Center for Independent Experts Independent Peer Review Report

Gulf of Alaska Rockfish – Pacific ocean perch

Virtual Meeting. March 30 – April 1, 2021

Dr. Noel Cadigan

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

Executive Summary

- The Pacific ocean perch (POP) stock assessment represents the best available science and any deficiencies that were identified were also addressed during the review.
- The bottom trawl survey indices for POP stock size are reliable.
- There is a retrospective pattern in assessment model estimates of stock size that is related to higher-than-expected recent survey biomass indices.
    - This pattern may be related to conflict between survey indices and age compositions.
    - The model may be over-fitting the age compositions relative to the survey biomass indices.
- It was premature to use VAST biomass indices in the POP stock assessment.
- More years of acoustic survey data are needed before deciding how it could be included in the POP assessment.

Important research recommendation highlights are:

1. Investigate a bootstrap procedure to estimate uncertainty in survey age compositions, to better understand how much weight these data should get in the assessment model.
2. Provide VAST diagnostics to better understand the potential for trawlable abundance model bias in the survey sampling context.
3. A state-space stock assessment model should be developed.
4. Continue and improve research on the sources of uncertainty and possibly bias in abundance estimates from summer acoustic-trawl data.

Background

The stock assessment for Gulf of Alaska 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 this integrated stock assessment was requested by the Alaska Fisheries Science Center's (AFSC) Auke Bay Laboratories Division (ABL). The goal of the review was to ensure that the stock assessment represents the best available science and that any deficiencies are 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. Geoff Tingley, New Zealand; Dr. Saang-Yoon Hyun, South Korea), and the chair was Dr. Paul Spencer from the AFSC. Documents and presentations were provided and presented by B. Williams, C. Lunsford, P. Hulson, W. Palsson, J. Conner, T. Holland, C. Gburski, D. Anderl, and D. Jones. The support of all of 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, in particular, assessments developed with software such as ADMB. 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 the application of spatio-temporal models to population index estimation;
- Experience with application of acoustic data collection within stock assessment;
- 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 an independent peer-review report.

Role of reviewer

On March 20, 2021, the Panel chairperson provided assessment documents and supporting materials via a webpage. These documents are listed in **Appendix 1**. I reviewed the background documents I was provided and compiled a list of issues to get clarification at the Panel review meeting. I attended the entire Panel review 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.

I ranked recommendations as high (**), medium (*) and low (no asterisk).

ToR 1. Evaluate the data used in the assessments, specifically trawl survey estimates of biomass, and recommend how data should be treated within the assessment model.

<u>Evaluate data</u>

A substantial amount of time during the Review Panel was focused on more fully understanding the basic stock assessment model inputs. Many of my questions about model inputs were satisfactorily addressed during the Review. I do not describe all of these. However, the following questions/issues were discussed during the Panel meeting and are linked to research recommendations.

Issue 1: Tow times have varied between 15-45 minutes, although since 1996 the standard has been 15 minutes. Survey indices in 1990 and 1993 included some tows with durations > 15 minutes. How was this accounted for?

The response was that survey biomass indices are based on swept area, and no adjustment has been applied based on differences in tow times.

I conclude that no adjustment is more appropriate when the trawl catch-per-minute is approximately constant and not affected by possible pelagic catches during the trawl descent and ascent, initial vertical herding, etc. Recommendation 15 below is related to this issue. However, during 1990 and 1993 the tow times were more different than since then, and two of the vessels (83 and 92; see table below) had horsepower (HP) much less than vessels used since then. This is the motivation for Research Recommendation 1 below.

Issue 2: Are vessel/skipper effects accounted for in the survey analysis? Can a table of vessels/skippers for each year be provided?

The table was provided:

Table 1. Number of hauls by year and vessel made during the GOA bottom trawl survey along with vessel characteristics. Source: Wayne Palsson AFSC 3/30/2021.

Vessel Number	19	21	23	24	57	58	60	66	83	90	92	94	100	134	143	147	148	159	176	178	555	560	
Length		65.5	37.8	24	37	28	38	26	30.5	38	35.4	37.8	45	49.4	37.5	38	47.2	47.2	52	56.7		50	
Horsepower		2200	1700	500	1710		1750	565	565	1750	1285	1725	2000	1800	1710	1725	1800	1800	2160	1500		3400	Total Haul:
YEAR																							
1984		133			296	166															334		929
1987				166				218														399	783
1990	294	166							248														708
1993		153					315			246	60												774
1996			255									286	266										807
1999			278		206							280											764
2001					244							245											489
2003														275	276	258							809
2005														288	298	251							837
2007												246			287	283							816
2009												280			272			271					823
2011															336		334						670
2013															280				268				548
2015															303				290	178			771
2017															293		243						536
2019															269		272						541
All Years	294	452	533	166	746	166	315	218	248	246	60	1337	266	563	2614	792	849	271	558	178	334	399	11605

The response was that there has been no specific analysis of vessel effects; however, tows are monitored (e.g., scanmar/marport) based on the survey standards. Vessels are randomized within strata so vessel effects may cancel. The model uses indices since 1990.

All survey indices since 2013 have positive residuals from the assessment model; however, I do not think this can be caused by different survey vessels, because vessel 143 has been used since 2003 for about 50% of the survey tows. Recommendation 15 below is related to this issue.

Issue 3: Incomplete survey coverage seems to be a problem because the 500-1000m strata have not been consistently covered, although the 500-700m strata have been surveyed more consistently than the 700-1000m strata. How much POP are found at these depths?

The following table was provided in response. A negligible fraction of the survey biomass has been found in the 500-1000m strata. Wayne Palsson also provided a table of the number of hauls by stratum and year for the GOA Bottom Trawl Survey. Coverage seemed generally good except for strata in the Yakutat and Southeast regions in 2001. Hence, I conclude that incomplete survey coverage is not an assessment problem for POP. Incomplete survey coverage issues seem minimal especially compared to sampling variability. Recommendation 15 is related to this issue.

Table 2. Design-based biomass (mt) estimates by depth zone for POP in the Gulf of Alaska. Source: Wayne Palsson AFSC using the GOA_BIOMASS_DEPTH table 3/30/2021.

Depth Zone (m) Year	0-100	101-200	201-300	301-500	501-700	701-1000	Annual Total	Sum >=501 m	% >501 m
1996	5042.5	373486.5	337720.6	55163.0			771412.6	0.0	0.0
1999	1736.9	622059.5	85593.0	17396.5	216.0	61.6	727063.5	277.6	0.0
2001	18463.2	531232.9	121296.6	2162.3			673155.0	0.0	0.0
2003	6383.1	283872.5	158748.7	8406.7	10.5		457421.5	10.5	0.0
2005	14102.0	384276.8	347576.8	18763.0	137.5	44.7	764900.8	182.2	0.0
2007	4036.2	276953.0	365541.0	41641.4	8.2	0.0	688179.8	8.2	0.0
2009	4459.0	390420.4	231925.0	22418.5	97.5	128.4	649448.8	225.9	0.0
2011	13824.0	504727.1	244358.2	15624.6	135.8		778669.7	135.8	0.0
2013	19163.4	864574.8	395102.8	19429.4	172.8		1298443.2	172.8	0.0
2015	10485.2	870210.9	228059.8	31574.5	76.2	0.0	1140406.6	76.2	0.0
2017	525788.1	545614.2	458319.1	40625.8	11.6		1570358.8	11.6	0.0
2019	14467.6	953031.6	195611.9	47176.4	1857.1		1212144.6	1857.1	0.2
Total Over Time	637951.2	6600460.2	3169853.5	320382.1	2723.2	234.7	10731604.9	2957.9	0.0

Note: 500-1000m strata were also not surveyed in 1990 or 1993 (Wayne Palsson, 3/30/2021).

