biomass indices that were higher than the model could predict. There is some lack of fit to a few cohorts (1997, 1999, and 2001) that I did not understand. Despite this, the status evaluations are reliable.

Recommendations are provided under ToR 5.

Critique of the NMFS review process, including suggestions for improvements of both process and products

Overall, I find the NMFS review process to be rigorous and at a very high standard. The BSAI POP review process provided sufficient background material relevant to the assessment, but at the same time did not over-whelm reviewers with details.

I was unsure why the review was not required to produce a summary report. I am not suggesting that this is necessary, but I do feel it would have been useful for the panel to make a conclusion on whether the stock assessment was sufficiently complete, the science reviewed is the best scientific information available, and assessment results are robust and reliable as the basis for fisheries management decisions.

The virtual review format worked well for me. Most presentations were provided before the review meeting. However, in the virtual format there is no opportunity to learn more about the assessment in chats during breaks and meals, etc. This may be more of a disadvantage for new reviewers.

Appendix 1: Bibliography of materials provided for review

Assessment Document

Paul D. Spencer and James N. Ianelli. Dec 2020. Assessment of the Pacific ocean perch stock in the Bering Sea/Aleutian Islands

Meeting Presentations:

1. Pete Hulson. Overview of rockfish biology, fishery, and history of assessment.
2. Paul Spencer. Current management of Alaska rockfish.
3. Ned Laman, Paul Spencer. Survey data, Abundance, distribution, and age composition.
4. James Thorson. Update on model-based abundance.
5. Raul Rameriz, Paul Spencer. Fishery data – Catch, observer program, ages, lengths.
6. Delsa Anderl, Paul Spencer. Age determination, lengths, maturity, and growth.
7. Darin Jones. MACE Program. Gulf of Alaska Acoustic-Trawl Survey POP Abundance and Catchability. Presentation.
8. Paul Spencer. Assessment model: Model structure, likelihoods, data weighting, parameter estimates, data fit, diagnostics.
9. Paul Spencer. Model Developments: Alternative data weighting, Alternative specification for natural mortality, Other miscellaneous model developments.
10. Paul Spencer. Response to CIE requests. May 12, 2022
11. Paul Spencer. Response to CIE requests. May 13, 2022

Background Materials

Past assessments and review documents

1. Paul D. Spencer and James N. Ianelli. Dec 2016. Assessment of the Pacific ocean perch stock in the Bering Sea/Aleutian Islands.
2. Paul D. Spencer and James N. Ianelli. Dec 2018. Assessment of the Pacific ocean perch stock in the Bering Sea/Aleutian Islands.
3. Sven Kupschus. 2013. CIE Review of Alaska Rockfish Assessments.
4. Neil Klaer. 2013. CIE Reviewer's Independent Report on Alaska Rockfish Stock Assessment
5. Cathy Dichmont. 2013. AFSC Rockfish Assessment Review.
6. Summary and response to the 2013 CIE review of AFSC rockfish

Relevant research

1. Francis, R.C., 2011. Data weighting in statistical fisheries stock assessment models. Canadian Journal of Fisheries and Aquatic Sciences, 68(6), pp.1124-1138.
2. Jones, D.T., Rooper, C.N., Wilson, C.D., Spencer, P.D., Hanselman, D.H. and Wilborn, R.E., 2021. Estimates of availability and catchability for select rockfish species based on acoustic-optic surveys in the Gulf of Alaska. Fisheries Research, 236, p.105848.

3. Legault, C.M. and Palmer, M.C., 2016. In what direction should the fishing mortality target change when natural mortality increases within an assessment?. Canadian Journal of Fisheries and Aquatic Sciences, 73(3), pp.349-357.
4. Szuwalski, C.S., Ianelli, J.N. and Punt, A.E., 2018. Reducing retrospective patterns in stock assessment and impacts on management performance. ICES Journal of Marine Science, 75(2), pp.596-609.
5. Then, A.Y., Hoenig, J.M., Hall, N.G., Hewitt, D.A. and 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), pp.82-92.
6. Thorson, J.T., Johnson, K.F., Methot, R.D. and Taylor, I.G., 2017. Model-based estimates of effective sample size in stock assessment models using the Dirichlet-multinomial distribution. Fisheries Research, 192, pp.84-93.

Additional References

Björn Björnsson, Jón Sólmundsson, Pamela J Woods, Natural mortality in exploited fish stocks: annual variation estimated with data from trawl surveys, ICES Journal of Marine Science, 2022, fsac063, https://doi.org/10.1093/icesjms/fsac063

Cadigan, N. 2020. A Simple Random-Effects Model to Smooth and Extrapolate Weights-at-Age for 3Ps Cod. DFO Can. Sci. Advis. Sec. Res. Doc. 2020/xxx. v + 25p. In press.

Candy SG, Constable AJ, Lamb T,Williams R (2007) A von bertalanffy growth model for toothfish at heard island fitted to length-at-age data and compared to observed growth from mark-recapture studies. CCAMLR Sci 14:43–66

Cope, J. M. and Punt, A. E. (2007) Admitting ageing error when fitting growth curves: an example using the Von Bertalanffy growth function with random effects. Can. J. Fish. Aquat. Sci., 64, 205–218.

