Independent Peer Review of *Gulf of Alaska Pacific Cod Stock Assessment*

Prepared for:
The Center for Independent Experts

July 2018

Dr Kevin Stokes
Stokes.net.nz Ltd
59 Jubilee Rd
Khandallah
Wellington 6035
New Zealand
Ph: +64 (04) 973 7305
E-mail: kevin@stokes.net.nz

Executive Summary

The report must be prefaced with an Executive Summary providing a concise summary of the findings and recommendations, and specify whether or not the science reviewed is the best scientific information available.

The terms of reference for the review of Gulf of Alaska Pacific cod do not explicitly consider typical matters such as reference point definition, projections, and whether the science is best available to inform management. Rather, they primarily seek input on matters related to complexity that are being considered in development of the model to inform management.

Issues considered include use of additional surveys now and in the future, weighting, and the appropriate degree of model complexity to account for changes in natural mortality, selectivity, and survey catchability.

Data for the assessment are limited and there are problems with ageing protocols. Though not considered in the terms of reference, resolving ageing protocols and ageing historical ageing samples would be highly advantageous. There are currently no fishery samples aged, and age data are only available for the tri/biennial AFSC trawl survey. ADFG and IPHC surveys provide some possibility of improving the fishery independent data available. Including the ADFG surveys does not appear currently to offer any advantages. The IPHC setline survey, however, in principle could offer the best prospects for improving the assessment, especially if ageing is undertaken. Work will need to be undertaken in 2018 and 2019 to confirm the utility of the survey.

The degree of appropriate complexity for any model depends on its purpose. Stock assessment models are carried out to inform management, and determining an appropriate level of complexity depends on the management utility provided. The stock assessment team has provided a high quality SAFE report in December 2017 that carefully considers model building with scrutiny of diagnostics at each step. The development has looked at model fit, but has thus far not considered appropriate model complexity in terms of utility.

My view is that while some degree of blocking is appropriate, the data and utility considerations should limit any final model to the extent that advice is robust. Too much complexity, or over-fitting, can lead to very precise but quite wrong results, and it is better in general to use simpler models with associated large confidence intervals. More complex models also provide challenges as to reference point definition (and meaning) and difficulties in providing projections.

Background

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.~~

Gulf of Alaska (GOA) Pacific Cod (*Gadus macrocephalus*) is distributed from northern California to the Eastern Bering Sea (EBS). While migrations are known to occur, discrete stocks are recognised in the EBS, Aleutian Islands and GOA. The GOA stock is managed as a single unit by the North Pacific Fisheries Management Council (NPFMC). Its life history is unremarkable, and it is managed using standard approaches.

The species, however, is known to fluctuate historically with recruitment and abundance tied to climate variations. It is distributed from close inshore to about 500m depth. Spawning occurs between 40-290m and eggs remain close to the bottom until hatching, which is dependent on defined environmental conditions. Larvae are pelagic, and growth and survival are heavily dependent on temperature, presumably related both to metabolic considerations plus climate induced variations in prey and predator fields.

The now well-known warming of the GOA from 2013/14 onwards and the observed decline in Pacific cod abundance/availability in all surveys and fisheries is a potential challenge for the assessment and management. However, data are limited and fitting the decline and determining stock status is relatively uncontentious given robust signals from multiple model configurations reported in the most recent SAFE chapter from December 2017. More challenging, depending on the choice of model complexity, is how to project the impact of alternative management actions if the assessment includes variable or blocked estimates of natural mortality and/or selectivity, and perhaps even using covariates in the estimation process. This is not explored in the SAFE chapter. The ToR for the review do not explicitly address model selection, status reporting, and projections, but are rather set up to seek advice on what degree of assessment complexity is appropriate.

The ToR also do not include anything explicit on data sources. It is worth noting in advance that the fisheries on GOA Pacific cod are split by season and gear, are interactive with other fisheries/stocks, and have been subject to changes in regulatory and economic drivers over time. Selectivity could well change through time. The fisheries, however, are subject to extensive and high-quality data acquisition with excellent scientific and compliance observation. The industry is involved and proactive, and the data available to the assessment are considered highly reliable. It is unfortunate that there have been difficulties in ageing of Pacific cod, and there is a clear need to resolve protocols and age historical fishery otoliths. Absent fisheries and ideally annual survey-derived age composition data, the assessment is reliant on length composition data and limited tri/biennial age data from just the AFSC trawl survey (for which ageing protocols have changed over time). In considering the ToR, it should not be overlooked that perhaps the most important thing to do to improve the assessment would be to resolve ageing protocols and expand the ageing database. Only with those data is it really possible to explore more widely natural mortality estimation and selectivity.

From the outset, it is worth noting that the ToR effectively seek advice on the degree to which mortality, selectivity, availability, and catchability can be estimated and might be included in the assessment. These alias each other and even with extensive data, there will

be no definitively correct model as there is little or no information to distinguish between underlying processes.

Because it may not emerge clearly in the consideration of ToR, it is worth mentioning here that the SAFE report includes a careful approach to model development, with clear consideration of the detailed fit of each model run and what can be concluded. This approach was continued by the lead analyst (Barbeaux) during the review.

Review Process and Activities

d. Reviewers shall provide a critique of the NMFS review process, including suggestions for improvements of both process and products.

The review of the Gulf of Alaska Pacific Cod Stock Assessment was held at the AFSC, Seattle, from 1-4 May, 2018. The Terms of Reference (ToR) are shown at Appendix 2. The list of participants and final agenda are at appendix 3.

There is no standardised NMFS review process common to all regional fisheries management arrangements. The process varies by region. In most regional review systems (e.g., SARC, STAR, SEDAR), the processes are highly formalised and require close adherence to Terms of Reference (ToR), with review products of direct relevance to fisheries management decision making processes. Most reviews conducted through the AFSC are less formal, and do not lead directly into formal decision-making processes, instead being typically used internally within AFSC for further assessment development and deliberations with the relevant Plan Team. The ToR for this review reflect such use. Any comments and suggestions made here are in this context. CIE ToR are coloured blue, while AFSC ToR are coloured green. Recommendations are coloured red.

Reviews which feed into management decision making processes require public notification and opportunities for input. My understanding is that this review is not part of such a formal process, but it is normal AFSC practice to still make a public notification. Industry bodies were contacted to provide an opportunity to make input and were represented during the review meeting with an open opportunity to make comment. For a review of this type, I am comfortable that opportunities and arrangements for public engagement were reasonable.

More formal NMFS reviews conducted through the CIE, for other fisheries centers, include requirements for CIE reviewers to contribute to Summary Reports, which become important process outputs that are used in further management processes. Like many (perhaps nearly all) AFSC reviews, this review did not include a requirement for a Summary Report. This conflicts somewhat with CIE ToR, but is not in practice a problem. From a reviewer perspective, it leaves "loose ends" from the review meeting and creates a different need for further consideration during report writing, but, again, this is not a problem.

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.~~

The review was chaired by Grant Thompson from the AFSC. CIE reviewers were Henrik Sparholt, Jean-Jacques Maguire, and Kevin Stokes. Materials sent in advance (see Appendix 1) were read by all CIE reviewers. The first morning and early afternoon consisted of

background talks on surveys, observers, management history, and other matters. A half day was then provided for presentation of and discussion on the GOA Pacific cod stock assessment. Further presentations and discussion, plus "homework" setting occupied the second day. Day 3 was used for further consideration of model runs and responses, but was delayed somewhat due to discovery during the meeting of an error in a data file, identified when comparing runs with a changed maturity calculation which displayed changes in fit when none should have occurred. While this caused some apparent embarrassment for the AFSC staff, my view is that such errors arise and the AFSC staff dealt well with the matter, identifying the problem and ensuring all model runs affected were rerun and all materials completely checked. I commend the lead analyst and AFSC staff for dealing effectively with the matter.

