Center for Independent Experts Independent Peer Review Report of Gulf of Alaska rex, dover, and flathead sole Assessments

Prepared for:
Center for Independent Experts
Northern Taiga Ventures, Inc. (NTVI)

By:
M. Kurtis Trzcinski, PhD
University of British Columbia
Department of Forest & Conservation Sciences
3004 - 2424 Main Mall
Vancouver, BC V6T 1Z4

Executive Summary

This report is an independent review of Gulf of Alaska rex, dover, and flathead sole assessments conducted for the Center for Independent Experts (CIE). My review including the documents and presentations that were submitted for review prior to and during the meeting held from April 29 to May 3, 2019, in Seattle, Washington, at the Alaska Fisheries Science Center.

I have a reasonable degree of confidence in the model, with the available data, to produce reliable estimates of stock status. My degree in confidence is based on several factors: 1) there are a large amount of high-quality data being collected with good scientific rigor (fisheries data, survey data, fish, lengths, ages) for these three flatfish species, 2) the model is a standard separable catch at age model, and therefore contains reasonable and commonly held assumptions for fisheries models, 3) the model fits the data reasonably well for all three flatfish species. I conclude that the science reviewed meets a high standard, incorporates the best scientific information available, and that the assessment team made considerable effort to make the best use of the data available. In my opinion, the results have provided a sound basis for management advice. I recommend that future assessments be improved by dropping the 1980s survey data, and estimating survey catchability (q).

My comments in this report focus on ways to i) more thoroughly understand the survey and fishery data entering the model, ii) further explore model sensitivity and uncertainty ranging from observation error, to process and model error, and iii) improve future research by exploring the data further using statistical models. This last point is best achieved by stating research priorities and discussing a medium to long-term vision for the assessment program.

Table of Contents

1.0 Background

This document contains my independent review of Gulf of Alaska rex, dover, and flathead sole assessments. A Center for Independent Experts (CIE) review meeting was held April 29-May 3, 2019, in Seattle, Washington, at the Alaska Fisheries Science Center (AFSC). Prior to the meeting, the review committee was provided with a Statement of Work (Appendix 2), including the Terms of Reference (ToR), background material, and assessment documents. Presentations and additional information were provided via Google Drive during the meeting. There was a general consensus among the CIE reviewers that the assessment was done with a high level of professionalism and care. As noted in my comments on the review process, the meeting was more casual in nature, which had the positive effect of creating a collegial environment for discussion not often found in working groups. This is in stark contrast to the way some CIE reviews of National Marine Fisheries Service (NMFS) assessments are conducted in other regions. The Terms of Reference (ToR) are for my report and its content and not for the assessment scientists, which is sometimes the case in other reviews. It is important to note that there was some discussion of the Terms of Reference. In particular, the ToR tasked us to review the "model," given the available data. We did not limit our review to the structure of the model *per se*, but rather, to the entire assessment process which includes i) an evaluation of the data input into the model, ii) the appropriateness of the assumptions made by the assessment scientist and structure of the model, and iii) the quality of the fit of the data to the model. In the sections that follow, I outline my positive and negative impressions and critique the science to date. Where possible, I offer suggestions and areas of future thought and research.

2.0 Individual Reviewer Activities

Prior to the meeting, I reviewed the assessment and background documents provided for review. All three reviewers equally shared the responsibility of a complete, thorough, and independent review of the Gulf of Alaska rex, dover, and flathead sole assessments. I participated in the review meeting on April 29-May 3, 2019, in Seattle, Washington, at the AFSC. The other two CIE reviewers were Patrick Cordue, and Geoff Tingley. The most current assessments were led by Carey McGilliard, with co-authors Wayne Palsson, and Benjamin Turnock. Review panel membership and a list of participants are listed in Appendix 3. During the meeting, the Panel asked questions of clarification and critiqued the work. Panel members were required to prepare their individual and independent reports after the meeting addressing the ToR as outlined in Appendix 2.