Frater, P.N. and Stefansson, G., 2020. Comparison and evaluation of approaches aimed at correcting or reducing selectivity bias in growth parameter estimates for fishes. Fisheries Research, 225, p.105464.

Kumar, R. 2020. “Making SPAY plot using R”. RPubs: https://rpubs.com/rajeevkumar/SPAY (Last updated: “2021-06-11”)

Perreault, A.M., Zheng, N. and Cadigan, N.G., 2020. Estimation of growth parameters based on length-stratified age samples. Canadian Journal of Fisheries and Aquatic Sciences, 77(3), pp.439-450.

Regular, P.M., Buren, A.D., Dwyer, K.S., Cadigan, N.G., Gregory, R.S., Koen-Alonso, M., Rideout, R.M., Robertson, G.J., Robertson, M.D., Stenson, G.B. and Wheeland, L.J., 2022. Indexing starvation mortality to assess its role in the population regulation of Northern cod. Fisheries Research, 247, p.106180.

Zheng, N., Cadigan, N. and Morgan, M.J., 2020. A spatiotemporal Richards–Schnute growth model and its estimation when data are collected through length-stratified sampling. Environmental and Ecological Statistics, 27(3), pp.415-446.

Appendix 2: CIE Statement of Work

Performance Work Statement (PWS)

National Oceanic and Atmospheric Administration (NOAA) National Marine Fisheries Service (NMFS)

Center for Independent Experts (CIE)

Program External Independent Peer Review

May 9-13, 2022

BSAI Rockfish – Pacific Ocean perch

Background

The National Marine Fisheries Service (NMFS) is mandated by the Magnuson-Stevens Fishery Conservation and Management Act, Endangered Species Act, and Marine Mammal Protection Act to conserve, protect, and manage our nation's marine living resources based upon the best scientific information available (BSIA). NMFS science products, including scientific advice, are often controversial and may require timely scientific peer reviews that are strictly independent of all outside influences. A formal external process for independent expert reviews of the agency's scientific products and programs ensures their credibility. Therefore, external scientific peer reviews have been and continue to be essential to strengthening scientific quality assurance for fishery conservation and management actions. Scientific peer review is defined as the organized review process where one or more qualified experts review scientific information to ensure quality and credibility. These expert(s) must conduct their peer review impartially, objectively, and without conflicts of interest. Each reviewer must also be independent from the development of the science, without influence from any position that the agency or constituent groups may have. Furthermore, the Office of Management and Budget (OMB), authorized by the Information Quality Act, requires all federal agencies to conduct peer reviews of highly influential and controversial science before dissemination, and that peer reviewers must be deemed qualified based on the OMB Peer Review Bulletin standards.
(http://www.cio.noaa.gov/services_programs/pdfs/OMB_Peer_Review_Bulletin_m05-03.pdf).

Further information on the CIE program may be obtained from www.ciereviews.org.

Scope

The stock assessment for Bering Sea/Aleutian Islands Pacific ocean perch provide the scientific basis for the management advice considered and implemented by the North Pacific Fisheries Management Council. An independent review of these integrated stock assessments is requested by the Alaska Fisheries Science Center's (AFSC) Resource Ecology and Fisheries Management (REFM) Division. The goal of this review will be to ensure that the stock assessments represent the best available science to date and that any deficiencies are identified and addressed.

The goal of this review will be to ensure that the stock assessments represent the best available science to date and that any deficiencies are identified and addressed. The specified format and contents of the individual peer review reports are found in **Annex 1**. The Terms of Reference (TORs) of the peer review are listed in **Annex 2**. Lastly, the tentative agenda of the panel review meeting is attached in **Annex 3**.

Requirements

NMFS requires three (3) reviewers to conduct an impartial and independent peer review in accordance with the PWS, OMB guidelines, and the TORs below. The reviewers shall have a working knowledge and recent experience in the application of age-structured stock assessment methods in general and, in particular, application of ADMB in stock assessment.

Additionally, the CIE reviewers shall have:

- Expertise with measures of model fit, identification, uncertainty, forecasting, and biological reference points;
- Familiarity with federal fisheries science requirements under the Magnuson-Stevens Fishery Conservation and Management Act;
- Familiarity with groundfish fisheries and management;
- Working knowledge of trawl survey design and estimation of stock biomass
- Excellent oral and written communication skills to facilitate the discussion and communication of results.

Tasks for Reviewers

1) Review the following background materials and reports prior to the review meeting. Two weeks before the peer review, the NMFS Project Contact will send by electronic mail or make available at an FTP site to the CIE reviewer all necessary background information and reports for the peer review. In the case where the documents need to be mailed, the NMFS Project Contact will consult with the CIE on where to send documents. The CIE reviewer shall read all documents in preparation for the peer review.

2) Attend and participate in the panel review meeting. The meeting will consist of presentations by NOAA scientists, including the stock assessment authors and survey team members to facilitate the review, provide any additional information and answer questions from the reviewers.