The meeting was scheduled for four days but only three were required with agreement reached on further work to be run after the third day and be made available to reviewers. The meeting therefore concluded at the end of the third day, allowing time on the fourth for reviewers to work independently. I have in the past been critical that reviews typically try to shoehorn too much in to too little time. It was a welcome change to have time to sit and talk through issues with no time pressure.

The review was greatly helped by the early provision of the main materials (see appendix 1), including presentations by the main analyst (Barbeaux). Like the SAFE report, these were well-considered and of a high standard. Materials were made available using Google Drive. Presentations by other ADFG and AFSC staff were made available at the time or soon after they were made. All were of a good standard and the organisation of materials by AFSC was excellent.

All reviewers participated fully in the meeting and interactions with the lead analyst and other AFSC staff were courteous and focused on the terms of reference. I am aware this was the first review for the lead analyst (Barbeaux), and thank him for his good preparation and openness throughout.

Terms of Reference

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.

See ToR below.

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.

The review was not set up to result in final, consistent views or opinions. Rather, for the GOA Pacific cod stock assessment under consideration, the review comprised of presentations to enable discussion and formulation of views on the ToR after the meeting; it is unclear to what extent the individual panellist's views will converge or diverge.

c. Reviewers should elaborate on any points raised in the summary report that they believe might require further clarification.

The AFSC ToR did not call for a Summary Report. No attempt was made to reach consensus on specific ToR.

Response to AFSC-specified ToR

~~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.

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.

The primary background paper for the review is the stock assessment chapter of the December 2017 SAFE report, which contains management-related information. The large majority of CIE reviews include ToR on management-related issues, such as the appropriateness of reference points, characterisation of uncertainty in management-related quantities, scope of projections, and whether or not the assessment constitutes the best available scientific information. The ToR for this review do not include these matters, but are focused instead on specific questions to assist development of a final stock assessment.

1. Evaluate and provide recommendations on data used in the assessment models. In particular:

a. What are the benefits vs disadvantages of including data from the ADFG small-mesh trawl and the IPHC longline surveys in the assessment?

The current stock assessment for GOA Pacific cod is informed by the AFSC longline and trawl surveys. Both surveys are used to provide relative abundance indices and length compositions. The trawl survey is additionally the sole source of age composition and conditional age-at-length data.

Neither survey, however, is designed for Pacific cod. The AFSC trawl survey is a general purpose multi-species survey of groundfish and invertebrates, covering the full extent of the GOA across a wide depth range, and carried out by two or three commercial vessels under contract each year. From its inception in 1984, the survey was triennial. It became biennial after 1999 (the last triennial/first biennial year). While using standardised protocols, the survey has varied somewhat through time with different vessels, a change from 30-minute to 15-minute duration tows (in 1996), numbers of stations and maximum depth covered. Nevertheless, and notwithstanding that the survey cannot sample a large amount of untrawlable ground, it provides a useful sampling tool for biological materials, environmental changes, and relative abundance and distribution of Pacific cod.

For Pacific cod, a relatively fast-growing species with moderately variable recruitment, it is

unfortunate that the survey is not annual, not just to provide an annual index, but especially to provide annual age compositions.

The AFSC longline survey, in contrast, is annual and does sample untrawlable grounds. However, it is designed primarily for sablefish and other species, uses a systematic as opposed to random design, and does not sample shallow waters (starting at 125m cf the trawl survey sampling from 15m depth). It does, though, include a 150-200m stratum, added specifically to sample Pacific cod.

With known weaknesses in the existing, available survey indices, it is natural to consider alternatives/additions. The potential alternative information sources on GOA Pacific cod have been identified as the ADFG large-mesh trawl surveys, and the IPHC longline survey (as noted in the ToR). Potential benefits to be considered include improved depth/ground coverage, annual availability, improved age samples, and signal detection for young fish.

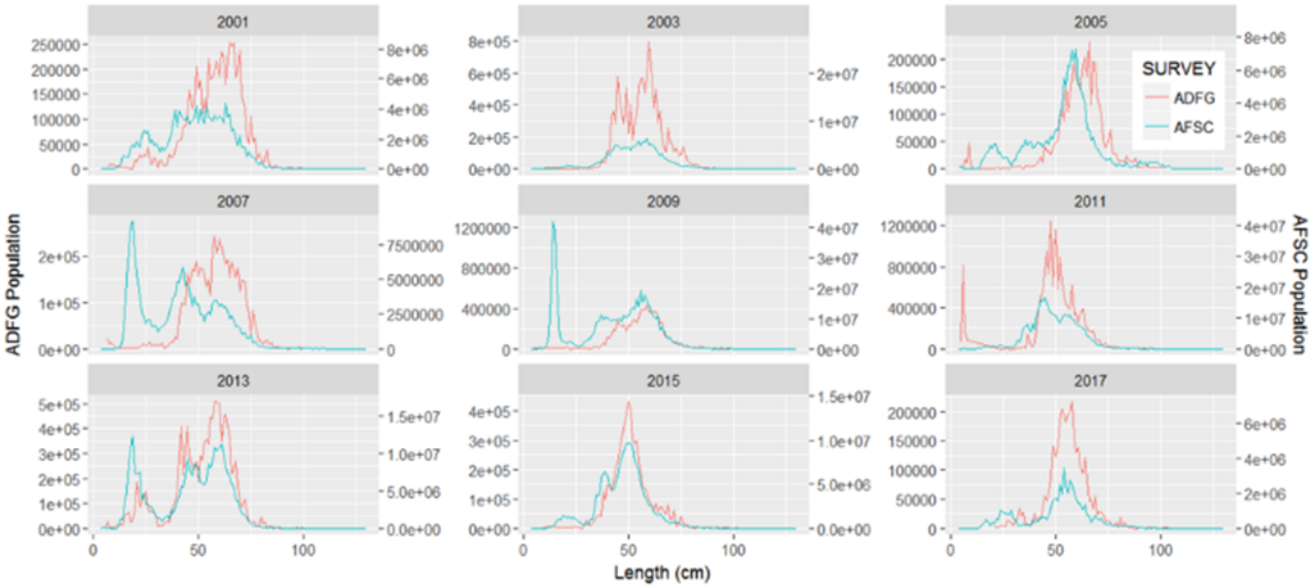
During the review meeting, presentations were made on these surveys.

The ADFG survey has separate Central Region and Westward Region components.

The ADFG Westward Region survey, extending from around Kodiak Island westward to the Aleutian Islands, has taken place annually since 1988. Multiple vessels are used annually with gear and protocols standardised. However, changes in stations/areas have taken place through time and budget constraints appear to have impacted sampling around Kodiak Island in recent years; it is unclear if these changes are of significance for Pacific cod. While originally focused on Tanner crab (*Chionoecetes bairdi*), since 1998 sampling of Pacific cod has generally been at 100%, though sub-sampling sometimes is used. Length compositions are available, and an index is derived annually and is available for inclusion in the stock assessment. The survey covers an extensive area using multiple vessels, with the large majority of tows in the Kodiak Island area. Sampling starts in early June northeast of Kodiak Island, continuing in the area until mid-July before moving west to the Aleutian Islands. The western side of the Kodiak Islands is not sampled until late August through September. It is unclear how this design potentially impacts population sampling and any interaction with annual or seasonal variations in growth or distribution.

The derived index, shown in the SAFE report (Fig 2.42), displays a similar trend from 1998 onwards as the AFSC trawl survey (Fig 2.30 of the same report). A priori, with the trends in the AFSC and ADFG westward survey matching, the only benefit of adding the survey information is if there is information in the annual length compositions.

Length frequencies from the survey, compared to the AFSC trawl, are shown below (note non-AFSC trawl survey years are not shown and scaling varies by year and survey). The ADFG survey appears to sample larger fish in many years and generally, but not always, does not select the smaller fish evident in the AFSC surveys (e.g. in 2005, 2007, 2009, 2015).