Issue 4. Stock and fishery weights-at-age (W@A). Are these different from each other, and have they changed over time? The different age-selectivity of the fishery and survey estimated with the assessment model suggests that there may be differences in the length-selectivity of these two fleets as well. If this is the case, then for POP the W@A of younger fish in the fishery may be greater than for the survey (i.e., the proxy for stock W@A). Conversely, there is evidence of a dome in age-selectivity for older ages in the fishery. If this selectivity is length-based, then W@A for older ages in the fishery may be less than stock W@A. Also, fishery age-selectivity has changed over time (blocks) which suggests the potential for a change in fishery W@A over time. Stock W@A can change over time as well, possibly in different ways than changes in fishery W@A.

Length-at-age (L@A) estimates from surveys since 1990 (Figure 1) were provided to address this. I agree that these estimates do not indicate that there have been large changes in stock size-at-age over time. However, the 2019 values at young ages (< 8) tend to be near the maximums of the time-series, and the values at older ages (> 20) tend to be near the lowest in the time-series. This is a surprisingly common pattern I see in many fish stocks, with stable or small increases in the size of young fish over time (i.e., years) but larger decreases in the length and weight of older fish. Hence, there is evidence of potential small changes in L@A and likely W@A. This may have some impact on estimates of stock biomass and research recommendation 2 is related to this. A potentially confounding issue is that the age samples were obtained using random sampling in 2019 whereas they were collected using length-stratified age sampling (LSAS) before 2019. Estimating Von Bertalanffy growth model parameters based on mean L@A that have been corrected for LSAS bias still produces some bias in parameter estimates (see Perreault et al., 2019), and better estimation procedures are available.

Issue 5. Maturity at age. Does this change over time and/or space?

The response was that the fishery and survey are not at the time of spawning, so it is difficult to get information on this, which is the motivation for research recommendation 3.

Issue 6. How are age and length compositions estimated.

We did not spend much time during the Review meeting on this issue. A brief response was that survey age compositions were developed using an age-length key and survey length compositions, which were weighted by stratum estimates of abundance. This seemed appropriate. However, because of potentially high within-stratum variation in catch rates, I provide a research recommendation to better understand the precision of the composition information which may also be informative about the weighting these data get in the assessment model estimation.

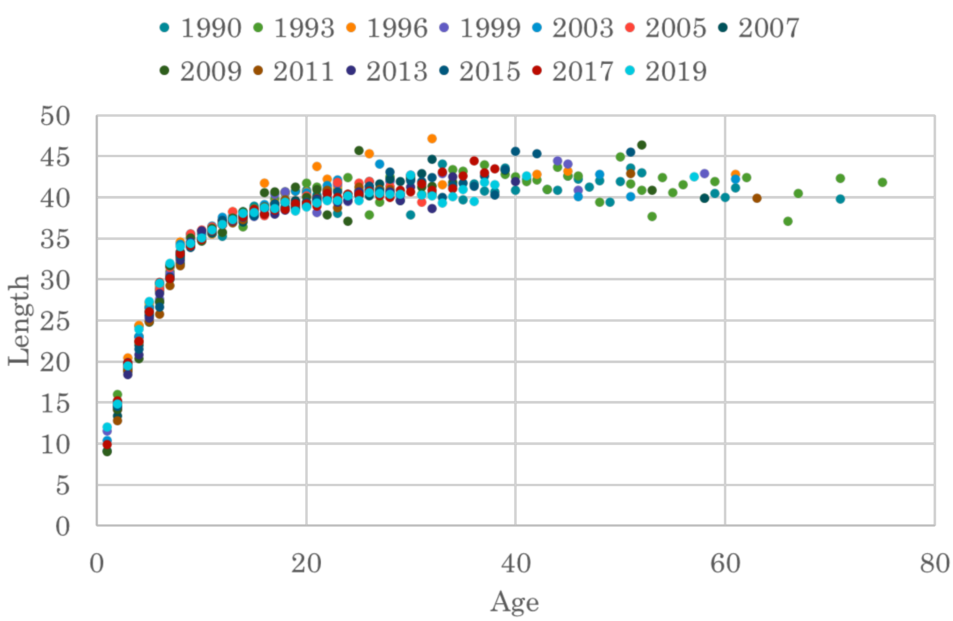

Figure 1. Survey mean length-at-age. Symbol colors indicate years.

<u>Recommend how data should be treated within the assessment model</u>

Recommendation 1. Conduct an assessment model sensitivity analysis to the 1990 and 1993 survey biomass indices and age compositions. This involves excluding the indices from the model fitting. Provide complete diagnostics for this run.

This sensitivity run was presented during the Review meeting. The run did not substantially alter the stock assessment results and management advice.

Recommendation 2. Investigate if stock weights-at-age from the survey are significantly (i.e., in the statistical sense) different than fishery weights-at-age. Also, investigate if there is significant temporal variation in both stock and fishery weights-at-age. Provide figures of how mean weight-at-age changes over time, with different panels for groups of ages (i.e., 1-5, 6-10, 10+). Consider using more efficient and less bias methods for analyzing size-at-age from length-stratified age samples (e.g., Perreault et al., 2019). Investigate spatiotemporal variation in weight as a function of length.

Recommendation 3*. Consider new sampling programs to collect information on POP maturity.

Recommendation 4.** Investigate a bootstrap re-sampling procedure (e.g., Jourdain et al., 2020) to estimate uncertainty (i.e., covariance) in survey age compositions. This could also be considered for fishery compositions, although I recognize that it may be less straight-forward if there is data-borrowing for unsampled fishery "strata" (i.e., gears, areas, seasons, etc.).

ToR 2. Evaluate the stock assessment model for GOA Pacific ocean perch in general and comment on appropriateness of parameter estimates to assess stock status determinations.

GOA Pacific ocean perch in general

The model is a standard 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 not state-of-the art, which now involves state-space models including process errors in population dynamics and time varying fishing mortalities. State-space models involve random effects plus an integrated likelihood which is more appropriate for variance parameter estimation and data weighting. 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 a stochastic equilibrium age distribution based on M and the estimated value of recruitment variation (σ_R). Typically, an initial value for Z is used or estimated but for POP, M is used 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. The F in 1961 was estimated to be about 50% of $F_{35\%}$. F in at least a few years prior to 1961 could be of a similar magnitude. 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 (see Research Recommendation 5).