3.0 Review of the Gulf of Alaska rex, dover, and flathead sole Assessment

This is a review of the stock assessment and research activities of three flatfish species found in the Gulf of Alaska: rex sole, dover sole and flathead sole. The lead stock assessment scientist was Cary McGillard. I was one of three CIE reviewers; the other two reviewers were Patrick Cordue and Geoff Tingley. Jim Ianelli chaired the meeting and broadly outlined the assessment

cycle and how CIE reviews fit in this process. I found this overview particularly important to setting the tone of the meeting and in helping the reviewers understand how their review might be used and how our comments could be most useful. Jim, and the other reviewers, noted how the process at the AFSC was different from other regions. I find that the current process allows some mental space and time to explore different statistical methods and assumptions, as well as to help prioritize future research. This approach to stock assessment review is a strong point of the process and culture established at the AFSC, and I encourage it be continued. I believe that dover sole will be assessed again this fall, and that rex and flathead sole are scheduled to be assessed in 2020.

The three flatfish species reviewed are part of a diverse fish community in the Gulf of Alaska. The productivity of the ecosystem has been high, and catches of the principal commercial species, halibut, pacific cod, walleye pollock and salmon, have fluctuated greatly over time with the prosecution of these fisheries and changes in environmental conditions. This is the broad context in which these three flatfish species exist. It appears that rex, dover and flathead sole stocks are at relatively low abundance when compared to other commercial species. While still desirable for sale and consumption, there appears to be little, if any, directed fishery for these species, and that by and large, they are caught as bycatch. The assessment of species principally caught as bycatch poses some challenges when assessing stock status and collecting data to support the assessment.

The general stock assessment approach was typical for a species that is caught in a multi-annual survey and where the fishery catch is sampled for length and age. Cary (the assessment lead), went through previous assessments and changes made to improve the assessments from 2011 to the most recent assessment (2015 for dover, 2017 for rex and flathead). These overviews provided context for how challenges in the assessment and data were met, including changes in the modeling platform (custom built ADMB model to Stock Synthesis 3). These changes and the general evolution of each assessment were, in my view, sequential improvements. That is, they were good and necessary to better understand stock dynamics, but none drastically changed our view of stock status and trends. This relative insensitivity of stock trends to changes in approach and/or data inputs was surprising to me and counter to most of my stock assessment experience. I will explore the reasons and consequences for the stability of stock trends in this report. Many of my comments apply to the assessment of all three species.

All three species are relatively long-lived. The maximum age observed for rex sole is 48 years. Dover sole are the hardest to age and the maximum estimated age observed is 59 years. Flathead sole have a maximum estimated age of 33 years. Long-lived and slow growing species can be difficult to age and be difficult to model stock dynamics. The difficulty in modeling these species is in following cohorts in an age structured model. Several things need to happen for the data to inform the model. Population dynamics need to be different enough among years to provide some contrast in the data where cohorts can be followed through time. When the model can track cohorts, we are then able to estimate years of successful reproduction and high recruitment and measure the impact of the fishery. Such success in modeling is predicated on the quality of the length and age data collected from the survey and fishery. In the case of these three flatfish assessments, a strong effort has been made to make the most of the data available. The presentations on the GOA survey, and ageing program were strong. The GOA is enormous and