Attempts to fit models using the ADFG westward data (index and length comps) were discussed during the review. My understanding is that model runs with no age data (not shown in detail) were not possible with the ADFG included with or without other survey length compositions. This perhaps indicates a lack of compatibility between AFSC and ADFG survey compositions. Further, model runs without age but using only ADFG length compositions had poorly defined likelihood surfaces and estimates of natural mortality were pushed to the upper bound of the prior being used. The implication is that the ADFG length compositions, reflecting the apparent sampling of larger cf smaller fish, contain little information on natural mortality.

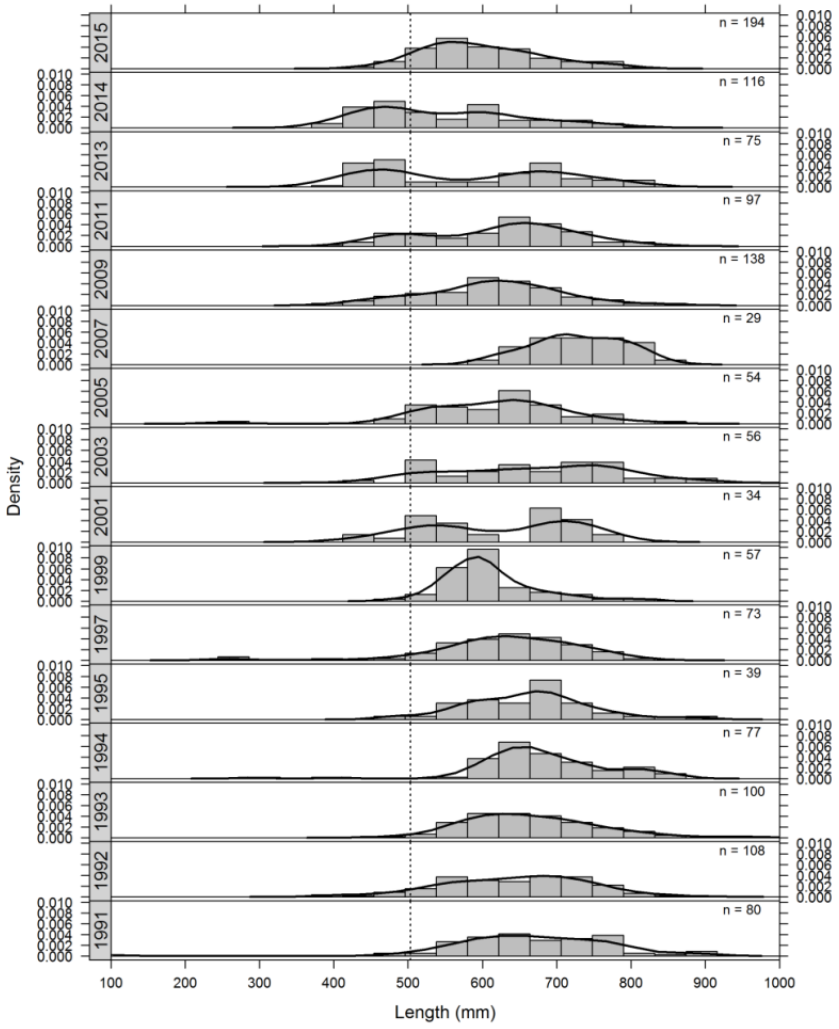
Runs using ADFG and other survey data were able to be fit using age composition data. Model runs made during the review (Model18.09.42_biasSTATE and Model18.09.41_bias) both include ageing error and a bias correction in ageing pre-2007 due to changes in age reading protocols. The bias correction is itself exploratory and needs further work, as does the method of dealing with it in the stock assessment (it is currently done by a simple external bias correction rather than using any fitting procedure). However, these runs are the only ones available to help interpret the value/impact of using the ADFG westward survey data in the assessment. The likelihoods for the models are difficult to compare and the models are also all untuned; care is needed not to over-interpret. Nevertheless, it appears the age composition data are considerably worse fit when the ADFG length compositions are used, there is conflict between survey length compositions, and the estimate of M on older fish is increased. These results are not surprising given the attempts to fit without age data noted above.

At this stage, the ADFG westward survey data seem to be inconsistent with the only available age data and possibly also with other length composition data. There is therefore no clear advantage to using the survey at this time in a stock assessment to inform management. Further exploratory analyses could prove otherwise. The disadvantage of using the survey data is that model fitting will be compromised at a time when robust advice is required.

The ADFG Central Region survey has had full accounting of Pacific cod since 1991, covering

Prince William Sound, Kachemak Bay and Kamishak Bay. Two vessels have consistently been used annually/biannually with gear and protocols standardised. Length compositions and indices are available for each area. The consideration of this survey is focused primarily on its potential to provide information on small fish and incoming recruitment. The presentation during the review showed core stations (all in State waters) and ancillary stations (in State and Federal waters). The depth ranges sampled were not clear, but appear to include shallow waters in which smaller fish might be expected. Length compositions for Prince William Sound only (see below), however, suggest selectivity (or availability or catchability) of small fish is not high. I am not aware of any model runs using the ADFG Central Region survey data and it is hard to make strong conclusions. My understanding is that the apparently low selectivity for small fish is such that there may be little utility in trying to use the index in stock assessment, because it does not add information as required and might lead to conflicting (though weak) signals between indices and length compositions. It is, however, very difficult to interpret the length compositions (below) relative to the AFSC and ADFG westward surveys (see above).

Overall, I can see no advantage in trying to fit the survey in the stock assessment at this time.



The IPHC Setline survey (https://iphc.int/management/science-and-research/fishery-independent-setline-survey-fiss/61-fiss-design-and-implementation) is an extensive survey

using approximately 1200 stations on a 10nm grid annually during late May to late August from northern California, Oregon, Washington state, through the Gulf of Alaska and Aleutian Islands and the Bering Sea. It has been operating using contracted commercial vessels (circa 14) across all areas since 1998 using fixed setline stations. It includes standard stations and, through time, various "expansion areas" to include known halibut areas not covered by the fixed design.

The IPHC is open to collaboration with other agencies, including NMFS/AFSC for sampling Pacific cod. There is an intention during 2019 to use expansion areas in areas which are relevant for Gulf of Alaska Pacific cod, though the choice of areas will be driven primarily by the sampling requirements for halibut.

Sampling already includes all species (not just halibut) from the first 20 hooks of each 100-hook line, including species caught and, in the past, Pacific cod lengths in given areas using the first 15 fish caught per setline even if from more than the first 20 hooks. As of yet, there are no data available to compare length compositions, but an index derived by the lead analyst (SAFE Fig 2.39) shows very similar trends to the AFSC and ADG westward trawl surveys and the AFSC longline survey.

Given the wide area and depth coverage of the setline survey, existing good biological sampling of Pacific cod and intended expansion, there is considerable potential to drive a highly informative annual index with associated length and potentially age compositions. However, it is apparently the case that the lengths of Pacific cod sampled in the Bering Sea trawl survey are different to those sampled in the IPHC setline survey. This may be due to the large hook size used in the setline survey being too large to sample smaller Pacific cod. Discussion during the review on this issue was inconclusive.

The key needs for the assessment are for an annual index and for information on young fish/recruitment. Until the information on length compositions becomes available for analysis (in 2018 or 2019?) the utility of the setline survey compositions to provide information on young fish is moot. However, a reliable annual index will have high value and annual compositions for length, and ideally age, should provide greater information than the current AFSC surveys if and when they are available, even if small fish are not sampled.