Issue 8. Recruitment. There were strong temporal patterns in recruitment deviations, with several time-blocks in which all deviations have the same sign. However, in the last several years of the assessment this temporal pattern has largely disappeared. There is some evidence of a long-term increasing trend in recruitment during the same time-period that SSB has usually been increasing. This suggests that the model fit may be improved by including a stock-recruitment relationship plus auto-correlated errors in recruitment. In fact, it seems like there could be a small number of recruitment "regimes", or blocks of years with higher or lower POP recruitment, which is probably a better way to account for autocorrelation than a simple AR(1) process. I am aware that a practical approach to fit a hidden Markov stock-recruitment model will soon be published and the POP assessment may benefit from such a model in the future. However, this will not be easy to embed in the assessment model, but even external stock-recruit model fitting may be informative. Also, research to identify factors that may contribute to periods of good or poor recruitment regimes for POP will be useful and may provide additional perspective of the reliability of recent estimates of recruitment.

Issue 9. Data weighting. This is a difficult area in stock assessment. In general, I do not like subjective weighting for fitting like the POP model because it means statistical inferences based on standard errors and likelihood profiles are also subjective. However, the POP model gives the survey biomass indices a weight of one, and other data components are weighted relative to this, so the data weighting for the POP model fitting is not entirely subjective. In general, the weighting should be designed so that the total variation of Pearson residuals is about one for different data components. The Panel was not provided with survey biomass index Pearson residuals, but my sense was that these will look OK in terms of having residual variation close to one. The Panel was provided with Pearson residuals for survey and fishery age compositions, and these had variation far less than one which suggests to me than the model is over-fitting the age compositions relative to the survey biomass indices. However, maybe I mis-interpreted what the residuals were. This could be important because the POP model does not fit the recent survey biomass indices very well. The four observed indices during 2013-2019 are substantially greater than model predictions. No information was presented to the Panel to indicate that this could be caused by a change in survey catchability. A possible explanation is that there has been improved recruitment, including the 2006 year-class, that has been under-represented in the age compositions, especially in 2013, 2017 and 2019. However, a model that fits the survey indices well will likely not fit the survey and catch age compositions as well. I find it difficult to decide how much lack of fit to the compositions is reasonable because I do not understand the sampling precision of these data. This is linked with my **Research Recommendation 4** above.

The POP model gives relatively large weight to the catch weight time-series which I conclude is appropriate because the assessment team has good confidence in the reliability of this information. A small weight was used for F deviations. I am not sure why this was necessary given the high weight given to fishery catch weights.

Issue 10. Priors.

- I conclude the M prior was appropriate. No evidence was presented during the Review to indicate otherwise.
- The prior mean for survey catchability (Q) seemed appropriate and was derived from the estimate of Q in Jones et al. (2021). However, the Q prior CV seemed large relative the small CV in Jones et al. (2021). This allowed the model to estimate Q = 1.80. This is high compared to the estimate in Jones et al. (2021), but entirely plausible if survey trawl swept-area is larger than the values used to expand mean biomass-per-tow to the entire survey area. This was based on wing-spread and tow distance, but if there is horizontal herding beyond the wings then the actual swept area of a tow may be greater. Door-spread is sometimes considered to provide an upper bound on the horizontal part of area fished, and information provided during the review panel suggested door-spread has occasionally been monitored and was 3 times wing-spread. In this context, Q=1.8 is a plausible value.
- A prior was used for the standard deviation of recruitment deviations (σ_R). The prior mean for σ_R (1.7) is very large relative to the CV (20%) and the model estimate of σ_R (0.77). I assume the model would estimate a much lower σ_R if the σ_R prior was modified with a lower mean. This would have a substantial impact on longer-term projection uncertainty because recruitment variability is the only source of variability in the projections, and in the long-term. I am not sure what was the basis of the σ_R prior or why it was used.

Issue 11. Selectivity. A separable fishing mortality model is used, with different age-effects for blocks of years. The age-pattern in survey catchability was assumed to be logistic and asymptotic. Model results indicate substantial domed age patterns in F's for some blocks of years. Convincing rationale/evidence for the dome (i.e., why the fishery does not catch the older POP as well as the survey) was not provided. Most fishery age composition Pearson residuals at ages 13-15 were negative, and most were positive at age 10 and 22-25. This possibly indicates that the gamma selectivity function used in the POP assessment model is mis-specified or not flexible enough. Research Recommendation 8 is related to this.

Issue 12. Plus group. A substantial amount of survey and fishery catches have ages >= 25 which is the plus group age. However, given that the sizes of these fish are about the same and therefore fishery and survey selectivity may be the same, and considering the ageing error for older POP, I do not see a need to expand the plus group age. However, if an analysis of age-compositions (see Research Recommendation 10) indicates that year-classes are tracking through the compositions for ages >=25 then the plus group should be expanded.

Issue 13. A likelihood profile over Q and M was not as smooth as I expect. See Research Recommendation 11.

Recommendation 5. Investigate a sensitivity model run with an initial age-structure derived using the assumed M and a few years of F like that estimated for 1961. For example, initial cumulative Z = a*M + min(a,3)*Finit will be appropriate if the stock experienced Finit fishing mortality for three years prior to the start of the assessment model.

Recommendation 6. Consider including a stock-recruit model with autocorrelated errors to improve the fit of the POP assessment model. Investigate possible drivers of patterns in recruitment deviations.

Recommendation 7. Consider removing priors for F Regularity and σ_R.

Recommendation 8.** A research (i.e., exploratory) state-space stock assessment model, run in tandem with the current stock assessment model, should be developed. This could include a more flexible F model, population dynamics process errors, etc. The goal should be a model with less external data-weighting and priors, although some priors (i.e., M) may still be useful. This may produce improved quantification of the uncertainty of estimates and more realistic stochastic projections, especially considering that survival of POP may vary over time for these long-lived fish. However, process errors in a state-space model are better identified when there are multiple surveys. Process errors affect the population dynamics and stock size and can produce common residual patterns among multiple surveys. If there is only one survey index (like POP at present) then process errors may be less well identified. Nonetheless, there are single survey state-space assessment models in use, and the WHAM model (Stock and Miller, 2021) is a good option to consider as a start for POP. However, WHAM will need to be modified to include a stochastic growth curve to derive population numbers-at-length to fit to fishery length compositions.

Recommendation 9. Consider including fishery length composition information in off-years when ages are not measured. However, this may not provide much additional information about recent recruitment trends because of the low selectivity of the fishery for ages less than seven.

Recommendation 10*. Evaluate the quality of fishery and survey age compositions for tracking cohorts. I find SPAY plots helpful for this (see https://rpubs.com/rajeevkumar/SPAY).

Recommendation 11*. Provide convergence diagnostics, including the maximum absolute gradient and the results of a jitter test.

Appropriateness of parameter estimates to assess stock status determinations

Under Issue 10 above I indicated that I was uncertain about the priors for Q and σ_R which will affect the scale of the stock size estimates and levels of F (Q prior) and the amount of process error that is propagated into projections (σ_R). However, changes in the CV of the Q prior may have little impact on estimation of $B/B_{X\%}$ or $F/F_{X\%}$ (X=35,40). Changes in the σ_R prior may affect probabilities about stock status determinations.