the challenges of surveying this ecosystem are great. I found Wayne Palsson's presentation very clear, and it demonstrated the thought and care put into making the survey data as valuable as possible in understanding stock dynamics and ecosystem change. I wish to take this opportunity to commend the bottom trawl survey team for their high standard and continued effort to understand and improve the survey. The logistics of running a survey are enormous, and I would support efforts to understand the data collected from the survey. I encourage Wayne (and others) to put proposals on the table for understanding gear effects, vessel effects or whatever he and the survey team feel is important to explore further. Similarly, Beth Matta presented the ageing program carefully, including ageing challenges and the ageing error associated with these three species. Without a world-class survey and ageing program at the AFSC, the uncertainty in the stock dynamics of these fish would likely be an order of magnitude greater. I encourage further work on age validation, and new technologies like Near Infrared Spectroscopy (NIR). It is, of course, hard to meet the ageing needs of all species and all assessments. In this report, I suggest less frequent assessment of these three flatfish species. However, a less frequent assessment can be balanced with more intensive sampling and ageing in the one to two years just prior to the assessment. Strategically, it may be best to rotate intensive sampling and ageing of these fish every 5 years or so. I can't tell given the data and our discussions at the meeting if this is the best approach, but I raise it here with hopes it will be discussed further by the assessment team.

The multi-species trawl survey provides an estimate of the distribution and abundance of the benthic fish community. It will, of course, sample some species better than others. Some thought and effort were put into understanding the catchability of these three flatfish species, both before (previous catchability and gear studies) and during the meeting. I was uncomfortable with the assumption that catchability was directly in proportion to the area surveyed (q=1). My discomfort regarding the assumption that q=1 extends to all surveys and to all species. In general, I think it is better to estimate catchability from the data. In the particular case of these three flatfish species, it appears that the prior strongly influences the final estimate of catchability, but regardless, it is worth the effort, and thus incorporates an important source of uncertainty into the final estimates of stock status. The issue of catchability was discussed at length at the meeting, and I was pleased that we were able to come up with a reasonable prior to catchability, and run a model estimating it. I would encourage the GOA survey team to continue research on gear and catchability, as it is likely to be an issue for multiple species. Catchability is affected by the gear used, but it is also affected by the area sampled and the proportion of the population(s) covered by the survey. Of particular concern was the catchability of dover sole, and how much of the population exists at depths deeper than 700 meters, which were infrequently sampled during the survey. We recommended that stations thatwere greater than 700m depth be dropped and that the survey series be divided into years which were sampled up to 500m and years which were sampled up to 700m, and two catchabilities estimated. The benefits and trade-offs of treating the data this way should be more closely examined. Jim Ianelli performed a quick statistical analysis estimating what proportion of the population might exist at depths greater than 700m, and it looks like a small fraction. I encourage further exploration of the distribution of all three species, but especially dover sole with respect to survey coverage.

As mentioned earlier, I was particularly concerned about the ability of the model to follow cohorts in the length and/or age data. There is considerable ageing error for these species, and, combined with their life history, it is difficult to follow cohorts and estimate cohort strength. I

was pleased to see an age validation study of rex sole, and an analysis of reader error. Maybe there hasn't been much variation in recruitment and cohort strength over the length of the survey (early 1990s to the present). Maybe these species have consistent low-level recruitment, and, because they are not heavily harvested, there is little contrast in population dynamics. An alternative view might be that given the sampling and ageing error, the length and age data are not very informative and the model does its best to follow the biomass survey trends by producing whatever recruitment is necessary. How would one test for such a possibility? As a simulation, what would happen if one doubled or tripled the abundance (proportion) of a particular cohort, would the model be able to follow it? How big would a cohort have to be for this length and age-based model to detect it? At the heart of this assessment problem (if there is indeed one) is the life history of the species and the error associated with the data collected. Perhaps we have not yet witnessed an exceptionally large recruitment event in these long-lived species. I suspect there is some differentiation in the length at age in young fish prior to the age at maturity, but very little after maturity. If so, this creates a piling up of fish at older ages with similar lengths. Add to this situation large ageing error, spatial and temporal variation in growth, and low contrast in cohort strength, and the age and length-based model doesn't tell us much more than biomass survey trends. So where does that leave us?