Model runs reported in the SAFE and considered during the review meeting (Model17.09.35 cf other Model17.09 runs) suggest the IPHC index can be well fit. However, the index is very similar to the AFSC indices, the runs considered are untuned, and there is considerable flexibility to fit indices through selectivity blocking and annual variation, and covariance with temperature. That the index is well fit is therefore neither surprising nor necessarily supportive of its inclusion in final runs. The primary rationale for inclusion is the prior expectation that the wide area and depth coverage and annual availability of the index should provide the single best index available for GOA Pacific cod relative population numbers.

Similarly, as long as Pacific cod sampling in the survey is representative and sufficient, and especially if ageing of samples is undertaken, it should provide the most representative data on population composition.

Together with work on extending the ageing database, consideration of the IPHC survey to be included in future stock assessments is a clear priority.

2. Evaluate and provide recommendations on model structure, assumptions, and estimation procedures.

In particular:

a. How would you evaluate the appropriate level of complexity in the stock assessment model given that we have historically used simple and more complex models to manage this stock?

The purpose of stock assessment is to inform management decision-making. To that end, the need is to investigate the trade-off between model complexity and management utility.

Stock assessments need to deliver consistently interpretable metrics, and ideally result in advice using those metrics that is robust to uncertainties or at least exposes how uncertainties impact advice. Advice needs not just to consider current/recent stock status, but also forecasts of how management interventions will impact future status. There are no ToR explicitly related to these management metrics (e.g., estimates of reference points, OFL) and projections, and the review focused more on the feasibility of estimation rather than on management utility.

The "appropriate level of complexity" is also dependent on the management regime in place. For the North Pacific, the appropriate level of complexity needs to take account of how the SSC advises and the NPFMC operates, the technical requirements of the Tier system in use, etc. However, it also needs to take account of the limitations of the models in use to interpret data, and the types and quality of data available.

In my (hackneyed) view, management is better served by roughly right than by precisely wrong assessment.

Specific, potential complexities are considered in the following ToR on selectivity blocking and the dependency of natural mortality on temperature. The single other issue of structural complexity that arose during the review (and prior) is stock distributional change/migration.

Pacific cod has quite clearly fluctuated historically and the recent large decline in availability to surveys and fishing coincides with a major environmental signal ("The blob"). Interpretation of the decline in availability as a population decline is the simplest from a stock assessment and immediate management perspective. An alternative suggestion is that the decline is entirely or partially due to movement or behavioural change associated with the environmental change. If such a change were short-term, then fish would be expected to reappear, and composition data should confirm the fate of cohorts. However, for a relatively short-lived species and an extended change period, especially with alternative fishing options and biannual surveys, confirmation might be confounded. Also, a major distributional change would be expected in commercial fisheries – as I understand it, no such change is clearly evident (though I note during review commercial distributions and other fishery bycatch were discussed).

From a modelling perspective, availability is reflected either as a change in mortality or as a change in selectivity, given there are no migration or other spatial aspects to the model, nor

any data to drive such modelling. Natural mortality is poorly defined given limited age compositions and ageing difficulties, though most model fits do suggest it is of the order of 0.4-0.45, consistent with the maximum age estimated using revised ageing protocols. Exploration of natural mortality related to environmental change is covered at ToR 3.

Changes in selectivity is the alternative mechanism for exploration, but is always problematic given multiple gear types, changes in fishery regulations and economic drivers, changes in survey methodology/gear, etc. Selectivity blocking is considered at ToR 2c.

Ultimately, the level of complexity of the stock assessment model must be driven by data, model fit, and management utility. In this case, there are good commercial catch data for the three gear types (trawl, longline, pots), consistent fishery independent trend signals from AFSC longline and trawl surveys (and from the IPHC setline survey and ADFG Western Region trawl survey), length composition data (annual from commercial fisheries and the AFSC longline survey, tri/biennial from the AFSC trawl), and age information as compositions and conditional at-length compositions from just the tri/biennial trawl surveys. Ageing protocols have been changed and there is a need to revise the age data pre-2007 or find a way to deal with the change within the model. Ageing error has not been quantified, but is assumed. Overall, while the signals of decline are clear and consistent, the data simply do not currently exist to drive a complex model that can clearly disentangle the cause of change. Like most stock assessments, the model results may reasonably show the pattern of historical fluctuation, but cannot definitively explain how and why those fluctuations came about or will progress.

My recommendation is to keep it simple to inform management. Use a fixed natural mortality but explore estimation thereof to check for robustness. Only change selectivity if there is either a compelling prior reason or if close scrutiny of, for example, the likelihoods by year for the age or length components of a gear/fleet show systematic patterns that match a known driver. If selectivity is blocked (or forced to be domed, etc.) then again, explore for robustness of i) status metrics, and ii) projection implications of alternative management actions. If selectivity is allowed to vary annually, consider very carefully if this compromises the ability to forecast at all.

A final caution. The model comparisons presented during review (Model 18 series) have a number of variations in data sets (e.g., surveys included or not, all age compositions cf no pre-2007 age compositions, etc.). Comparisons cannot directly use NLL or AIC as a measure of fit and only by looking at the detailed fits within models is it possible potentially to interpret the most appropriate level of complexity (see also Tor 2c). In the same vein, comparisons between tuned (e.g. the Model 16.10 series) and cf untuned (Model 16.0-16.08, 17 and 18 series) models need to be treated cautiously. (I note on re-reading that the SAFE report provides a similar caution.)

b. What factors should be considered in data weighting and how should we assess the appropriateness of current methods applied for this stock?

The paper by Francis (2011) covers elegantly the principles of weighting with strong arguments as to why his proposed method gives priority to abundance data over composition data. The argument essentially is that abundance is what matters most in terms of status and management advice. This is true in general, but if models are fit which include variable natural mortality and selectivity, then composition data become more

important both in the fitting of models and in projecting. It may, therefore, be appropriate to use abundance-oriented (Francis-type) weighting for some models while using more composition-oriented weighting (e.g. Mcallister and Ianelli) for others. As ever, it depends on how the model outputs are to be used. That is, there is an interaction with the type of complexity introduced and whether the focus is on status or forecasting.

Weighting is also impacted by other matters. For example, in high F fisheries with extensive and long-term commercial compositional data, there is a lot of information on mortality and even abundance. In such fisheries (e.g., on many Northeast Atlantic stocks), surveys may be relatively unimportant in estimating abundance and status but may be critical in defining incoming recruitment. In contrast, low F fisheries with poorly sampled or otherwise challenged compositional data, may rely largely on abundance signals from fishery-independent surveys. The form of weighting, if any, therefore in practice relies on many considerations.

Where weighting is used then multi-stage weighting is generally appropriate to account for first observation and then, iteratively, process error. The SAFE report clearly outlines the weighting approaches taken and I see no problems with the assumed initial multinomial sample sizes or explanation as to why only a single-stage weighting was used.

Limited lessons for the current assessment may be found by looking at specific runs, with and without weighting. Available runs are limited but Models 16.10.11, 16.10.20, 16.10.23, and 16.10.25 have unweighted counterparts, as does Model 17.09.36. I have not found any runs with weighting cf non-weighting for models with constant or estimated M but with fixed selectivities.

Comparing 17.09.36 and 17.09.35, in which all selectivities are allowed to be dome shaped, there is a trawl and longline fishery selectivity block for 2005-2006, there is a 2015-16 block on M, selectivity for fisheries can vary annually, and there is a covariate on q with temperature. In this case, the Francis weighting makes a notable difference on NLL, but very little difference at all to management-related quantities. What is not clear, is how the improvements in NLL, largely gained in the fits to length compositions (due to the multinomial single pass weighting), translate to population structure and impact on projections. However, as the model already includes maximum annual and time block variability and a covariate, it is in any case difficult to know how projections would be made.