There is also a retrospective pattern in the model, such that biomass estimates have been larger as years of data are added to the assessment. This is caused by the higher than predicted survey biomass indices since 2013. The retrospective patterns are not severe in that confidence intervals for SSB usually contain the values estimated as assessment data are added in subsequent years. However, SSB estimates do increase systematically in the last several years.

Recommendation 12*. Provide a retrospective analysis of current status evaluations. This will provide additional information on the reliability of the status evaluations.

ToR 3. Evaluate the strengths and weaknesses in the stock assessment model for GOA Pacific ocean perch, and recommend any improvements to the assessment model.

Strengths

The age structured POP assessment model:

1. Utilizes a stochastic Von Bertalanffy growth model to also fit catch at length data, thus increasing the amount of information the model "integrates" for the stock assessment.
2. Estimation utilizes some information about the precision of age-compositions. Although this is common in the US, in many assessments in Canada and Europe this information is not readily available for stock assessment and is not used for model estimation.
3. Utilizes additional information about Q in the form of a data-based prior.
4. Accounts for uncertainty in the value of M.
5. Accounts for aging error.
6. Accounts for uncertainty in the maturation ogive in statistical inferences about SSB.

Weakness

1. Not state-of-the-art, which I think is state-space with stochastic population dynamics and mortality processes, and appropriate observation models (e.g., Schnute, 1994; Aanes et al., 2007; Nielsen and Berg, 2014; Cadigan, 2015; Aeberhard et al., 2018; Perreault et al., 2020; Stock and Miller, 2021).
2. The multinomial likelihood for age and length composition data does not adequately account for the over-dispersion and correlation structure that commonly occurs.

However, better alternatives are still an active area of research and may require a state-space model framework including random effects.
3. The model does not account for between-age-year variation in M.
4. Parametric selectivity models may be too simplistic, and there is some evidence of lack of fit in the POP model composition Pearson residuals.

Improvements

See Research Recommendations throughout this report.

The Review Panel was presented with some information about the "Science-Industry Rockfish Research Collaboration in Alaska (SIRRCA)". A possibly useful research activity is to collect fine-scale spatiotemporal data on fishery catch rates and then utilize a depletion model based on local declines in CPUE to estimate local stock size. This could be compared with survey and acoustic biomass indices that would need to be conducted in the same area prior to the fishery, or after the fishery, to get additional information on the catchability of these surveys. If a substantial amount of this information can be provided, then this could be a valuable input in a future spatial stock assessment. However, there may be many challenges to address (e.g., Cadigan, Wade, and Nielsen, 2017) about how CPUE relates to stock size, that may still lead to an improved understanding of fishery impacts on POP.

ToR 4. Evaluate and recommend how survey data are used for biomass indices within the assessment. Specifically, advise on trawl survey indices arising from design-based methods versus model-based approaches.

Design-based methods versus model-based approaches

There is an extensive amount of statistical literature on design- versus model-based approaches in survey sampling research, which unfortunately I have not been keeping up with completely. There are also hybrid approaches, broadly referred to as model-assisted approaches (e.g., Särndal et al., 2003; Chen et al., 2004) that offer a good compromise between model-efficiency and the model-robustness of the design-based approach. I use notation similar to Chen et al. (2004) to describe these approaches.

Assume the survey area is divided into *N* distinct tow sites and that the catch variable of interest at site *i* is y_i and that the population average is $\bar{y}_N = N^{-1}\sum_{i=1}^{N} y_i$. Note that *N* is usually very large and it is impossible to sample every site. Assume that *n* << *N* sites are sampled in a survey using a probability sampling design where the probability of sampling at site *i* is $\pi_i = \Pr(i \in s) > 0$ and the sample *s* are the *n* sites chosen to trawl at. A generic designed-based estimate of $\bar{y}_N$ is

$$\bar{y}_s = N^{-1}\sum_{i=1}^{n} y_i/\pi_i.$$

In statistics this is called the Horvitz-Thompson estimator. For example, if a simple random sampling design is used then $\pi_i = \frac{n}{N}$ and $\bar{y}_s = \bar{y}$ is the ordinary sample mean. If the population is divided into *H* strata and stratified simple random sampling is used with n_h samples in stratum

$h = 1, \dots, H$ then $\pi_{i \in h} = n_h/N_h$, $\bar{y}_s = N^{-1} \sum_{i=1}^{n} N_h \bar{y}_h$, and $\bar{y}_h$ is the ordinary sample mean for stratum h.

In fisheries surveys there will typically be measurement error in the catches, and the best we can hope for is that the trawl catch is unbiased for the local trawlable abundance at site *i* which I denote as μ_i; that is, $E_M(y_i) = \mu_i$. I assume the stochastic processes that generate the catch have some probability distribution function that is used for the model-based expectation E_M. I use the subscript D to denote design-based expectations where the average is with respect to all possible samples *s*. Typically the model will use auxiliary covariates (e.g., latitude, longitude, depth) that are known for all tow sites *1,…,N* and parameters that must be estimated using the sampled catches $\{y_i\}_{i \in s}$ to estimate the μ_i at all tow sites. Let $\hat{\mu}_i$ denote the estimate. The purely model-based estimate of $\mu_N = N^{-1} \sum_{i=1}^{N} \mu_i$ is

$$\hat{\mu}_N = N^{-1} \sum_{i=1}^{N} \hat{\mu}_i.$$

This is the VAST approach, where a model is used to predict μ_i at all tow sites. However, for various reasons the model may not always provide unbiased predictions, and this can create design-bias in $\hat{\mu}_N$ as an estimate of μ_N. Sometimes the bias can be severe (Chen et al., 2004).

An operational disadvantage of purely model-based approaches is that model assumptions must be appropriate and model estimation must be sufficiently reliable. This requires examining model diagnostics for each survey variable of interest, and typically there are many variables of interest for a species (i.e., number per tow, weight per tow, number per tow and length class, etc.) and many species of interest in fisheries surveys. It will usually be impractical to examine model goodness-of-fit for many variables. A single best model may not be apparent either, or model selection statistics such as AIC may guide us to the incorrect model (Thorson et al., 2021). Opsomer et al. (2007) referred to *generic* inference as the problem of making sensible estimates for many variables in a straightforward and internally consistent way, and they referred to *specific* inference in which custom models are built for a few variables and the dataset at hand. I suggested that the human resource limitations typical of almost all fisheries science organizations that support stock assessment means that our focus should be on generic inference. There is therefore an understandable reluctance to specify statistical models for the behavior of all the variables of interest in the population (Breidt and Opsomer, 2017), whereas the design-based approach provides a simple and robust all-purpose statistical framework. A disadvantage of the design-based approach is that the resulting estimators can be inefficient, and sometimes dramatically so.

