Center for Independent Experts (CIE) Independent Peer Review of "Assessments of the Pacific ocean perch (*Sebastes alutus*) and other rockfish stocks in the Gulf of Alaska"

Virtual meeting

30 March – 1 April, 2021, Pacific Daylight Time

Prepared by Saang-Yoon Hyun

Prepared for the Center of Independent Experts

College of Fisheries Sciences

Pukyong National University

Busan, 48513, Korea

Email: shyun@pknu.ac.kr (alternative email: shyunuw@gmail.com)

Contents

Executive Summary

I accepted an invitation from the Center for Independent Experts (CIE) to a review of the assessments of Pacific ocean perch (POP) (*Sebastes alutus*), northern rockfish (*Sebastes polysppinis*), and dusky rockfish (*Sebastes ciliatus*) stocks in the Gulf of Alaska (GOA). The review meeting was held virtually from 10 am to 5 pm during March 30 - April 1, 2021, Pacific Daylight Time (March 31 - April 2, Korea time, my location). Three CIE reviewers (Noel Cadigan, Geoff Tingley, and myself) participated in the virtual meeting. The assessment team ("they", hereafter) in the Alaska Fisheries Science Center responded well to CIE reviewers' questions and requests during the virtual meeting. The meeting agenda were presented well as planned.

Findings by Term Of Reference (TOR)

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

POP were sampled with other groundfish by the NMFS bottom trawl survey, which had been deployed triennially during 1990-1999, and biennially during 2001-2019. NMFS used a stratified sampling design for the bottom trawl survey, where 59 strata were defined and at least two stations (sampling units) per stratum were allocated. Overall their calculation of a survey index (a relative population size in number or weight) and the likelihood function for the annual survey indices were not different from the universal practice. I raise two issues here: (i) whether the current trawlable area was sufficient, and (ii) the uncertainty of the survey indices.

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

Their age-structured assessment model was rooted in the traditional statistical catch at age analysis (SCAA). Some features of their assessment model different from the traditional SCAA included using data about lengths of fish caught by the commercial fishery. They estimated a total of 164 free parameters. They have six likelihood functions, given data about the fishery catches, age and length compositions in the fishery catches, survey indices, and maturity-at-age. Further they have five penalized likelihood functions for annual recruitment deviations, annual fishing mortality deviations, recruitment variability, natural mortality, and survey catchability. Those penalized likelihood functions could be viewed as priors in Bayesian statistics. They used software ADMB (Fournier et al. 2012) for numerical optimization where they estimated not only parameters but also the uncertainties of parameter estimates. I acknowledge and support their stock assessment model, making a few suggestions: (i) priors; (ii) typos in all penalized likelihood functions in the assessment report; (iii) the prior form is questionable even after the typo is corrected; and (iv) model goodness-of-fit and unnecessary/confusing expressions.

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.

Overall, their methodology was thorough, and they have continued to improve their assessment with state-of-the-art methods, adding new features such as length-to-age transition information. The process of determining ages of fish is laborious and costly, extracting otoliths, breaking-burning them, and reading their annuli. For these reasons, it is appropriate that they determined fish ages every other year instead of every year. When actual age composition data were not available in years, they (internally in the model framework) inferred age compositions of a population using the length-to-age transition information in those years. They used ADMB as optimization software. ADMB and TMB are sophisticated software, but users of the software are required to be proficient in computer language syntax as well as mathematical statistics used in the stock assessment model. They should be praised for this high proficiency. However, the assessment report lacked formal model validation, and they did not incorporate the fishery cpue data. I also suggest they should do a retrospective error analysis for estimates of annual fishing mortality as well.

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.

Survey data used in the POP assessment (Hulson et al. 2020) were mainly from the bottom trawling, which had been made under a stratified sampling design, triennially during 1990-1999, and biennially during 2001-2019 (Figure 1). One of the major purposes of the survey was to detect a relative population size (a survey index) and to collect biological information such as body size- and age-compositions. There is room for improvement in the calculation of the survey index because a considerable proportion of the population area was not trawled (see the above section, "Whether or not the current trawlable area was sufficient" under TOR 1) and the spatial distribution and density of POP were heterogeneous over the entire population area.