First, I think it is important to have gone through all the analysis of data and modeling effort that has brought us to this point. The next question is where to go from here? Well, I can see two options. One could dive further into the data and collect new data to understand stock structure, distribution, and life history (spawning timing, growth, age at maturity, etc.). For example, one could run statistical models on the current data to better understand the spatial and temporal distribution of the fishery length and age samples and their overall representativeness. This could lead to post-stratification and possibly better sampling design for these species. Another option is to take a step way back, keep an eye on survey biomass trends, and only assess these species over longer assessment cycles. The stocks seem to be in good shape, that is as far as we can tell, they are not over exploited. We do not have an estimate of unfished biomass, and are unlikely to get one anytime soon given the data. So, the stocks could be overfished and stable or underfished and stable. All we know from the survey is that the population seems stable for all three species. If the stocks seem "okay," then perhaps effort is best spent elsewhere, as long as some data and management guidelines are put in place. For example, survey biomass can still be calculated every year, and only if survey trends are drastically increasing or decreasing beyond some predetermined threshold, would further sampling and analysis take place. Another might be if more boats start directing their fishing activity toward these species, greater sampling and assessment effort can take place. I understand the AFSC has to balance both sampling effort and stock assessment expertise. I think one of the current weaknesses with particular respect to these three flatfish species, and perhaps to bycatch species in general, is a good picture of when and where they are caught in the fishery and how well they are sampled. What are the captains directing for when they catch these flatfish? What species are caught together? Is there some association? How can this information be used in assessment, and in determining the timing of the next assessment? I am not advocating for another assessment approach, the one currently used is among the best available. Rather, I am suggesting setting up some simple monitoring and triggers that would indicate the need for a deeper look at stock status. One simple monitoring tool might be the mean length and mean age of the stock from samples collected on the survey. If the GOA survey is capable of maintaining its high standard of consistency and regular sampling,

a significant change in the mean age of the population could indicate an ageing population that is not being replaced by recruitment, or conversely a younger mean age could indicate some new recruitment and/or the loss of older age classes. Either outcome could trigger a full assessment.

If the assessment team wishes to dive deeper and understand the stock dynamics better, here are a few suggestions (in no particular order).

1. Stock structure. Understanding stock structure is important for managing commercial species, and I don't think we have a good idea of the stock structure of these three flatfish species. The timing of spawning, the locations of spawning, the transport of eggs and larvae, all have consequences for our understanding of stock productivity and whether the stock is being managed properly. Of particular interest are both thought experiments and simulations that would test assumptions about stock structure on the current modeling paradigm and stock status. In particular, how does the mismatch between the distribution and abundance of rex sole relative to the survey, management zone and fishery affect the assessment and our understanding of stock status? For example, if the data were truncated such that there was a better spatial match between the spatial distribution of the population (primarily western and central GOA), survey coverage, and fishery, does that affect our view of stock status and the impact of the fishery? Just because management needs a quota at a given (larger) spatial scale doesn't mean that we shouldn't run this type of simulation to see if it has any impact on our view of the state of the population.

2. Patterns in growth and maturity. There is a considerable amount of growth data for these three species and there appears to be considerable cohort variation in growth. Further studies (spatial and/or temporal statistical models) of growth will help our understanding of stock structure and environmental influences on growth and possibly stock productivity. The maturity data collected for dover sole is sparse (only near Kodiak Island?); I am unsure of the quality of the data for the other two species. So, I recommend more work on maturity. All three species are sexually dimorphic. What is the sex ratio at length? Is that changing over time? Is there segregation of the sexes in space or time? These kinds of studies could be good for a Master's student and could help in stock assessment.