The interaction between complexity and weighting can be seen comparing i) Model 16.10.11 with 16.08.11, and ii) 6.10.23 with 16.08.23. The 10 series models include Francis weighting, while the 08 series are unweighted. In (i) all fishery selectivities are allowed to be dome shaped and vary annually. In (ii), all selectivities are blocked rather than varying annually. There is little difference in the management-related quantity estimates between (i) and (ii) or between weighted and unweighted runs. As above, however, it is unclear how the detailed fits and assumptions would be carried through into projections.

Overall, weighting needs to be considered alongside complexity. If a very simple model is used, then an unweighted run will balance the influence of composition and abundance data. The abundance signals are strong and consistent, and unless there are strong counter-signals form the composition data, weighting may have little impact. It is not clear in this case that the composition data are highly informative. Results in fact seem to be determined primarily by whether or not selectivity is assumed asymptotic or allowed to be domed. If assumed asymptotic, status will be assessed higher. So long as selectivity is

allowed to be domed, all model runs seem to result in similar status results regardless of the complexity of M and selectivity estimation and of weighting--at least as seen in the limited exploration. As noted above, the degree of complexity used to advise will need to consider how any fitted model can be used to project. Fitting annually varying domed selectivity and estimating M, perhaps with blocks or a covariate, seem somewhat over-demanding of the limited data, and will make assumptions for projections debatable, but the more complexity introduced by annual variation and blocking, the less likely weighting is to have any impact.

c. How can we evaluate the appropriate level of time variability and appropriate pattern (i.e. blocking vs random walk) in fishery and survey selectivity patterns?

The true selectivity of a fishery or survey is a property of underlying availability of fish which varies in space and time, and to fishing operations that also vary in space and time, interacting with the vulnerability of fish by age or length to the gear and how it is operated. In modelling, the need is to capture as best possible the average selectivity over appropriate dimensions. Given models do not typically include the detailed spatio-temporal dynamics of fish and fleets, selectivity is aliased with other processes, and potentially relevant composition data are typically limited, estimation of selectivity is both contentious and problematic. Worse, how selectivity is modelled affects determination of management-related quantities such as %SPR and impacts on projections of management impacts.

Evaluating the appropriate level of time variability and pattern therefore interacts with ToR 2a and 2b. The balance of any final model or suite of models used to inform management will need to take account of data availability, whether confounding processes are allowed to vary, data weighting, etc. There is no one right answer and analyst judgment cannot be avoided.

For all commercial fleets in the Pacific cod assessment, age data are currently unavailable, and selectivity can only be modelled based on size. The appropriate pattern (asymptotic or domed) will of course depend on the gear type and can be informed by a priori considerations. Trawl selectivity may be asymptotic, but for a variety of fishery and species-specific reasons may be domed. Longline selectivity may similarly display either pattern depending on hook set and depth, etc. Pot selectivity is typically domed. For the longline and trawl surveys, the same underlying considerations apply. The specific form may be informed by commercial size compositions cf other commercial data, and by survey compositions, as well as by comparison to other fisheries. The form and potential changes through time may be informed by management interventions and operational changes driven by regulatory, economic, or other commercial drivers.

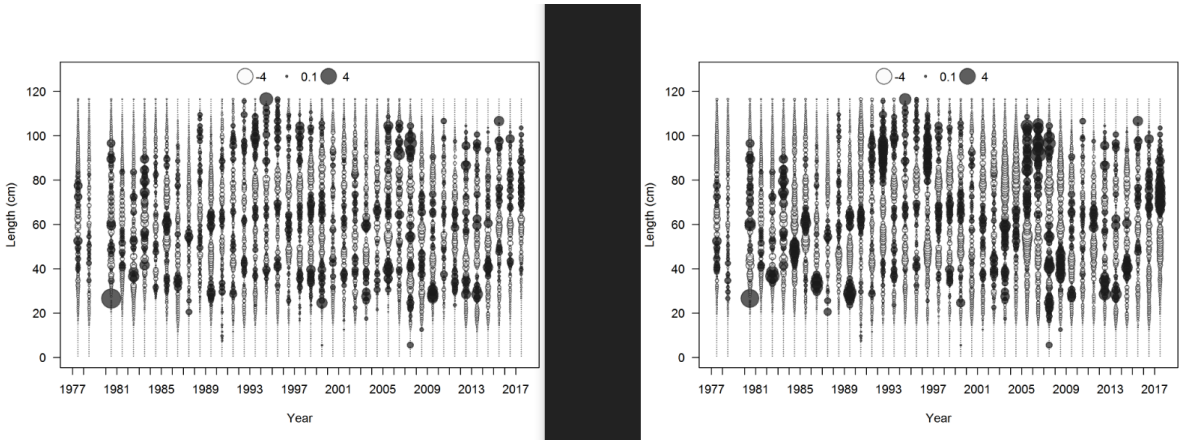
Deciding on adoption of appropriate selectivity patterns and blocks in models used for advice needs to take account of these prior considerations, but also needs to be based on model fitting. For selectivity, the key thing is to look in detail at the residual patterns for age or size by fleet through time, and perhaps at likelihood components on the same scale (not just on total length or age compositions). Always, however, it is not just a matter of getting the best fit, care is needed to consider how final blocking may affect reference point definition and ability to project. There is little utility to management in fitting annually variable selectivity for a gear, and then not being able to project forward due to lack of knowledge about future selectivity. Over-fitting will also likely mask the true process and estimation error, and there is probably better utility to management in more averaging and

allowing correspondingly higher estimation of confidence intervals on reference points and projections. We come again, therefore, to the need to investigate the trade-off between model complexity and management utility.

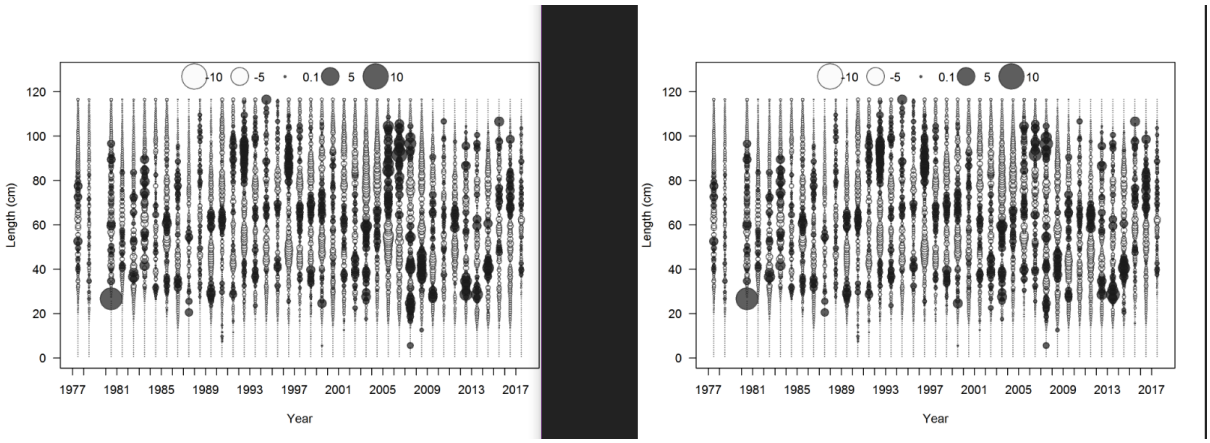
It is difficult to comment in detail. However, I have tried to compare two model sets to aid understanding. First, Model 16.08.23 cf Model 16.08.11, in which the difference is the addition of blocked fishery and survey selectivity as opposed to annually varying selectivity. Second, Model 17.09.35 cf Model 17.09.31, in which the only difference is the added selectivity block for trawl and longline for 2005-06. I do not suggest any of these models are appropriate final choices, but use them only to look at changes in gross likelihood components and residuals.

For the first model comparison, noting these are not models that were fully developed or proposed for adoption, the AIC decreases because while there is a decrease in parameter number by 217, the NLL increase is 245. Length composition data are, of course, fit worse, but all other data are fit better (as judged by the likelihood component values). Status determination is little affected. Residuals (below) on length are, of course, worse with some strong patterns (e.g. at 2005-06).