There are many model-assisted approaches that have been designed to provide design-unbiased estimates of μ_N (see Skinner and Wakefield, 2017; Breidt and Opsomer, 2017) but also improved efficiency compared to purely design-based estimators. Design consistency of model-assisted estimators is guaranteed, under very weak assumptions, and in particular consistency does not depend on strong modeling assumptions (Skinner and Wakefield, 2017). An intuitive approach is the difference estimator of μ_N,

$$\hat{\mu}_{Diff} = N^{-1} \sum_{i=1}^{N} \hat{\mu}_i + \sum_{i \in s} \frac{y_i - \hat{\mu}_i}{\pi_i} = \hat{\mu}_N - bias(\hat{\mu}_N),$$

where $-\sum_{i\in s}\frac{y_i-\hat{\mu}_i}{\pi_i}$ is a design-based estimate of the bias in $\hat{\mu}_N$ as an estimate of $\bar{y}_N$, and is an approximately design- and model-based estimate of the bias in $\hat{\mu}_N$ as an estimate of μ_N. This estimator will be approximately design-unbiased and is design-consistent regardless of any potential misspecification of the model. However, if the model-based $\hat{\mu}_i$ are highly correlated with the y_i then the design-variance of $\hat{\mu}_{Diff}$ will be much smaller than the design-based estimator, $\bar{y}_s$.

A problem for stock assessment is that $\hat{\mu}_{Diff}$ can be negative and for that and other reasons alternative model-assisted approaches have been proposed. For example, Liang et al. (2017) proposed Bayesian model calibration with the pseudo-empirical likelihood framework to produce improved Blue Crab (*Callinectes sapidus*) abundance indices. Model calibration approaches are an active area of research in survey sampling (e.g., Wu and Thompson, 2020).

As a first step, I think it will be useful to investigate if there is evidence of substantial VAST model bias. This only requires evaluating stratum size-weighted averages of the VAST ordinary raw residuals (observed minus model predicted). This is not an analysis of the VAST model goodness of fit, but rather just an evaluation if the VAST model predictions given unbiased predictions of trawl catches at sampled sites.

Recommendation 13.** Provide the stratum size-weighted averages of the VAST ordinary raw residuals.

This is the $bias(\hat{\mu}_N)$ term in the above equation. If the absolute average bias is large, then additional and detailed examination of the VAST assumptions and estimation will be necessary. If VAST provides biased predictions of the trawl catches at the sample sites on average, then this casts doubt on the reliability of the VAST predictions for unsampled sites.

The standard designed-based estimator for a stratified random survey will be the same as a stratum-effects model-based approach, in which each stratum*year combination is a separate parameter in a statistical model. If the mle of the mean is the sample mean (i.e., Normal, Gamma, Poisson, Negative Binomial, delta-Gamma distributions) then the strata size-weighted average of model predictions will be the same as the design-based estimator. Hence, in the sense of Firth and Bennett (1998), the stratum-effects model is design-consistent which is a desirable property.

Note that estimation of the measurement error variance is a problem with the stratum effects model when there are many strata and low sample sizes within strata. MLE's of variance parameters have a known bias that is not ignorable when the number of parameters is large relative to the sample size. This is a problem for statistical inferences about stock size (Cadigan, 2011). Also, the stratum-effects model cannot be used directly to interpolate trawlable densities in incomplete surveys in which not all strata are sampled in some years, although this is not an important problem for POP. Hence, I am not advocating for the stratum-effects model, but I just use this as an example of a desirable model property.

POP biomass indices

The Review Panel was provided with density maps of VAST model predictions of survey biomass per tow (I think) that indicated high biomass existed at the offshore edge of the prediction region. This was not consistent with observations that POP are mostly found at depths less than 500m. This needs to be clarified (see Research Recommendation 14).

During the Review meeting it was suggested that VAST may provide better POP abundance estimation in untrawlable areas compared to simply assuming that POP density is the same in trawlable and untrawlable areas which is an implicit assumption in the design-based approach. It is not clear to me that VAST has much additional information about differences in POP densities in trawlable and untrawlable areas so I am uncertain that this could be the source of the recent differences in VAST versus design-based POP biomass indices.

Another possible reason why VAST indices may differ from design-based ones involves the different effects of large catches on these indices. However, in recent years there have been fewer large catches in the POP bottom trawl surveys, so this does not seem to help explain the differences in the VAST and design-based indices.

Another discussion topic involved how to get length and age compositions based on VAST outputs. These compositions are survey strata abundance-weighted and how to do this with VAST needs more research. One approach is to do a length-structured VAST and then aggregate estimates across the survey domain to get survey abundance-at-length. A spatially aggregated age-length key could be applied to derive abundance-at-age. Alternatively, and perhaps more appropriately, a spatial age length-key (e.g., Babyn et al., 2021) could be applied to abundance-at-length to estimate spatial abundance-at-age which could then be aggregated over the spatial domain. This is what is done in some ICES assessments (e.g., North Sea cod). However, a length-structured VAST will be computationally demanding and perhaps prohibitive. Nonetheless, spatial variation in POP density may depend on spatial variation in length-structure and this should be investigated and accounted for in a VAST model. It is somewhat accounted for in the design-based method which is usually applied independently to survey catch-at-length.

I conclude that it was premature to use VAST biomass indices in the POP stock assessment. There are several diagnostic analyses that need to be explored. I understand what the designed-based estimator is doing, but I am less certain about the efficacy of the VAST POP indices.

Recommendation 14.** Provide trawlable biomass values aggregated over survey strata. This should include time-series of maps indicating strata, where each stratum is colored to indicate the area-expanded VAST biomass. Also useful are time-series plots of VAST biomass aggregated over sets of strata for standard depth ranges shown in Table 2. It will also be informative if this could be further divided into trawlable and untrawlable grounds.

Recommendation 15*. Account for potential vessel and tow time effects in a VAST model. Examine the statistical significance of vessel and tow duration effects. Consider including vessel as a random effect.

Recommendation 16. In conjunction with **Recommendation 15**, consider including the 1984 and 1987 survey catches in the VAST model, to extend the survey biomass indices back to those years. This VAST model should include those effects that were different or less standardized in the 1984 and 1987 surveys. Consider the potential confounding of year effects with other effects.

Recommendation 17*. Investigate methods to produce length and size compositions that are weighted by VAST spatial density estimates.

ToR 5. Evaluate abundance estimates from summer acoustic-trawl data, and recommend how it may be used within the assessment.

Availability of POP to this off-bottom survey may be low compared to the bottom trawl survey, although the survey covers the spatial extent of POP. The lower availability is obvious when comparing acoustic and bottom-trawl biomass indices. The acoustic biomass indices are much lower. POP and pollack are difficult to differentiate by acoustic backscatter alone so 110-115 pelagic trawls (usually) are conducted to provide information on species/length/age compositions and the length-weight relationship. It seems that acoustic survey personnel can sometimes differentiate when POP may be present in acoustic backscatter, but I was unclear about how this works and how the information may be used when estimating POP abundance. During the meeting it was described that uncertainty evaluation only involves the spatial distribution of total biomass from acoustic backscatter. We discussed sources of uncertainty and possibly bias related to:

1. length-weight relation
2. length-frequency spatial distribution
3. trawl selectivity
4. fish avoidance
5. species classification errors
6. TS-length relation

However, all these issues require **further research**. The acoustic biomass index series is still short for stock assessment – four estimates in 2013, 2015, 2017, and 2019. This only covers 6 years of POP dynamics.