3) After the review meeting, reviewers shall conduct an independent peer review report in accordance with the requirements specified in this PWS, OMB guidelines, and TORs, in adherence with the required formatting and content guidelines; reviewers are not required to reach a consensus.

4) Each reviewer should assist the Chair of the meeting with contributions to the summary report, if required in the terms of reference.

5) Deliver their reports to the Government according to the specified milestones dates.

Foreign National Security Clearance

When reviewers participate during a panel review meeting at a government facility, the NMFS Project Contact is responsible for obtaining the Foreign National Security Clearance approval for reviewers who are non-US citizens. For this reason, the reviewers shall provide requested information (e.g., first and last name, contact information, gender, birth date, passport number, country of passport, travel dates, country of citizenship, country of current residence, and home country) to the NMFS Project Contact for the purpose of their security clearance, and this information shall be submitted at least 30 days before the peer review in accordance with the NOAA Deemed Export Technology Control Program NAO 207-12 regulations available at the Deemed Exports NAO website: http://deemedexports.noaa.gov/ and http://deemedexports.noaa.gov/compliance_access_control_procedures/noaa-foreign-national-registration- system.html. The contractor is required to use all appropriate methods to safeguard Personally Identifiable Information (PII).

Place of Performance
The place of performance shall be at the contractor's facilities, and in Seattle, WA.

Period of Performance
The period of performance shall be from the time of award through July 2022. The CIE reviewers' duties shall not exceed 14 days to complete all required tasks.

Schedule of Milestones and Deliverables
The contractor shall complete the tasks and deliverables in accordance with the following schedule.

Schedule	**Milestones and Deliverables**
Within two weeks of award	Contractor selects and confirms reviewers
Approximately 2 weeks later	Contractor provides the pre-review documents to the reviewers
May 9-13, 2022	Panel review meeting
Approximately 3 weeks later	Contractor receives draft reports
Within 2 weeks of receiving draft reports	Contractor submits final reports to the Government

Applicable Performance Standards
The acceptance of the contract deliverables shall be based on three performance standards:

(1) The reports shall be completed in accordance with the required formatting and content; (2) The reports shall address each TOR as specified; and (3) The reports shall be delivered as specified in the schedule of milestones and deliverables.

Travel

All travel expenses shall be reimbursable in accordance with Federal Travel Regulations (http://www.gsa.gov/portal/content/104790). International travel is authorized for this contract. Travel is not to exceed $8,000.

Restricted or Limited Use of Data
The contractors may be required to sign and adhere to a non-disclosure agreement.

Project Contact(s):
Paul Spencer
NMFS, Alaska Fisheries Science Center 7600 Sand Point Way NE, Seattle WA 98115 Phone: 206-526-4248
paul.spencer@noaa.gov

Appendix 2.3: DRAFT AGENDA

All times below are Pacific Daylight Time
Daily breaks at 11:30AM and 3:45PM, Lunch 1PM-2PM

Tuesday, March 30	
10:00 AM – 11:30 AM: Introduction/Background	
Introductions and agenda	Paul Spencer
Overview of rockfish biology, fishery, and history of assessment	Ben Williams
Current management of Alaska rockfish	Chris Lunsford
11:45 AM - 1:00 PM: Discussions	
2:00 PM – 3:45 PM: Input data	
Survey data	
Abundance, distribution, and age composition	Pete Hulson, Wayne Palsson
Model-based abundance	Pete Hulson, Jason Conner
Fishery data – Catch, observer program, ages, lengths	Pete Hulson, Tom Holland
Age determination, lengths, maturity, and growth	Pete Hulson, Chris Gburski, Delsa Anderl
4:00 PM - 5:00 PM: Discussions	
Wednesday, March 31	
Field-based catchability	
10:00 AM – 11:30 AM: Assessment model	Pete Hulson
Model structure, likelihoods, data weighting, parameter estimates, data fit, diagnostics	
11:45 AM - 1:00 PM: Discussions	
2:00 PM - 3:45 PM: Parameters, priors, diagnostics Pete Hulson	
Catchabilities, selectivities, model fits, diagnostics	
4:00 PM - 5:00: Discussion	
Thursday, April 1	
Pre-recorded presentations to review: Acoustic survey	
10:00 AM – 11:30 AM: Model developments	
Incorporation of acoustic information	Pete Hulson, Darin Jones
Incorporation of model-based index	Pete Hulson
Internal review model scenarios	Pete Hulson
11:45 AM - 1:00 PM: Discussion	
2:00 PM - 3:45 PM: Requested topics/model runs	
4:00 PM - 5:00 PM: Summarize, revisit Terms of Reference	
Friday, April 2	
10 AM - 2 PM: Additional time, as needed	

Appendix 3: Panel membership or other pertinent information from the panel review meeting

Chair	Pete Hulson	AFSC, Seattle
Members	Noel Cadigan	Memorial University, St. Johns, Newfoundland, Canada
	Geoff Tingley	Gingerfish Ltd, Wellington, New Zealand
	Matthew Cieri	Maine Department of Marine Resources, US

Attendees: N. Laman, J. Thorson, R. Rameriz, and D. Anderl