3. Spatial and temporal dynamics of the fishery. Studying the fishery is a forgone conclusion for a targeted and commercially important species. It tends to get less attention for bycatch species. These three flatfish species are often not targeted by the fishery and are not often caught. This scenario makes sampling and understanding the impact of the fishery all the more difficult. A better understanding of when, where, and how the commercial fishery operates (the fleets which happens to catch these flatfish) is important for understanding the representativeness of the biological samples collected from the fishery (observer and port sample programs) and the impact of the fishery on these flatfish stocks. A statement was made at the meeting that "total TAC is never caught, because halibut bycatch closes the fishery... [which species? all three?]" How would this management decision affect our understanding of stock status and our estimates of the impact of the fishery? A table or graph of when the fishery was closed each year might influence our understanding. I might have missed such a table, but

bringing it forward into the discussion of the history of the fishery and into the presentations would help set this context.

A close look at the protocol for biological sampling of these three flatfish species is warranted. I understand it is difficult logistically to sample fish that are not targeted or the primary species in the catch. One approach might be to rotate the intensity of sampling bycatch species on a 5-year (or some other) cycle that could match and better inform the stock assessment cycle. Focused years of sampling and ageing could be better than doing less every year. Some simulation modeling with the assessment model(s) could help guide decisions about these trade-offs.

Also, a short comment on port samples. Most of our discussion of sampling the fishery was about the observer program. While observer data is generally of a higher quality and resolution than port samples, I do think port sample data can be valuable and useful in understanding the fishery and its impacts. I encourage a deeper discussion of the value of port samples. One of the issues to consider, similar to the comments below on observer sampling, is how to scale-up port samples to the trip and possibly to the catch.

4. Model sensitivity. Model sensitivity has been explored, to some degree, by comparing different models, but I would like to see more exploration and to have those sensitivity tests discussed more in the assessment document. What is learned about the model and data when a sensitivity test is run? There was some discussion about data weighting of length composition data (McAllister and Ianelli, Francis, etc.), which I think is important, and I will let the other reviewers comment more on. What I would like to encourage is a more thorough analysis of the representativeness of the biological samples of the catch. This could be done in simple statistical models outside of the assessment model, that may (or may not) suggest alterations to current sampling programs, and may (or may not) lead to post-stratification of samples. Of course, the goal is to get an un-biased sample of the population, which is then scaled-up to the proper strata and ultimately to the catch. Current assessments used the length samples scaled-up to the haul. This may be the best method, but I would like to suggest the assessment scientists go through the process of scaling up from all the hauls to the entire catch as a heuristic process. One may return to the haul-level data and analysis, but I suspect by doing the full expansion we would find that certain samples (in space and time) have a greater impact on our view of the size structure of the catch. If so, this would be an argument for not expanding from the haul to the entire catch, but then these results would lead to a discussion of how to better sample the catch in space and time. So, I am suggesting a model/data sensitivity run on catch data expanded to the entire catch. Does this change our view of spawning stock biomass and the impact of the fishery? Perhaps tangentially, but I would be curious if the model estimate of the catchability of the survey (q) changes at all. Hopefully not, but this would be worth checking.

   Some sensitivity tests are more for building confidence in the model than getting us closer to a better estimate of stock status. For example, it seems like the trawl survey is informing the model, what happens if the survey is dropped? Would the model give a similar result? If so, why? Where are the data pointing in the same direction (coherence)?

Where is there dissonance? What are the likelihood components of each dataset? Which data set (and likelihood) is contributing the most to the final results? Similarly, if a cohort is doubled and the ageing error remains the same, can the model track it well? Is it really that useful to estimate growth within the model? What is gained or lost by including it into the model versus having it external to the model? Not every sensitivity test needs to be shown in the document in detail, but a paragraph discussing model tests would help build confidence in the model and the results.

5. "Components of Uncertainty." The assessment authors have taken a good step in including uncertainty in various places (in the data and model), but it is somewhat diffuse. I think that a section on uncertainty would help both the assessment team and reviewers assess the components of uncertainty. This could range from variance estimates used in the model, data weighting, the consequences of ageing error, to how well parameters are estimated within the model (results from MCMC runs would be great to see), to much larger issues like process error (e.g. variation in natural mortality, variation in fishery selectivity), model error (influence of model specification and structure), to large-scale environmental influences on growth and reproduction. Some of these have been done, and more can be done and are, I find, critical to the development of an assessment. In some ways, a separate section on uncertainty would force us to discuss the different components of error and how they affect our view of stock trends and management decisions.