I conclude that more years of acoustic survey data are needed before deciding how it could be included in the POP assessment. However, having an additional fishery-independent abundance index, and in particular an acoustic survey of the off-bottom (i.e., 0.5m) water column, can be quite valuable for detecting changes in availability of POP to the bottom-trawl survey.

Recommendation 17**. Continue and improve research on the sources of uncertainty and possibly bias listed above. This should include quantification and incorporation of these sources of uncertainty into acoustic biomass and age/size compositions.

Additional References

Aeberhard, W.H., Mills Flemming, J. and Nielsen, A., 2018. Review of state-space models for fisheries science. Annual Review of Statistics and Its Application, 5, pp.215-235.

Aanes, S., Engen, S., Sæther, B.E. and Aanes, R., 2007. Estimation of the parameters of fish stock dynamics from catch-at-age data and indices of abundance: can natural and fishing mortality be separated?. Canadian Journal of fisheries and aquatic sciences, 64(8), pp.1130-1142.

Babyn, J., Varkey, D., Regular, P., Ings, D. and Flemming, J.M., 2021. A gaussian field approach to generating spatial age length keys. Fisheries Research, 240, p.105956.

Breidt, F.J. and Opsomer, J.D., 2017. Model-assisted survey estimation with modern prediction techniques. Statistical Science, 32(2), pp.190-205.

Cadigan, N.G., 2011. Confidence intervals for trawlable abundance from stratified-random bottom trawl surveys. Canadian Journal of Fisheries and Aquatic Sciences, 68(5), pp.781-794.

Cadigan, N.G., 2015. A state-space stock assessment model for northern cod, including under-reported catches and variable natural mortality rates. Canadian Journal of Fisheries and Aquatic Sciences, 73(2), pp.296-308.

Cadigan, N.G., Wade, E. and Nielsen, A., 2017. A spatiotemporal model for snow crab (*Chionoecetes opilio*) stock size in the southern Gulf of St. Lawrence. Canadian Journal of Fisheries and Aquatic Sciences, 74(11), pp.1808-1820.

Chen, J., Thompson, M.E. and Wu, C., 2004. Estimation of fish abundance indices based on scientific research trawl surveys. Biometrics, 60(1), pp.116-123.

Firth, D. and Bennett, K.E., 1998. Robust models in probability sampling. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 60(1), pp.3-21.

Jourdain, N.O.A.S., Breivik, O., Fuglebakk, E., Aanes, S. and Vølstad, J.H., 2020. Evaluation of sampling strategies for age determination of cod (*Gadus morhua*) sampled at the North Sea International Bottom Trawl Survey. ICES Journal of Marine Science, 77(3), pp.859-869.

Liang, D., Nesslage, G., Wilberg, M. and Miller, T., 2017. Bayesian calibration of blue crab (*Callinectes sapidus*) abundance indices based on probability surveys. Journal of Agricultural, Biological and Environmental Statistics, 22(4), pp.481-497.

Nielsen, A. and Berg, C.W., 2014. Estimation of time-varying selectivity in stock assessments using state-space models. Fisheries Research, 158, pp.96-101.

Opsomer, J.D., Breidt, F.J., Moisen, G.G. and Kauermann, G., 2007. Model-assisted estimation of forest resources with generalized additive models. Journal of the American Statistical Association, 102(478), pp.400-409.

Perreault, A.M., Zheng, N. and Cadigan, N.G., 2019. Estimation of growth parameters based on length-stratified age samples. Canadian Journal of Fisheries and Aquatic Sciences, 77(3), pp.439-450. https://doi.org/10.1139/cjfas-2019-0129.

Perreault, A.M., Wheeland, L.J., Morgan, M.J. and Cadigan, N.G., 2020. A state-space stock assessment model for American plaice on the Grand Bank of Newfoundland. J. Northw. Atl. Fish. Sci, 51, pp.45-104.

Särndal, C.E., Swensson, B. and Wretman, J., 2003. Model assisted survey sampling. Springer Science & Business Media.

Schnute, Jon T. A general framework for developing sequential fisheries models. Canadian Journal of Fisheries and Aquatic Sciences 51, no. 8 (1994): 1676-1688.

Skinner, C. and Wakefield, J., 2017. Introduction to the design and analysis of complex survey data. Statistical Science, 32(2), pp.165-175.

Stock, B.C. and Miller, T.J., 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, p.105967.

Thorson, J.T., Cunningham, C.J., Jorgensen, E., Havron, A., Hulson, P.J.F., Monnahan, C.C. and von Szalay, P., 2021. The surprising sensitivity of index scale to delta-model assumptions: Recommendations for model-based index standardization. Fisheries Research, 233, p.105745.

Wu C., Thompson M.E., 2020. Calibration Weighting and Estimation. In: Sampling Theory and Practice. ICSA Book Series in Statistics. Springer, Cham. https://doi.org/10.1007/978-3-030-44246-0_6.

Conclusions and Recommendations

I have ranked recommendations as high (**), medium (*) and low (no asterisk).

ToR 1. Evaluate the data used in the assessments, specifically trawl survey estimates of biomass, and recommend how data should be treated within the assessment model.

- Many vessels have been used in the bottom trawl surveys for POP. However, vessels are randomized within strata so vessel effects may cancel. All survey indices since 2013 have positive assessment model residuals; however, I do not think this can be caused by different surveys vessels, because vessel 143 has been used since 2003 for usually about 50% of the survey tows.
- Incomplete survey coverage is not an assessment problem for POP. Incomplete survey coverage issues seem minimal especially compared to sampling variability.
- Length-at-age estimates do not indicate that there have been large changes in stock size-at-age over time, but there is evidence of potentially small changes in length-at-age and likely weight-at-age. This may have some impact on estimates of stock biomass.

Recommendation 1. Conduct an assessment model sensitivity analysis to the 1990 and 1993 survey biomass indices and age compositions. This sensitivity run was presented during the Review meeting. This run did not substantially alter the stock assessment results and management advice.

Recommendation 2. Investigate if stock weights-at-age from the survey are significantly (i.e., in the statistical sense) different than fishery weights-at-age. Also, investigate if there is significant temporal variation in both stock and fishery weights-at-age.

Recommendation 3*. Consider new sampling programs to collect information on POP maturity.

Recommendation 4.** Investigate a bootstrap re-sampling procedure (e.g., Jourdain et al., 2020) to estimate uncertainty (i.e., covariance) in survey age compositions. This should also be considered for fishery compositions.

ToR 2. Evaluate the stock assessment model for GOA Pacific ocean perch in general and comment on appropriateness of parameter estimates to assess stock status determinations.