In this case, I would ignore the worse length composition residuals (though might be tempted to at least look at 2005-06) and suggest adopting the blocked selectivity as being i) more parsimonious, and ii) of greater utility as reference point definition and projecting are more straightforward.



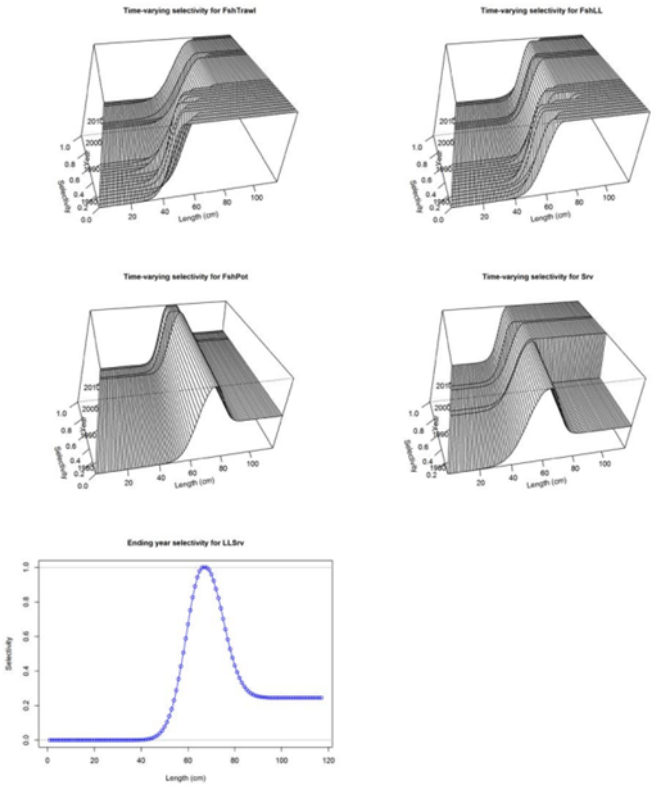
For the second model comparison, Model 17.09.35 has a lower AIC, balancing an increase in parameter number of 7 against a NLL reduction of circa 35 units. All likelihood components increase except, unsurprisingly, that for length compositions, which of course are better fit due to the selectivity (at length) block introduced to reflect a change in the B season fishery in those years. Looking at the residuals just for the trawl fishery (below: LH panel Model 17.09.31; RH panel Model 17.09.35) there is a clear improvement in residuals for the years in question; in the no-block model there is a very strong negative-positive pattern on smaller-larger fish which disappears in the blocked model.



Is it then appropriate to add the block? The two models result in almost identical management-related outputs, including status determination. Fits on the restricted years for the length compositions are improved, but all other fits degrade very slightly. There is no clear advantage to including the block and its inclusion i) adds to complexity while providing no management utility; ii) slightly degrades fits to other data for no good reason; and iii) adds a degree of complexity which might interact with other changes to fit/weighting, all of which would then need to be explored. In this case, therefore, despite an a priori reason for a block and improved fit to data, given there is no utility, I would not persevere with the 2005-06 block in a final model used for providing advice. Of course, if the block were in the "blob" period or influenced status determination, that conclusion might change.

The ToR question remains: what is the appropriate level of time variability? The foregoing confirms the analyst explorations as being helpful in defining some degree of blocking.

Results of blocked and time varying selectivity runs are represented by the output from Model 17.09.35 (below) showing fitted selectivity by fleet (commercial x3 and survey x2). The commercial blocks are based on provided, known fishery changes, with Longline selectivity blocks 1978-1990 (annual devs), 1991-2004, 2005-2006, 2006-2016, 2017, Trawl selectivity blocks 1977-1990 (annual devs), 1991-2004, 2005-2006, 2006-2016, 2017, and Pot selectivity blocks 1990 - 2013, 2013-2017. The survey blocks are based on changes of e.g. tow duration and net mensuration.



My inclination is: I) Not fit annual varying selectivity for early periods. It adds complexity and will reduce estimated error in early population size. Fitting these as single blocks with error seems preferable. II) Given the history of the fisheries and clear indications of change in selectivity, a block from 1991 to 2016 for trawl and longline fisheries is appropriate. However, noting the difference in fitted selectivity for trawl and longline pre-2005 and post-2006 (when a 2005-06 block is included), it is unclear if a single block is best and exploring the fits and management implications is still needed. III) For pots, the change in sampling in 2013 suggest a possible need for a new block, but estimation must be confounded by the overlap in time of the "blob". It may be simpler to use just the single block from 1990 onwards. IV) It is very unclear how to deal with 2017 until further information is available to the stock assessment. I am unclear as to the intention, but understood from the review that pot selectivity would be fit for 2013-17, while for trawl and longline there is an intention to fit 2017 selectivity separately. With changes in ABC and fishing operations, there is likely to be a real change in selectivity. This may not impact status determination, but any projections will need to take account of potential changes in practice. It is unclear how this would be done.

3. Evaluate how ecosystem indicators are used in the assessment and provide recommendations how they can be better integrated into model development and stock management.

a. Should environmental indices be used to model natural mortality in the model? Is it appropriate to use a time block for the extremely warm period to adjust natural mortality?

As alluded to in ToR 2a, fish stock assessment models need both to explain the past in

simplified terms and provide predictive power for the future. The greater the complexity of models used to interpret the past, in general, the greater the difficulty for prediction.

My strong view is that advice and management can and should always take account of information such as changes in environmental conditions, but these may not typically be best included in the models. I note the standard approach in the North Pacific to providing an ecosystem chapter with multiple standard indices and additional information. Data collection and research on North Pacific environmental and ecosystem changes is arguably the best in the world. Nevertheless, consideration for management decision-making is handled as a separate (integrated) step. Changes in indices can often be ascribed to different reasons through time, and interpretation can be diametrically opposed in some cases.

The issue of using environmental information to inform recruitment was considered by Francis 2006; https://academic.oup.com/icesjms/article/63/4/594/691541). The conclusions extend to this ToR. The abstract of the paper is as follows:

There has recently been considerable interest in establishing relationships between environmental variables and annual recruitment to fish stocks. Such relationships have the potential to reduce the uncertainty in the assessment of the stocks. When many environmental variables are considered, it is easy to draw conclusions that exaggerate the ability to predict recruitment. One technique to protect against this is cross-validation. This technique has usually been incorrectly applied, in that it has not included predictor screening (the selection from a large set of potential predictors of a smaller set to use in prediction). A simulation experiment is used to show that this omission can cause chance correlations to be wrongly identified as useful, and the reliability of useful predictors to be overestimated. It also shows that the mistaken use of chance correlations to predict recruitment can be worse than the use of the default predictor (the mean of previous recruitments), and that our ability to measure the reliability of recruitment predictors is typically poor.

Natural mortality is in truth variable by year across ages due to changes in growth patterns, predator fields, food availability, all of which may have complex density-dependent bases as well as environmental ones. It may well be possible to fit a model that relates natural mortality to e.g. temperature or cumulative temperature in given years, but there is no guarantee at all that the fitted relationship will hold and have predictive power – it may be the underlying process changes or that the fit is unreliable. The fundamental problem is that while we can create pictures of the past, we can only predict if we understand the processes that will give rise to the future.

The ToR asks specifically if it is appropriate to use a time block for the recent warm period to adjust natural mortality. Natural mortality is estimated either externally or within the model using a prior. Estimation is informed by composition data (either age directly or length via growth). Based on model fitting reviewed, the composition data contain little information on natural mortality. Information on small/young fish is a general weakness of the model and information on natural mortality for medium to large fish is contained in few cohorts, smeared in the data as time progresses. Any chance of estimating natural mortality over the recent 3-4 years block would rely on few data, and I see little realistic chance of reliably estimating it until more data have accrued. Again, assessments develop pictures of the past and only passing time allows reliable estimation. In any case, a separate, recent natural mortality estimate would further confuse potential reference point definition and create a

quandary for defining projections.