The process for a CIE review at the AFSC is particularly constructive and collegial. One way to improve the process is for the assessment lead to write a summary document which also lays out a vision for the program. We were given previous assessments to review. Carey McGilliard, the assessment lead, presented a summary of the assessment history and model development, which was good, and led to some good discussions. However, I think a document that summarized the lessons learned from past work would provide a stronger foundation for review. As it turned out, it was up to the reviewers to collect the pieces together. This approach has some advantages, as the CIE reviewers formulate an independent view of evolution of the science and assessment program. The advantage of pulling together a summary of past assessments is that it directly challenges the reviewers to agree or disagree with the "lessons learned" and vision forward.

As a final comment, I would like to commend the assessment team for both the high level of professionalism at the meeting and quality of the work on these three flatfish assessments. A CIE review can help improve an assessment in many ways, and hopefully suggest some avenues forward that were not previously thought of. Stock assessment leads and program leads, such as those leading the trawl survey program and ageing program, have ideas on how to improve their program and the data coming from it. I would encourage all three groups to put proposals on the table, just a verbal one – "stick their neck out," so to speak, and voice ways they would like to see their program develop. By doing so, they solicit the CIE reviewers' comment on, for example, a trawl study or an ageing study and its potential to improve stock assessment for the species under consideration. It could be a way to test ideas out with the knowledge of the people in the room and for CIE reviewers to help empower research and development in stock assessment science at the AFSC.

Terms of Reference

Gulf of Alaska Rex Sole

1. *Evaluation of the ability of the stock assessment model for GOA rex sole, with the available data, to provide parameter estimates to assess the current status of rex sole in the Gulf of Alaska*

   The data used and model presented were, in my view, adequate for providing parameter estimates to assess the current status of rex sole in the GOA. Going forward, I find that it is necessary to i) drop the 1980s trawl survey CPUE data (age and length data could be retained), because these data differ too much from later surveys to be treated similarly, and ii) estimate catchability (q) from the data when possible and use an informative prior. With these changes, I have confidence in the model, data, and parameter estimates to produce an accurate assessment of the status of rex sole with an appropriate level of uncertainty.

2. *Evaluation of the strengths and weaknesses in the stock assessment model for GOA rex sole*

   In my report I raise several concerns, the 1980s survey data and the estimate of catchability being the most important. Perhaps my biggest concern is the ability of the model to follow cohorts. If this is not possible giving the life-history of this fish and the uncertainty in the data, how useful is this model? I think it still can be useful, but I urge the assessment team to consider this question.

3. *Recommendations for improvements to the assessment model.*

   My recommendations have more to do with the data put into the model than the model structure *per se*. I recommend understanding the fishery and the samples collected from the fishery better, and statistical analysis of growth patterns.

Gulf of Alaska Dover Sole (Deepwater flatfish)

1. *Evaluation of the ability of the stock assessment model for GOA Dover sole, with the available data, provide science advice to inform the management of Dover sole in the Gulf of Alaska*

   The data used and model presented were, in my view, adequate for providing parameter estimates to assess the current status of dover sole in the GOA. Going forward, I find that it is necessary to i) drop the 1980s trawl survey CPUE data (age and length data could be retained), because these data differ too much from later surveys to be treated similarly, and ii) estimate catchability (q) from the data when possible and use an informative prior. In particular, the survey time series should be divided into two, surveyed up to 500m and surveyed up to 700m, stations beyond 700m should be dropped, and a separate catchability estimated for each survey. With these changes, I have confidence in the model, data, and parameter estimates to produce an accurate assessment of the status of dover sole with an appropriate level of uncertainty.