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

Acoustics are an efficient technique for detecting the sizes of a fish school, but acoustic measurements are merely supplementary in the stock assessment framework for the following reasons. Data from an acoustic survey without trawling provide limited information, because we cannot identify fish species, body sizes and ages. An acoustic survey even with trawling still fails to scan in close proximity to untrawlable seabed, and thus such an operation is not appropriate for detecting groundfish populations such as GOA POP. This problem could be relieved with an aid, deployment of lowered stereo camera (Jones et al. 2021).

Recommendations

I provide a list of recommendations.

Background

I accepted an invitation from the Center for Independent Experts (CIE) to a review of the assessments of Pacific ocean perch (POP) (*Sebastes alutus*), northern rockfish (*Sebastes polysppinis*), and dusky rockfish (*Sebastes ciliatus*) stocks in the Gulf of Alaska (GOA). The review meeting was held virtually from 10 am to 5 pm during March 30 - April 1, 2021, Pacific Daylight Time (2 am to 9 am, March 31 - April 2, Korea time, my location). The virtual nature of the meeting was due to the Coronavirus pandemic. Three CIE reviewers (Noel Cadigan, Geoff Tingley, and myself) participated in the virtual meeting. The assessment team ("they", hereafter) in the Alaska Fisheries Science Center responded well to the CIE reviewers' questions and requests during the virtual meeting. Of the assessment team members, Paul Spencer as the meeting chair, Pete Hulson as the assessment report's senior author, and Jim Ianelli attended all meetings. The other participants are listed in Appendix 3. The CIE reviewers were provided with materials about assessment of GOA POP and other rockfish a few days before the virtual meeting. Those materials included the GOA POP assessment report, ADMB files (TPL, DAT, CTL files), and Power Point files about data collection, biology, bottom trawl surveys, and acoustic surveys. The CIE reviewers had been also given oral presentations during the virtual meeting (see Appendix 1 for presentation titles and presenters).

Review activities

The CIE requested a review of the assessments of POP, northern rockfish, and dusky rockfish stocks in the GOA. However, the review meeting's focus was on the GOA POP stock assessment. The meeting agenda was presented as planned (see "Annex 3" under "Appendix 2. Performance Work Statement"). A glitch occurred during the 1st day of the virtual meeting because I was unfamiliar with software, "Cisco Webex", used for the virtual meeting. Except for the glitch moments, I was able to listen to and speak to participants and watch presentation materials through the software. The meeting was open to the public, but participants were limited to CIE reviewers and the assessment team.

Findings by Terms Of Reference

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

POP were sampled with other groundfish by the NMFS bottom trawl survey, which had been deployed triennially during 1990-1999, and biennially during 2001-2019. NMFS used a stratified sampling design for the bottom trawl survey, where 59 strata were defined and at least two stations (sampling units) per stratum were allocated. Overall, their calculation of a survey index (a relative population size in number or weight) and the likelihood function for the annual survey indices were not different from the universal practice. I raise two issues here: (i) whether the current trawlable area was sufficient, and (ii) the uncertainty of the survey indices.

(1) Whether or not the current trawlable area was sufficient

One of the major purposes of the bottom trawl survey was to sample POP (known data) to infer the population (unknown parameter). Under a good sampling design (with a random sampling and a large sample size), an estimate of the population size inferred from the samples (POP caught by the survey) would be close to the true (unknown) population size.

Unfortunately, 'random' sampling was impossible mainly because some areas such as high relief substrates were not trawlable. Their sampling unit was based on whether or not the area was trawlable as well as on the other criteria (e.g., abundance, economic value). In addition to the lack of a random sampling, the assumption was weak that the POP density would be the same between trawlable and untrawlable areas. Although the assumption was not explicitly stated, the extrapolation of a survey index from a trawlable area to an untrawlable area would not have been justified without the assumption.

One of the ways to overcome these above concerns would be to sample as much area as possible. But the possible sampling area in the latest survey year 2019 was 'at most' 40.9% of the total population area of the survey regions that was 320,000 km^2, because the known area was 50.1% and the trawlable area of the known area was 81.6%: i.e., 40.9% = (50.1/100) x (81.6/100). It would have been practically impossible to fully fill stations (sampling units) in the possible sampling area, and thus the sampled area was 'much less than' 40.9% of the total area.

(2) Uncertainty of the survey indices.