- The model framework is appropriate but not state-of-the art, which now involves state-space models including process errors in population dynamics and time varying fishing mortalities.
- Pearson residuals for survey and fishery age compositions had variation far less than one which suggests the model is over-fitting the age compositions relative to the survey biomass indices.
- The M prior was appropriate.
- The prior mean for survey catchability (Q) seemed appropriate and was derived from the estimate of Q in Jones et al. (2021). However, the Q prior CV seemed large relative the small CV in Jones et al. (2021).
- The basis of the σ_R prior or why it was used was not clear. The choice of prior will have a substantial impact on longer-term biomass projection uncertainty.
- The gamma selectivity function used in the POP assessment model may be mis-specified or not flexible enough.
- The plus group age specification seems good. However, if year-classes are tracking through the compositions for ages >=25 then the plus group should be expanded.
- Typically, an initial value for Z is used or estimated but for POP, M is used 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.
- Changes in the CV of the Q prior may have little impact on estimation of $B/B_{X\%}$ or $F/F_{X\%}$ (X=35,40).
- Changes in the σ_R prior may affect probabilities about stock status determinations.

Recommendation 5. Investigate a sensitivity model run with an initial age-structure derived using the assumed M and a few years of F like that estimated for 1961.

Recommendation 6. Consider including a stock-recruit model with autocorrelated errors to improve the fit of the POP assessment model. Research to identify factors that may contribute to periods of good or poor recruitment regimes for POP may provide additional perspective of the reliability of recent estimates of recruitment.

Recommendation 7. Consider removing priors for F Regularity and σ_R.

Recommendation 8.** A research (i.e., exploratory) state-space stock assessment model, run in tandem with the current stock assessment model, should be developed.

Recommendation 9. Consider including fishery length composition information in off-years when ages are not measured. However, this may not provide much additional information about recent recruitment trends because of the low selectivity of the fishery for ages less than seven.

Recommendation 10*. Evaluate the quality of fishery and survey age compositions for tracking cohorts. I find SPAY plots helpful for this (see https://rpubs.com/rajeevkumar/SPAY).

Recommendation 11*. Provide convergence diagnostics, including the maximum absolute gradient and the results of a jitter test.

Recommendation 12*. Provide a retrospective analysis of current status evaluations. This will provide additional information on the reliability of the status evaluations.

ToR 3. Evaluate the strengths and weaknesses in the stock assessment model for GOA Pacific ocean perch, and recommend any improvements to the assessment model.

<u>*Strengths*</u>

1. Integrates catch at length data in an age-structured model.
2. Estimation utilizes some information about the precision of age-compositions.
3. Utilizes additional information about Q in the form of a data-based prior.
4. Accounts for uncertainty in the value of M.
5. Accounts for aging error.
6. Accounts for uncertainty in the maturation ogive in statistical inferences about SSB.

<u>*Weakness*</u>

5. Not state-space.
6. The multinomial likelihood for age and length composition data does not adequately account for over-dispersion and correlation.
7. The model does not account for between-age-year variation in M.
8. Parametric selectivity models may be too simplistic.

<u>*Improvements*</u>

See Research Recommendations throughout this report.

Research progress will result in an improved POP stock assessment model in the future. However, I expect the application of a better model with the current data available will not produce substantially different inferences about stock status.

The stock assessment was sufficiently complete, and the science reviewed is the best scientific information available.

ToR 4. Evaluate and recommend how survey data are used for biomass indices within the assessment. Specifically, advise on trawl survey indices arising from design-based methods versus model-based approaches.

- An operational disadvantage of purely model-based approaches is that model assumptions must be appropriate and model estimation must be sufficiently reliable.

- A disadvantage of the design-based approach is that the resulting estimators can be inefficient, and sometimes dramatically so.
- Model-assisted approaches offer a good compromise between model-efficiency and the model-robustness of the design-based approach, and these approaches should be considered for POP stock size indices in the future.
- It was premature to use VAST biomass indices in the POP stock assessment. There are several diagnostic analyses that need to be explored.

Recommendation 13.** Provide the stratum size-weighted averages of the VAST ordinary raw residuals, which is a design-based estimate of potential VAST model bias.

Recommendation 14.** Provide trawlable biomass values aggregated over survey strata. Also useful are time-series plots of VAST biomass aggregated over sets of strata for standard depth ranges.

Recommendation 15*. Account for potential vessel and tow time effects in a VAST model. Examine the statistical significance of vessel and tow duration effects.

Recommendation 16. In conjunction with **Recommendation 15**, consider including the 1984 and 1987 survey catches in the VAST model, to extend the survey biomass indices back to those years.

Recommendation 17*. Investigate methods to produce length and size compositions that are weighted by VAST spatial density estimates.

ToR 5. Evaluate abundance estimates from summer acoustic-trawl data, and recommend how it may be used within the assessment.

- More years of acoustic survey data are needed before deciding how it could be included in the POP assessment.
- However, having an additional fishery-independent abundance index, and in particular an acoustic survey of the off-bottom (i.e., 0.5m) water column, can be quite valuable for detecting changes in availability of POP to the bottom-trawl survey.

Recommendation 17**. Continue and improve research on the sources of uncertainty and possibly bias in abundance estimates from summer acoustic-trawl data. This should include quantification and incorporation of these sources of uncertainty into acoustic biomass and age/size compositions.

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 POP review process provided sufficient background material relevant to the assessment, but at the same time did not over-whelm reviewers.

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, and I especially appreciated the voice-over narrations that were provided. 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

Meeting Materials:

1. Ben Williams. Overview of Pacific ocean perch biology, fishery, and history of assessment. Presentation.
2. Wayne Palsson. Gulf of Alaska Bottom Trawl Survey. Presentation.
3. Jason Conner. VAST estimates of Pacific Ocean Perch. Presentation.
4. Tom Holland. North Pacific Observer Program Alaska Fisheries Science Center Fisheries Monitoring and Analysis Division. Presentation.
5. Chris Gburski and Delsa Anderl. Age Determination of Gulf of Alaska Pacific Ocean Perch (Sebastes alutus), Northern Rockfish (Sebastes polyspinus), and Dusky Rockfish (Sebastes ciliatus) at the Alaska Fisheries Science Center (AFSC). Presentation.
6. Pete Hulson. GOA Pacific ocean perch. Input Data. Presentation.
7. Darin Jones. MACE Program. Gulf of Alaska Acoustic-Trawl Survey POP Abundance and Catchability. Presentation.
8. Pete Hulson. GOA Pacific ocean perch Model development. Presentation.
9. Pete Hulson. GOA Pacific ocean perch Model structure and results. Presentation.

Background Materials

Past assessments and review documents

1. Peter-John F. Hulson, Chris R. Lunsford, Ben Fissel, and Darin Jones. November 2020. Assessment of the Pacific ocean perch stock in the Gulf of Alaska.
2. Peter-John F. Hulson, Dana H. Hanselman, Chris R. Lunsford, Ben Fissel, and Darin Jones. November 2019. Assessment of the Pacific ocean perch stock in the Gulf of Alaska.
3. Peter-John F. Hulson, Dana H. Hanselman, Chris R. Lunsford, and Ben Fissel. November 2017. Assessment of the Pacific ocean perch stock in the Gulf of Alaska.
4. Summary and response to the 2013 CIE review of AFSC rockfish.
5. Sven Kupschus. 2013. CIE Review of Alaska Rockfish Assessments, 9-11 April 2013, Alaskan Fisheries Science Center, Juneau, Alaska.
6. Neil Klaer. 2013. CIE Reviewer's Independent Report on Alaska Rockfish Stock Assessment.
7. Cathy Dichmont. 2013. AFSC Rockfish Assessment Review.