2. *Evaluation of the strengths and weaknesses in the stock assessment model for GOA Dover sole*

In my report I raise several concerns, the 1980s survey data and the estimate of catchability being the most important. My biggest concern with all three assessment models is the ability of the model to follow cohorts. Some of the basic biology of dover sole is uncertain and is based only on a scant amount of data. For example, currently maturity data only come from one small area near Kodiak Island. So, it seems a maturity study is necessary. A start could be to determine gonad stages on the survey.

3. *Recommendations for improvements to the assessment model.*

   My recommendations have more to do with the data put into the model than the model structure *per se*. I recommend understanding the fishery and the samples collected from the fishery better, and understanding growth and maturity patterns better.

Gulf of Alaska Flathead Sole

1. *Evaluation of the ability of the stock assessment model for GOA flathead sole, with the available data, to provide parameter estimates to assess the current status of flathead sole in the Gulf of Alaska.*

   The data used and model presented here, in my view, is adequate for providing parameter estimates to assess the current status of flathead sole in the GOA. Going forward, I find that it is necessary to i) drop the 1980s trawl survey CPUE data (age and length data could be retained), because these data differ too much from later surveys to be treated similarly, and ii) estimate catchability (q) from the data when possible and use an informative prior. With these changes, I have confidence in the model, data, and parameter estimates to produce an accurate assessment of the status of flathead sole with an appropriate level of uncertainty.

2. *Evaluation of the strengths and weaknesses in the stock assessment model for GOA flathead sole.*

   In my report I raise several concerns, the 1980s survey data and the estimate of catchability being the most important. Perhaps my biggest concern is the ability of the model to follow cohorts. If this is not possible, giving the life-history of this fish and the uncertainty in the data, how useful is this model? I think it still can be useful, but I urge the assessment team to consider this question.

3. *Recommendations for improvements to the assessment model.*

   My recommendations have more to do with the data put into the model than the model structure *per se*. I recommend understanding the fishery and the samples collected from the fishery better, and understanding growth patterns better.

Appendix 1: Bibliography of materials provided for review

Gulf of Alaska Flathead Sole

- Turnock, B.J., McGilliard, C.R. and Palsson, W., J. 2017. Assessment of the Flathead Sole Stock in the Gulf of Alaska. In Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Gulf of Alaska. pp. 841-912. North Pacific Fishery Management Council, P.O. Box 103136, Anchorage AK 99510. https://www.afsc.noaa.gov/REFM/Docs/2017/GOAflathead.pdf
- McGilliard, C.R. and Palsson, W., J. 2015. Assessment of the Flathead Sole Stock in the Gulf of Alaska. In Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Gulf of Alaska. pp. 751-808. North Pacific Fishery Management Council, P.O. Box 103136, Anchorage AK 99510. https://www.afsc.noaa.gov/REFM/Docs/2015/GOAflathead.pdf
- McGilliard, C.R., Palsson, W., Stockhausen, W., and Ianelli, J. 2013. Assessment of the Flathead Sole Stock in the Gulf of Alaska. In Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Gulf of Alaska. pp. 612-756. North Pacific Fishery Management Council, P.O. Box 103136, Anchorage AK 99510. https://www.afsc.noaa.gov/REFM/Docs/2013/GOAflathead.pdf
- Stockhausen, W., Wilkins, M.E., and Martin, M.H. 2011. Assessment of the Flathead Sole Stock in the Gulf of Alaska. In Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Gulf of Alaska. pp. 753-820. North Pacific Fishery Management Council, P.O. Box 103136, Anchorage AK 99510. https://www.afsc.noaa.gov/REFM/Docs/2011/GOAflathead.pdf