My strong inclination is towards a simpler model, including a single natural mortality block, and the use of averaged information being used to define reference points and projections. Averaging across changeable periods should capture the large uncertainty associated with any reference points, status determination, and projections. The more complexity is introduced, the more this real uncertainty might be under-estimated. Again, to be repetitive, management is best served by robust advice and exposing uncertainty. While we can fit many things, roughly right models generally have more utility than precisely wrong ones.

b. Is the temperature-catchability relationship modeled for AFSC surveys being modeled appropriately?

(NB There seems to be an error in the summary spreadsheet provided post review in that the total likelihoods and AIC for Models 17.09.26 and 17.09.31 are slightly wrong. It makes no difference to interpretation, but I have not checked if the error carries through elsewhere.)

I see no problem with the modelling as such. It is well described in the SAFE, allowing catchability of the AFSC longline survey in the stock assessment to depend also on the CFSR for 10cm fish. (The SAFE description on polynomial fitting seems a red herring.) What is not clear is why, ultimately just the CFSR for the 10cm fish is used. I cannot find a clear rationale for this or testing of the implications of alternatives. Nor is it necessarily the case that catchability across the size range would be similarly impacted. The correlation between 10cm RPN and CFSR is only 0.30, suggesting a weak relationship – while an improvement in model fit is expected, it is not clear there will be any gain in predictive power.

The model fit is only marginally better overall, with the major gain in improvement to the survey indices (of course) and slightly on the length compositions. It is unclear if the fits to the trawl survey compositions are affected, but the survey age composition fits appear little changed suggesting all improvements are in the longline survey fits. The improvement in survey fit is entirely at the expense of alignment with the priors. The model fits better, as expected. However, it's inclusion has come after other model complexities which impact on interpretation of the data and has also not been explored with alternative tuning.

Appendix 1

Materials provided in advance using Google Drive

- ComFish_Gulf of Alaska PCOD_notes
- ecosysGOA_2017
- GOA_Cod_PlanTeam2016
- GOAPCOD_2016
- GOAPCOD_2017
- GOAPCOD_SEPT2017_PT
- T01x-05 SoW GOA Pac Cod

Presentations (most provided in advance)

1. Barbeaux. CIE Gulf of Alaska PACIFIC COD (Background, data, stock assessment)
2. Zador. GOAcodCIE_2018_Zador (The GOA's endless summer)
3. Hanselman. The AFSC Longline survey
4. Palsson. The AFSC Bottom Trawl survey
5. Spalinger. The ADFG westward region large-mesh bottom trawl survey
6. Byerly. The ADFG central region large-mesh bottom trawl survey
7. Concepcion. The North Pacific observer program
8. Pacific cod Catch Estimation and Management in Federal Groundfish Fisheries of the Gulf of Alaska – 2018
9. Hicks. The IPHC setline survey.

Other Reference materials provided using Google Drive

- 620_Final Report2007
- article about movement of P cod -- 19425120.2014.976680
- British_columbia_pacific_cod
- Canada P cod stocks 331687
- Charnov Gislason Pope 2013 Evolutionary assemblage rules for fish life histories
- DutilandLabert1998_M_poorcondition
- Francis2011_weighting
- Laureletal2008_larvae
- Laureletal2010_Larvae
- McAlister and Ianelli1998_Weighting
- RoseandKulka1998
- shimada_TAGGING
- Thorson2014_weighting

Appendix 2

Statement of Work

External Independent Peer Review by the Center for Independent Experts

Assessment of the Pacific cod stocks in the Gulf of Alaska

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[1]. Further information on the Center for Independent Experts (CIE) program may be obtained from www.ciereviews.org.

Scope

The Gulf of Alaska Pacific cod stock assessment has had a large number of alternative models over the years. In 2016, the model was rebuilt from scratch and greatly reduced in complexity from the previous model. Of particular concern is that this stock has experienced a precipitous decline since 2015 and there is concern that the simpler model may not adequately address the important biological complexities to appropriately manage this stock in the face of climate variability. However, review is requested of all aspects of the stock assessment models. The Pacific cod fisheries in the Gulf of Alaska is of great economic importance garnering $103 million ex-vessel value annually (29% of all Gulf of Alaska groundfish fisheries). The individual review reports are to be formatted with content requirements as specified in **Annex 1**. The Terms of Reference (ToRs) of the peer review are attached in **Annex 2**. The tentative agenda of the panel review meeting is attached in **Annex 3**.

[1] http://www.cio.noaa.gov/services_programs/pdfs/OMB_Peer_Review_Bulletin_m05-03.pdf

Requirements
Three CIE reviewers shall conduct an impartial and independent peer review in accordance with the SoW and ToRs herein. CIE reviewers shall have working knowledge and recent experience in the application of stock assessment methods in general, and in Stock Synthesis in particular.

Tasks for Reviewers
Each CIE reviewers shall complete the following tasks in accordance with the SoW and Schedule of Milestones and Deliverables herein.

Pre-review Background Documents: Two weeks before the peer review, the NMFS Project Contact will send (by electronic mail or make available at an FTP site) to the CIE reviewers the necessary background information and reports for the peer review. CIE reviewers are responsible only for the pre-review documents that are delivered to the reviewer in accordance to the SoW scheduled deadlines specified herein. The CIE reviewers shall read all documents in preparation for this peer review.

2016 Assessment of the Pacific cod stock in the Gulf of Alaska (150 p.)
2017 Assessment of the Pacific cod stock in the Gulf of Alaska (144 p.)
2017 Ecosystem Considerations Status of the Gulf of Alaska Marine Ecosystem (215 p.)
Comments on the final 2016 and 2017 Gulf of Alaska (GOA) Pacific cod assessments by the Plan Team and Scientific and Statistical Committee (*SSC*)

Panel Review Meeting: Each CIE reviewer shall conduct the independent peer review in accordance with this SoW and ToRs, and shall not serve in any other role unless specified herein. Each CIE reviewer shall actively participate in a professional and respectful manner as a member of the meeting review panel, and their peer review tasks shall be focused on the ToRs as specified herein. The NMFS Project Contact is responsible for any facility arrangements (e.g., conference room for panel review meetings or teleconference arrangements). The NMFS Project Contact is responsible for ensuring that the Chair understands the contractual role of the CIE reviewers as specified herein.

This review meeting will include three main parts: The first will consist of a series of presentations with follow-up questions and discussions by CIE reviewers, and will be chaired by an AFSC scientist or supervisor. The second will consist of real-time model runs and evaluations conducted in an informal workshop setting, and will be chaired jointly by the CIE reviewers. The third, time permitting, will consist of initial report writing by the CIE reviewers, with opportunity for additional questions of the assessment author.

Contract Deliverables - Independent CIE Peer Review Reports: Each CIE reviewer shall complete an independent peer review report in accordance with the SoW. Each CIE reviewer shall complete the independent peer review according to required format and content as described in **Annex 1**. Each CIE reviewer shall complete the independent peer review addressing each ToR as described in **Annex 2**.

Other Tasks – Contribution to Summary Report: Each CIE reviewer may assist the Chair of the panel review meeting with contributions to the Summary Report, based on the terms of reference of the review. Each CIE reviewer is not required to reach a consensus, and should provide a brief summary of the reviewer's views on the summary of findings and conclusions reached by the review panel in accordance with the ToRs.

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

http://deemedexports.noaa.gov/compliance_access_control_procedures/noaa-foreign-national-registration-system.html

Specific Tasks for CIE Reviewers: The following chronological list of tasks shall be completed by each CIE reviewer in a timely manner as specified in the **Schedule of Milestones and Deliverables**.