It was reasonable to consider heteroscedasticity in the likelihood for the annual survey indices (annual relative population sizes). That is, they allowed the variance of a survey index to differ by year instead of considering its constant variance over years. Such a good implementation of heteroscedasticity was possible because they calculated time-varying uncertainty (e.g., annual coefficient of variation (CV)) of the survey indices. However, the CV's of the annual survey indices are questionable because they did not consider the covariances of survey indices from neighbor strata when calculating the variance of a total of survey indices over all strata in a survey year. These CV values seemed too small (closed dots in Figure 1b, where all values were below 53% in 1999). For this reason, I suggest they should revise the calculation of the CV's of the annual survey indices.

However, if this revision is difficult for some reason, then they may return to the homoscedasticity likelihood for the survey data. Regardless of whether time-varying variances, or time-constant variance is considered, the variance term would 'eventually' become part of the weight of the likelihood, although they separately expressed the weight and the variance as $\lambda_{\hat{I}}$ and $\sigma^2_{I,y}$ respectively in their report (Table 9-13 of Hulson et al. (2020)). In other words, either heteroscedasticity or homoscedasticity is masked by the likelihood weight.

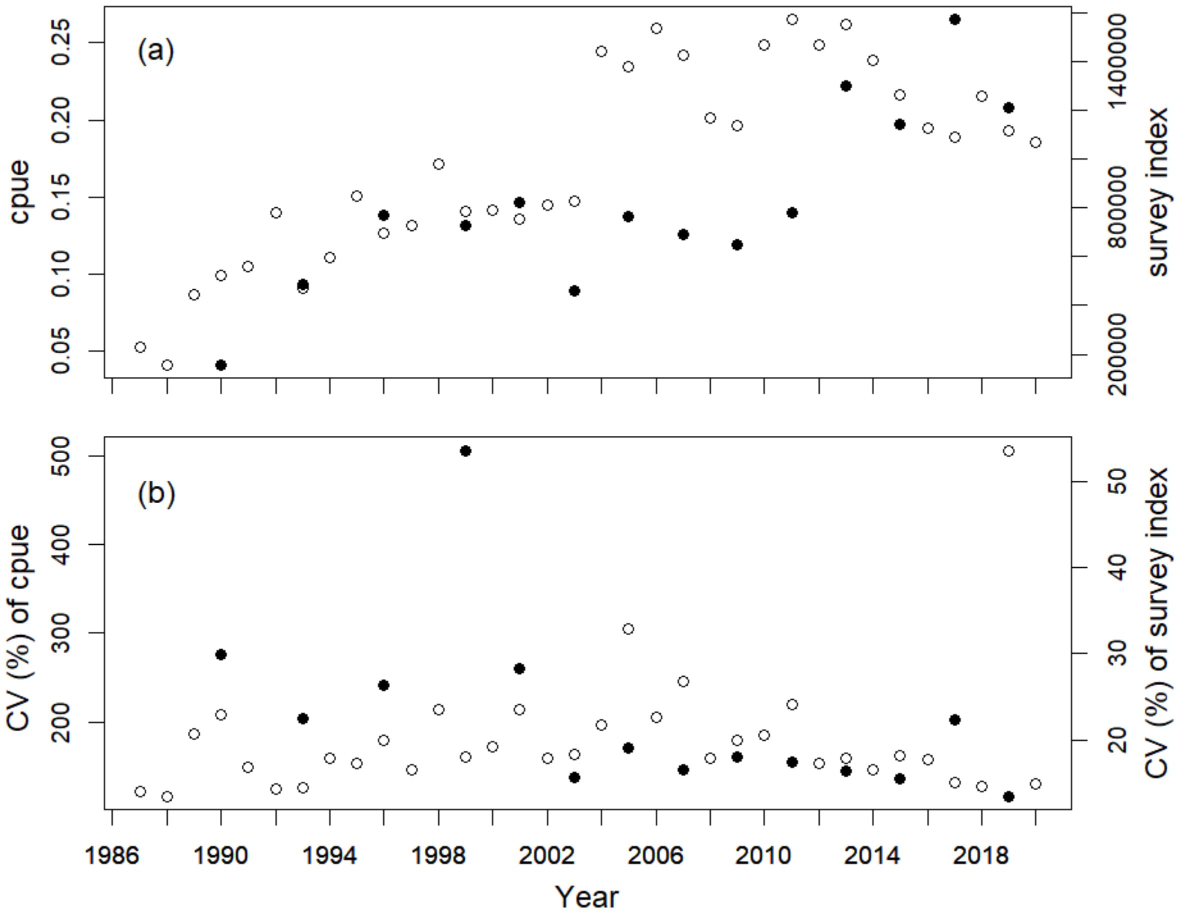

Figure 1. Comparison in relative population sizes (panel a) and their CVs (panel b) between the fishery cpue (ton/min) (left y-axis) and the bottom trawl survey index (right y-axis). Open circles represent the fishery cpue values and their CVs, while closed dots indicate the bottom trawl survey indices and their CVs.