Relevant research

1. 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.
2. Thorson, J.T., Cunningham, C.J., Jorgensen, E., Havron, A., Hulson, P.J.F., Monnahan, C.C. and von Szalay, P., 2021. The surprising sensitivity of index scale to delta-model

assumptions: Recommendations for model-based index standardization. Fisheries Research, 233, p.105745.

3. Thorson, J.T., 2019. Guidance for decisions using the Vector Autoregressive Spatio-Temporal (VAST) package in stock, ecosystem, habitat and climate assessments. Fisheries Research, 210, pp.143-161.
4. Thorson, J.T., Shelton, A.O., Ward, E.J. and Skaug, H.J., 2015. Geostatistical delta-generalized linear mixed models improve precision for estimated abundance indices for West Coast groundfishes. ICES Journal of Marine Science, 72(5), pp.1297-1310.

Other

1. Cahalan, J., J. Mondragon, and J. Gasper. 2010. Catch sampling and estimation in the Federal groundfish fisheries off Alaska. U.S. Dep. Commer., NOAA Tech. Memo. NMFS-AFSC-205, 42 p.

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

Virtual External Independent Peer Review

Gulf of Alaska 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 Gulf of Alaska Pacific ocean perch provides the scientific basis for the management advice considered and implemented by the North Pacific Fisheries Management Council. An independent review of this integrated stock assessment is requested by the Alaska Fisheries Science Center's (AFSC) Auke Bay Laboratories Division (ABL). The goal of this review will be to ensure that the stock assessment represents 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 statistical age-structured stock assessment methods in general and, in particular, assessments developed with software such as ADMB.

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 the application of spatio-temporal models to population index estimation;
- Experience with application of acoustic data collection within stock assessment;
- 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 make all necessary background information and reports available electronically 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. Prior to the peer review, the CIE reviewers will participate in a test to confirm that they have the necessary technical (hardware, software, etc.) capabilities to participate in the virtual panel in advance of the review meeting. The AFSC NMFS Project Contact will provide the information for the arrangements for this test.

3. Attend and participate in the panel review meeting. The meeting will consist of presentations and discussions with the stock assessment authors, NMFS observer program staff, and survey scientists to facilitate the review. 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 by the specified deadline.

Place of Performance

This review will be conducted via virtual meeting software.

Period of Performance
The period of performance shall be from the time of award through April 14 2020. 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.

Within two weeks of award	Contractor selects and confirms reviewers
Approximately 2 weeks later	Contractor provides the pre-review documents to the reviewers
March 30-April 1, 2021	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
No travel is necessary, as this meeting is being held remotely.

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

Project Contact(s):
Pete Hulson
Auke Bay Laboratories
NMFS, Alaska Fisheries Science Center
17109 Point Lena Loop Rd., Juneau, AK, 99801
Phone: 907-789-6060
pete.hulson@noaa.gov

Appendix 2.1: Peer Review Report Requirements

1. The report must be prefaced with an Executive Summary providing a concise summary of the findings and recommendations, and specify whether the science reviewed is the best scientific information available.
2. The report must contain a background section, description of the individual reviewers' roles in the review activities, summary of findings for each TOR in which the weaknesses and strengths are described, and conclusions and recommendations in accordance with the TORs.
   a. Reviewers must describe in their own words the review activities completed during the panel review meeting, including a brief summary of findings, of the science, conclusions, and recommendations.
   b. Reviewers should discuss their independent views on each TOR even if these were consistent with those of other panelists, but especially where there were divergent views.
   c. Reviewers should elaborate on any points raised in the summary report that they believe might require further clarification.
   d. Reviewers shall provide a critique of the NMFS review process, including suggestions for improvements of both process and products.
   e. The report shall be a stand-alone document for others to understand the weaknesses and strengths of the science reviewed, regardless of whether or not they read the summary report. The report shall represent the peer review of each TOR, and shall not simply repeat the contents of the summary report.
3. The report shall include the following appendices:

Appendix 1: Bibliography of materials provided for review.

Appendix 2: A copy of this Performance Work Statement

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

Appendix 2.2: Terms of Reference for the Peer Review

1. Evaluate the data used in the assessments, specifically trawl survey estimates of biomass, and recommend how data should be treated within the assessment model
2. Evaluate the stock assessment model for GOA Pacific ocean perch in general and comment on appropriateness of parameter estimates to assess stock status determinations
3. Evaluate the strengths and weaknesses in the stock assessment model for GOA Pacific ocean perch, and recommend any improvements to the assessment model.
4. Evaluate and recommend how survey data are used for biomass indices within the assessment. Specifically, advise on trawl survey indices arising from design-based methods versus model-based approaches.
5. Evaluate abundance estimates from summer acoustic-trawl data, and recommend how it may be used within the assessment.

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	Paul Spencer	AFSC, Seattle
Members	Noel Cadigan	Memorial University, St. Johns, Newfoundland, Canada
	Geoff Tingley	Gingerfish Ltd, Wellington, New Zealand
	Saang-Yoon Hyun	Pukyong National University, Busan, South Korea

Attendees

Patrick Ressler	AFSC, Resource Assessment and Conservation Engineering, Seattle
Denise McKelvey	AFSC, Resource Assessment and Conservation Engineering, Seattle
Darin Jones	AFSC, Resource Assessment and Conservation Engineering, Seattle
Kari Fenske	AFSC, Auke Bay Lab, Juneau
Dana Hanselman	AFSC, Auke Bay Lab, Juneau
Kristin McQuaw	Alaska Groundfish Data Bank, Newport, OR
Cindy Tribuzio	AFSC, Auke Bay Lab, Juneau
Madison Hall	AFSC postdoc, Resource Assessment and Conservation Engineering
Julie Bonney	Alaska Groundfish Data Bank, Kodiak
Dan Goethel	AFSC, Auke Bay Lab, Juneau
Chris Lunsford	AFSC, Auke Bay Lab, Juneau
Ben Williams	AFSC, Auke Bay Lab, Juneau
Pete Hulson	AFSC, Auke Bay Lab, Juneau
Wayne Palsson	AFSC, Resource Assessment and Conservation Engineering, Seattle
Jason Conner	AFSC, Resource Assessment and Conservation Engineering, Seattle
Tom Holland	AFSC, Resource Ecology and Fisheries Management, Seattle
Chris Gburski	AFSC, Resource Ecology and Fisheries Management, Seattle
Jim Thorson	AFSC, Habitat and Ecological Processes Research, Seattle
Jim Ianelli	AFSC, Resource Ecology and Fisheries Management, Seattle