1) Conduct necessary pre-review preparations, including the review of background material and reports provided by the NMFS Project Contact in advance of the peer review.
2) Participate during the panel review meeting scheduled in Seattle, WA during May 1 - 4, 2018.
3) Approximately three weeks after the conclusion of the panel review meeting, each CIE reviewer shall submit an independent peer review report addressed to the CIE. Each CIE report shall be written using the format and content requirements specified in **Annex 1**, and address each ToR in **Annex 2**.

Place of Performance

The place of performance shall be at the contractor's facilities, and Seattle, Washington.

Period of Performance

The period of performance shall be from the time of award through June 2018. Each reviewer's duties shall not exceed 14 days to complete all required tasks.

Schedule of Milestones and Deliverables: CIE shall complete the tasks and deliverables described in this SoW in accordance with the following schedule.

March 26, 2018	CIE selects and confirms reviewers. Reviewer contact information is sent to the NMFS Project Contact
April 16, 2018	NMFS Project Contact sends the reviewers the pre-review documents
May 1 - 4, 2018	Each reviewer participates and conducts an independent peer review during the panel review meeting
Approximately three weeks later	CIE receives draft reports
Approximately two weeks later	CIE 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 (3) The reports shall be delivered as specified in the schedule of milestones and deliverables.

Travel

All travel expenses shall be reimbursable in accordance with Federal Travel Regulations (http://www.gsa.gov/portal/content/104790). International travel is authorized for this contract. Travel is not to exceed $12,000.

Restricted or Limited Use of Data

The contractors may be required to sign and adhere to a non-disclosure agreement.

NMFS Project Contact:

Steven J. Barbeaux, Alaska Fisheries Science Center

7600 Sand Point Way NE

Seattle, WA 98115

Phone: 206-526-4211

Steve.Barbeaux@noaa.gov

Annex 1: Format and Contents of CIE Independent Peer Review Report

1. The CIE independent report shall 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 main body of the reviewer report shall consist of a Background, Description of the Individual Reviewer's Role 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 should describe in their own words the review activities completed during the panel review meeting, including providing 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, and especially where there were divergent views.

   c. Reviewers should elaborate on any points raised in the Summary Report that they feel 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 CIE independent 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 CIE independent report shall be an independent peer review of each ToRs, and shall not simply repeat the contents of the summary report.

3. The reviewer report shall include the following appendices:

   Appendix 1: Bibliography of materials provided for review
   Appendix 2: A copy of the CIE Statement of Work
   Appendix 3: Panel Membership or other pertinent information from the panel review meeting.

Annex 2: Terms of Reference for the Peer Review

Assessment of the Pacific cod stocks in the Gulf of Alaska

1. Evaluate and provide recommendations on data used in the assessment models. In particular:
    a. What are the benefits vs disadvantages of including data from the ADFG small-mesh trawl and the IPHC longline surveys in the assessment?
2. Evaluate and provide recommendations on model structure, assumptions, and estimation procedures. In particular:
    a. How would you evaluate the appropriate level of complexity in the stock assessment model given that we have historically used simple and more complex models to manage this stock?
    b. What factors should be considered in data weighting and how should we assess the appropriateness of current methods applied for this stock?
    c. How can we evaluate the appropriate level of time variability and appropriate pattern (i.e. blocking vs random walk) in fishery and survey selectivity patterns?
3. Evaluate how ecosystem indicators are used in the assessment and provide recommendations how they can be better integrated into model development and stock management.
    a. Should environmental indices be used to model natural mortality in the model? Is it appropriate to use a time block for the extremely warm period to adjust natural mortality?
    b. Is the temperature-catchability relationship modeled for AFSC surveys being modeled appropriately?

Annex 3: Tentative Agenda

CIE Review of the GOA Pacific cod stock assessment models

Alaska Fisheries Science Center

7600 Sand Point Way NE, Seattle, WA 98115

May 1 - 4, 2018

Building 4; Room 2039

Review panel chair: Grant Thompson, Grant.Thompson@noaa.gov

Senior assessment author: Steven J Barbeaux, Steve.Barbeaux@noaa.gov

Security and check-in: Sandra Lowe, Sandra.Lowe@noaa.gov (206)526-4230

Sessions will run from 9 a.m. to 5 p.m. each day, with time for lunch and morning and afternoon breaks. Discussion will be open to everyone, with priority given to the panel and senior assessment author.

Tuesday, May 1

Preliminaries:

09:00 Introductions and adoption of agenda—Grant Thompson

Data sources (current and potential):

09:10 Overview of data types used in the assessments—Steve

09:20 Catch accounting system and in-season management—AKRO SF Division (via WebEx)

09:50 Observer program—AFSC FMA Division

10:20 Break
10:30 GOA trawl survey—AFSC RACE Division

11:00 AFSC longline survey—AFSC Auke Bay Laboratory (via WebEx)

11:30 IPHC longline survey—IPHC

12:00 Lunch

13:00 ADFG surveys— ADFG (via WebEx)

13:30 GOA Ecosystem assessment—AFSC REFM – Stephani Zador

Assessment models:

14:00 Assessment history—Steve

15:00 Break

15:10 Current assessments—Steve

16:10 Discussion—Everyone

16:40 Assignments for models to be presented on Wednesday—Panel

Wednesday, May 2 and Thursday, May 3

Review of models assigned the previous day—Steve

Discussion, real-time model runs—Everyone

Assignments for models to be presented the following day—Panel

Friday, May 4

Review of models assigned on Thursday—Steve

Discussion, real-time model runs—Everyone

Report writing (time permitting)—Panel

Appendix 3

List of Participants provided by AFSC

List of presenters:

1. Steve Barbeaux (AFSC)
2. Mary Furuness (NMFS Alaska Region)
3. Marlon Concepcion (AFSC)
4. Wayne Palsson (AFSC)
5. Dana Hanselman (AFSC)
6. Allan Hicks (IPHC)
7. Kally Spalinger (ADFG)
8. Mike Byerly (ADFG)
9. Stephani Zador (AFSC)

List of CIE reviewers:

1. Jean-Jacques Maguire
2. Henrik Sparholt
3. Kevin Stokes

List of other in-person participants

1. Delsa Anderl (AFSC)
2. Jim Armstrong (North Pacific Fishery Management Council)
3. Craig Castelle (AFSC)
4. Anne Hollowed (AFSC)
5. Jim Ianelli (AFSC)
6. Sandi Neidetcher (AFSC)
7. Chad See (Freezer Longline Coalition)
8. Grant Thompson (AFSC)
9. Tom Wilderbuer (AFSC)

Final Agenda

Tuesday, May 1

Preliminaries:

09:00 Introductions and adoption of agenda—Grant Thompson

Data sources (current and potential):

09:10 Overview of data types used in the assessments—Steve

09:20 Catch accounting system and in-season management—AKRO SF Division (via WebEx)

09:50 Observer program—AFSC FMA Division

10:20 Break

10:30 GOA trawl survey—AFSC RACE Division

11:00 AFSC longline survey—AFSC Auke Bay Laboratory (via WebEx)

11:30 IPHC longline survey—IPHC

12:00 Lunch

13:00 ADFG surveys— ADFG (via WebEx)

13:30 GOA Ecosystem assessment—AFSC REFM – Stephani Zador

Assessment models:

14:00 Assessment history—Steve

15:00 Break

15:10 Current assessments—Steve

16:10 Discussion—Everyone

16:40 Assignments for models to be presented on Wednesday—Panel

Wednesday, May 2 and Thursday, May 3

Review of models assigned the previous day—Steve

Discussion, real-time model runs—Everyone

Assignments for models to be presented the following day—Panel

Friday, May 4

Review of models assigned on Thursday—Steve

Discussion, real-time model runs—Everyone

Report writing (time permitting)—Panel