On the other hand, I would appreciate their description of how they calculated the fishery cpue and its uncertainty (e.g., variance). The CVs of the fishery cpue's appeared too large: e.g., the mean of 180%, ranging from 116% to 505% (open circles in Figure 1b).

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

Their age-structured assessment model was rooted in the traditional statistical catch at analysis (SCAA). Some features of their assessment model different from the traditional

SCAA included using data about lengths of fish caught by the commercial fishery. They estimated a total of 164 free parameters. They had six likelihood functions, given data about the fishery catches, age and length compositions in the fishery catches, survey indices, and maturity-at-age. Further they had five penalized likelihood functions for annual recruitment deviations, annual fishing mortality deviations, recruitment variability, natural mortality, and survey catchability. Those penalized likelihood functions could be viewed as priors in Bayesian statistics. They used software ADMB (Fournier et al. 2012) for numerical optimization where they estimated not only parameters but also the uncertainties of parameter estimates.

However, they did not simultaneously estimate all parameters associated with the assessment within the model framework. Instead, they externally estimated body growth parameters, which were used for the calculation of a size (length or weight)-to-age transition. The practice is not rare though, where all parameters are not estimated at the same time within the model framework. Although some fastidious reviewers could criticize the practice as underestimating the uncertainties of parameter estimates (e.g., a previous CIE reviewer), I find it acceptable, because it would be extremely difficult to carry all uncertainties of parameter estimates in a complicated model such as the GOA POP assessment model. A practical solution to the problem would be to consider additional uncertainty as a buffer for parameter estimates, especially uncertainties in harvest management references. Most managers already apply this idea by having the allowable biological catch (ABC) be lower than but not equal to the maximum sustainable yield (MSY): i.e., ABC $<$ MSY rather than ABC $\leq$ MSY.

I acknowledge and support their stock assessment model, making a few suggestions.

(1) Priors

They applied priors of natural mortality (M), catchability of the bottom trawl survey (q), and recruitment variability (σ_r), respectively for the assessment. The prior point value and CV for M, q, and σ_r were as follows: 0.0614/year with CV of 10%, 1.15 with CV of 45%, and 1.7 with CV of 20%, respectively. I am a bit surprised that the larger weight (i.e., the lower CV) was given to M while the smaller weight (i.e., the higher CV) was assigned to q. Recalling that natural mortality (M) cannot be controlled to any degree and varies by year as well as by age while catchability (q) could be somewhat controlled and the survey gear has been standardized over time, I wonder about their reasoning behind the strong belief given to the prior of M vs. the weak belief assigned to the prior of q. Also the prior value of σ_r, 1.7 seemed to be too large because it was in the natural logarithm: e.g., $\log N_{2,y} = \mu_r + \varepsilon_y{}^r$, where ε^r ~ N(0, $\sigma_r{}^2$).

More importantly I suggest they should try to estimate M without adding its prior, treating it as a free parameter under a state-space framework. Natural mortality is one of the key parameters that derive the population dynamics. One of the merits of a state-space model framework lies in the estimability of M without incorporating its prior information.

(2) Typos in all penalized likelihood functions in the assessment report

Out of many typos in the assessment report (Hulson et al. 2020) that I found, I would like them to correct the typos in all penalized likelihood functions. I had been confused by them until I checked TPL file in ADMB (file name: pop.tpl). I show corrected equations below.

The left column (with typos) ➔ The right column (correct)

$$L_\theta = \frac{1}{2\cdot\sigma_\theta{}^2}\cdot\ln\left(\frac{\theta}{\theta_{prior}}\right)^2 \quad ➔ \quad L_\theta = \frac{1}{2\cdot\left[CV(\theta)\right]^2}\cdot\left[\ln\left(\frac{\theta}{\theta_{prior}}\right)\right]^2 \qquad (1)$$

$$L_r = \lambda_r\left(\frac{1}{2\cdot\sigma_r{}^2}\cdot\sum_Y \varepsilon_y{}^r + Y\cdot\ln\sigma_r\right) \quad ➔ \quad L_r = \lambda_r\left(\frac{1}{2\cdot\sigma_r{}^2}\cdot\sum_Y\left(\varepsilon_y{}^r\right)^2 + Y\cdot\ln\sigma_r\right) \qquad (2)$$

$$L_f = \lambda_f\cdot\sum_Y(\varepsilon_Y^f) \quad ➔ \quad L_f = \lambda_f\cdot\sum_Y(\varepsilon_Y^f)^2 \qquad (3)$$

(3) The prior form is still questionable even after the typo is corrected

What they meant with the prior form for M, q, and σ_r was at the right side in eq 1. If they don't agree with me, then they must check TPL file: e.g., see line 943 – 952 in "pop.tpl". Even after correcting the typo, I still raise a question why they directly put the "$[CV(\theta)]^2$" in the denominator. The form, L_θ in eq 1 looks based on the likelihood with log(θ) being treated as a normal random variable and constants with respect to θ being ignored. Then, they should have substituted "the variance of log(θ)" in place of "$[CV(\theta)]^2$".

The form is wrong even if they might argue it is a non-parametric form. First, there is a dimension problem. The variance of a random variable has its dimension and unit whereas the CV is dimensionless. For example, M_{prior} of 0.0614 is in "per year" while q_{prior} of 1.15 is dimensionless. Thus, the variance of M_{prior} is in "per year^2" while the variance of q_{prior} is dimensionless. Second, simply the form does not make sense in terms of a parametric likelihood function.

I suggest they should revise the prior form to be defensible. One of the efficient revision methods would be to analytically derive E{log(θ)} and Var{log(θ)} by delta method. They assumed the prior value of θ and its corresponding CV. And thus it would be reasonable to treat θ_{prior} as the mean of θ, and to calculate the variance of θ using the given CV. That is, we have that $E\{\theta\} = \theta_{prior}$ and $Var\{\theta\} = Var\{\theta_{prior}\}$. Then, assuming that log(θ) ~ *Normal*(E{log(θ)}, Var{log(θ)}), it is straightforward to identify these first and second moments by delta method: i.e.,

$$E\{\log(\theta)\} \approx \log(\theta_{prior}) - \frac{Var\{\theta_{prior}\}}{2\cdot\left(\theta_{prior}^{2}\right)}$$

$$Var\{\log(\theta)\} \approx \frac{Var\{\theta_{prior}\}}{\theta_{prior}^{2}} \tag{4}$$

Finally the resultant revised prior form is eq. 5 in terms of the "negative" log-likelihood function with constants with respect to θ being ignored:

$$L_{\theta} = \frac{\left(\log\theta - E\{\log(\theta)\}\right)^2}{2\cdot Var\{\log(\theta)\}} \tag{5}$$

Needless to say, substitute two moments in eq. 4 for the corresponding moments in eq. 5.

(4) Model goodness-of-fit and unnecessary/confusing expressions

The model's predicted values of trawl survey indices did not well reflect the fluctuation of those survey indices. I suggest that they should remove the uncertainty measures, such as their term, 95% confidence intervals (CIs) (shaded area) in Figure 9-4 in Hulson et al. (2020) for a few reasons. First, the illustration provided an incorrect impression that the 95% CIs (their term) were too narrow to cover the survey indices. Indeed, the CIs often failed to cover the survey indices in the figure, and I fear that the illustration would incorrectly imply a poor performance of their assessment model. Second, since survey indices were used as input data with observation errors in the assessment model framework, you should not use the term, "confidence interval" because it incorrectly implies the survey indices were parameters, not data.

However, it is interesting that their assessment report shows a good fit with the survey indices in Figure 9-18 (Hulson et al. 2020) without describing its background. They simply mention "random effects estimates" and "random effects model" in the legend and caption of Figure 9-18. If they used a random-effects model for the stock assessment, then they should have described in the report what random effects were, what parameters were as opposed to random effects, and how they optimized the model. The assessment team's oral presentation did not describe them either in the CIE virtual meeting. Even in Figure 9-18, the confusing the term, "confidence interval" still remains as described above.

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.

3.1. Strengths

They have continued to collect quality data about survey indices or fishery cpue's, fish age, fish body size (length, weight), and sexual maturity through survey and fishery operations, and laboratory work for the identification of fish ages. Given these data, they applied an age-structured assessment model. Overall their methodology was thorough and they have continued to improve their assessment with state-of-the-art methods, adding new features such as length-to-age transition information. The process of determining ages of fish is laborious and costly, extracting otoliths, breaking-burning them, and reading their annuli. For these reasons, it is appropriate that they determined fish ages every other year instead of every year. When actual age composition data were not available in years, they (internally in the model framework) inferred age compositions of a population using the length-to-age transition information in those years.

They used ADMB as optimization software. ADMB is sophisticated software but users of the software are required to be proficient in computer language syntax as well as mathematical statistics used in the stock assessment model. They should be praised for this high proficiency.

3.2. Weaknesses and suggestions

The assessment report lacks formal model validation. The assessment report (Hulson et al. 2020) has a section, "Model evaluation" under "Results", but the description in the section was about the model selection based on the model's goodness-of-fit using the total loglikelihood value (i.e., the objective function value), parameter parsimony, biological patterns of recruitment, catchability and selectivity, and a visual examination of the length and age compositions between the model's predicted values and data. I accept their model selection procedure. However, a model with a good balance (e.g., a low value of AIC) between goodness-of-fit and parameter parsimony does not guarantee low bias and high precision in parameter estimates. For this reason, I suggest that they should perform formal model validation, setting true values of free parameters, generating pseudo data, feeding those simulated data into the assessment model, estimating parameters, and comparing estimates of free parameters with the corresponding true values. Such model validation would help us to judge the reliability of the parameter estimates (and the resultant derived quantities) made by the model. The "Simulation study" of Miller and Hyun (2018) would be an example for model validation.

The assessment's time step is a year but their bottom survey has not been made every year (see closed dots in Figure 1). it is unclear why they did not synthesize the fishery cpue's and the survey indices especially when the annual survey indices were not available. Their ADMB code (TPL, DAT files) shows they did not use the fishery cpue data. Should they not synthesize all available data and information in a single model framework? Data collected by a survey, and caught by a fishery are samples of a population in a statistics context, and fishery cpue's are independent of survey indices. Thus, it is not surprising that such data (i.e., different samples) at a time often show different signals at the time (Figure 1a). Although the fishery catch is not a random sample, its cpue should be considered to be a relative population size. If my observation is correct, then they should have synthesized the two kinds of population size indices by incorporating the additional likelihood for the fishery cpue data.

Their retrospective error analysis was made for estimates of the annual spawning stock biomass. They argued that the model improved in terms of a retrospective pattern by showing that the magnitude of the resultant Mohn's rho was lower than before: e.g., |0.15| in 2020 vs. |0.27| in 2019 in the assessment report (Hulson et al. 2020) while |0.22| in 2019 in the virtual meeting. I suggest they should show such analysis for estimates of the annual fishing mortality as well for two reasons. It would be a good practice to confirm whether the signal of the fishing mortality was opposite to that of the annual spawning stock biomass in the relative difference of parameter estimates (Miller and Hyun 2018). Also a change in catches was well known to affect a retrospective pattern (Cadigan and Farrell 2005).

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.

Survey data used in the POP assessment (Hulson et al. 2020) were mainly from the bottom trawling, which had been made under a stratified sampling design, triennially during 1990-1999, and biennially during 2001-2019 (Figure 1). One of the major purposes of the survey was to detect a relative population size (a survey index) and to collect biological information such as body size- and age-compositions. There is room for improvement in the calculation of the survey index because a considerable proportion of the population area was not trawled (see the above section, "Whether or not the current trawlable area was sufficient" under TOR 1) and the spatial distribution and density of POP were heterogeneous over the entire population area.

The catchability of a fish is determined with its availability in a space and its vulnerability to a gear (Jones et al. 2021). If a heterogeneous spatial distribution of POP further differs by time (year), then time-varying catchability (q_t) would be more reasonable in the proportional relationship between a survey index and a population size (in number or weight): i.e., ‘ $I_t = q_t \cdot B_t$’ rather than ‘ $I_t = q \cdot B_t$’. Such time-varying q might improve the model’s overall performance as assessed with metrics such as goodness-of-fit and a retrospective error pattern. However, it would require additional data to implement time-varying catchability, because they would have more parameters (e.g., as many parameters as the number of the survey years) to be estimated. One of the effective ways to get around this problem would be to treat q as a random effect in a state-space model framework.

The assessment report (Hulson et al. 2020) did not show annual survey indices and the corresponding uncertainty made with the model-based (i.e., vector autoregressive spatio-temporal (VAST)) approach. I was not able to see them until the virtual meeting for the CIE review. Substantial differences were found in survey indices and their uncertainty between the design-based (i.e., stratified sampling) and VAST approaches. However they did not describe in detail the cause of the substantial differences.

The VAST method is supposed to eventually replace the design-based approach, because the former resulted in a less biased and more precise estimate of a population density than the survey with stratified sampling (Thorson 2019, Thorson et al. 2021). AFSC also started to implement it. However, it looks uncertain how soon the replacement will be made because there are still limitations of the VAST approach: e.g., sensitivity to spatial scales, trade-off between computation burden and spatial resolution (the number of “knots”), and prior information of catchability covariates for every datum. For this reason, they should continue to deploy the current survey with stratified sampling because the stratified sampling theory has long been established, and it is straightforward to interpret its estimates.

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

Acoustics are an efficient technique for detecting the sizes of a fish school, but acoustic measurements are merely supplementary in the stock assessment framework for the following reasons. Data from an acoustic survey without trawling provide limited information, because we cannot identify fish species, body sizes and ages using acoustic measurements alone. An acoustic survey even with trawling still fails to scan in close proximity to untrawlable seabed, and thus such operation is not appropriate for detecting groundfish populations such as GOA POP. This problem could be relieved with an aid, deployment of lowered stereo camera (LSC) (Jones et al. 2021). Jones et al. (2021) showed the benefit of acoustics and LSC survey tools, and further tried to infer the catchability (q) of a bottom trawl survey from a LSC survey because the value of q could be used as its prior for the stock assessment model framework. For the purposes, Jones et al. (2021) compared in fish density on an trawlable area between LSC and bottom trawl surveys. Unfortunately, there was a substantial difference especially in POP density between those two surveys (Jones et al. 2021). Data from acoustic and/or LSC surveys alone without a 'bottom' trawling survey must not be used for the stock assessment model framework. However, this limitation could be lifted as engineering in acoustics and LSC operation improves in future.

Recommendations

a. If the survey for the POP stock assessment continues to rely on a bottom trawl survey, they should consider increasing the current trawlable area.

b. They should revise the calculation of the CV of annual bottom trawl survey indices (annual relative population sizes) because they failed to consider the covariances of survey indices from neighboring strata when calculating the variance of the annual survey index.

c. They should incorporate the annual fishery cpue's into the assessment model framework.

d. They should improve the model fit to the survey indices. One of the efficient ways to improve the goodness-of-fit might be to consider process errors in state variables (random effects).

e. The penalized likelihood form (eq. 1) as the prior of M, q, and σ_r must be revised (beyond the typo). The revised form (eqs. 4 and 5), which I suggest above, might improve the model performance.

f. They should do formal model validation, setting true values of free parameters, generating pseudo data, feeding those simulated data into the assessment model, estimating parameters, and comparing estimates of free parameters with the corresponding true values. Such model validation would help us to judge the reliability of parameter estimates and the resultant derived quantities made by the model.

g. For the retrospective error analysis, they should also examine estimates of annual fishing mortality.

References

Cadigan, N. G., and P. J. Farrell. 2005. Local influence diagnostics for the retrospective problem in sequential population analysis. ICES Journal of Marine Science **62**:256-265.

Fournier, D. A., H. J. Skaug, J. Ancheta, J. Ianelli, A. Magnusson, M. N. Maunder, A. Nielsen, and J. Sibert. 2012. AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. Optimization Methods and Software **27**:233-249.

Hulson, P.-J. F., C. R. Lunsford, B. Fissel, and D. Jones. 2020. Assessment of the Pacific ocean perch stock in the Gulf of Alaska. NPFMC Gulf of Alaska SAFE.

Jones, D. T., C. N. Rooper, C. D. Wilson, P. D. Spencer, D. H. Hanselman, and R. E. Wilborn. 2021. Estimates of availability and catchability for select rockfish species based on acoustic-optic surveys in the Gulf of Alaska. Fisheries Research **236**:105848.

Miller, T. J., and S.-Y. Hyun. 2018. Evaluating evidence for alternative natural mortality and process error assumptions using a state-space, age-structured assessment model. Canadian Journal of Fisheries and Aquatic Sciences **75**:691-703.

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**:143-161.

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

Appendix 1. Bibliography of materials provided for review

(1) Documents

AFSC. 2013. Summary and response to the 2013 CIE review of AFSC rockfish.

Dichmont, C. 2013. AFSC Rockfish Assessment Review.

Hulson, P.-J. F., D. H. Hanselman, C. R. Lunsford, and B. Fissel. 2017. Assessment of the Pacific ocean perch stock in the Gulf of Alaska. NPFMC Gulf of Alaska SAFE.

Hulson, P.-J. F., D. H. Hanselman, C. R. Lunsford, B. Fissel, and D. Jones. 2019. Assessment of the Pacific ocean perch stock in the Gulf of Alaska. NPFMC Gulf of Alaska SAFE.

Hulson, P.-J. F., C. R. Lunsford, B. Fissel, and D. Jones. 2020. Assessment of the Pacific ocean perch stock in the Gulf of Alaska. NPFMC Gulf of Alaska SAFE.

Jones, D. T., C. N. Rooper, C. D. Wilson, P. D. Spencer, D. H. Hanselman, and R. E. Wilborn. 2021. Estimates of availability and catchability for select rockfish species based on acoustic-optic surveys in the Gulf of Alaska. Fisheries Research 236:105848.

Klaer, N. 2013. CIE Reviewer's Independent Report on Alaska Rockfish Stock Assessment.32.

Kupschus, S. 2013. CIE Review of Alaska Rockfish Assessments, 9-11 April 2013, Alaskan Fisheries Science Center, Juneau, Alaska.

Thorson, J. T., A. O. Shelton, E. J. Ward, and H. J. Skaug. 2015. Geostatistical delta-generalized linear mixed models improve precision for estimated abundance indices for West Coast groundfishes. ICES Journal of Marine Science 72:1297-1310.

Thorson, J. T. 2019. Guidance for decisions using the Vector Autoregressive Spatio-Temporal (VAST) package in stock, ecosystem, habitat and climate assessments. Fisheries Research 210:143-161.

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

Thorson, J. T., A. O. Shelton, E. J. Ward, and H. J. Skaug. 2015. Geostatistical delta-generalized linear mixed models improve precision for estimated abundance indices for West Coast groundfishes. ICES Journal of Marine Science 72:1297-1310.

(2) Oral presentation slides

Gulf of Alaska Bottom Trawl Survey. Presenter: Wayne Palsson

VAST estimates of Pacific Ocean Perch. Presenter: Jason Conner

North Pacific Observer Program. Presenter: Tom Holland

Age determination of Gulf of Alaska Pacific Ocean Perch (*Sebastes alutes*), Northern Rockfish (*Sebastes polyspinus*), and Dusky Rockfish (*Sebastes ciliatus*) at the Alaska Fisheries Science Center (AFSC). Presenters: Chris Gburski and Delsa Anderl

MACE Program, Gulf of Alaska Acoustic-Trawl Survey, POP Abundance and Catchability. Presenter: Patrick Ressler

(3) ADMB files via github

pop.tpl

goa_pop_2019.dat

goa_pop_2019.ctl

Appendix 2. Performance Work Statement

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

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) 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.

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 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

Annex 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.

Annex 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.

Annex 3: Tentative Agenda

Review of Gulf of Alaska Pacific ocean perch Stock Assessment
March 30 - April 1, 2021

Alaska Fisheries Science Center, Meeting link (virtual meeting)

For further information:
Pete.Hulson@noaa.gov, Paul.Spencer@noaa.gov

Presentations

Day 1:

Overview of rockfish biology, fishery, and history of assessment
Bottom trawl survey
VAST approach to survey indices
Fishery data
Age and growth
Input data version 2

Day 2 and 3:
Model presentation
Acoustic survey and catchability (video)
Acoustic survey and catchability (slides)
Vast history etc

Schedule

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

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