Gulf of Alaska Rex Sole

- McGilliard, C.R. and Palsson, W., J. 2017. Assessment of the Rex Sole Stock in the Gulf of Alaska. In Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Gulf of Alaska. pp. 657-742. North Pacific Fishery Management Council, P.O. Box 103136, Anchorage AK 99510. https://www.afsc.noaa.gov/REFM/Docs/2017/GOArex.pdf
- McGilliard, C.R., Palsson, W., and Stockhausen, W. 2015. Assessment of the Rex Sole Stock in the Gulf of Alaska. In Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Gulf of Alaska. pp. 625-674. North Pacific Fishery Management Council, P.O. Box 103136, Anchorage AK 99510. https://www.afsc.noaa.gov/REFM/Docs/2015/GOArex.pdfStockhausen, W., Wilkins, M.E., Martin, M.H. 2011. Assessment of the Rex Sole Stock in the Gulf of Alaska. In Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Gulf of Alaska. pp. 629-690. North Pacific Fishery Management Council, P.O. Box 103136, Anchorage AK 99510. https://www.afsc.noaa.gov/REFM/Docs/2011/GOArex.pdf

Gulf of Alaska Dover Sole (Deepwater flatfish)

- McGilliard, C.R. and Palsson, W. 2015. Gulf of Alaska Deepwater Flatfish. In Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Gulf of

Alaska. pp. 563-624. North Pacific Fishery Management Council, P.O. Box 103136, Anchorage AK 99510. https://www.afsc.noaa.gov/REFM/Docs/2015/GOAdeepflat.pdf

- McGilliard, C.R., Palsson, W., Stockhausen, W., and Ianelli, J. 2013. Gulf of Alaska Deepwater Flatfish. In Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Gulf of Alaska. pp. 403-536. North Pacific Fishery Management Council, P.O. Box 103136, Anchorage AK 99510. https://www.afsc.noaa.gov/REFM/Docs/2013/GOAdeepflat.pdf
- Stockhausen, W., Wilkins, M.E., Martin, M.H. 2011. Gulf of Alaska Deepwater Flatfish. In Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Gulf of Alaska. pp. 547-628. North Pacific Fishery Management Council, P.O. Box 103136, Anchorage AK 99510. https://www.afsc.noaa.gov/REFM/Docs/2011/GOAdeepflat.pdf

Presentations and Supplementary material were provided on a Google drive link including:

1. Fishery Management Plan for the Gulf of Alaska
2. GOAfmpAppendix
3. Survey_Averaging_Model_WorkingGroupWriteup_2013_draft
4. CIE Review Overview
5. Assessment presentation for GOA rex sole
6. Assessment presentation for GOA dover sole
7. Assessment presentation for GOA flathead sole
8. GOA FMP Management Areas
9. Age determination for GOA flatfish presentation
10. Gulf of Alaska bottom trawl survey
11. North Pacific Observer Program Sampling Design

Among several other notes, figures, primary papers and documents

Appendix 2: CIE Statement of Work

Performance Work Statement (PWS)
National Oceanic and Atmospheric Administration (NOAA)
National Marine Fisheries Service (NMFS)
Center for Independent Experts (CIE) Program
External Independent Peer Review

Gulf of Alaska flatfish - Dover sole, rex sole, and flathead sole

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 assessments for Gulf of Alaska Dover sole, rex sole, and flathead sole provide the scientific basis for the management advice considered and implemented by the North Pacific Fisheries Management Council. An independent review of these integrated stock assessments is requested by the Alaska Fisheries Science Center's (AFSC) Resource Ecology and Fisheries Management Division (REFM). The goal of this review will be to ensure that the stock assessments represent the best available science to date and that any deficiencies are identified and addressed. The specified format and contents of the individual peer review reports are found in **Annex 1**. The Terms of Reference (TORs) of the peer review are listed in **Annex 2**. Lastly, the tentative agenda of the panel review meeting is attached in **Annex 3**.

Requirements

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

Tasks for Reviewers

1) Review the following background materials and reports prior to the review meeting:

Gulf of Alaska Flathead Sole

Turnock, B.J., McGilliard, C.R. and Palsson, W., J. 2017. Assessment of the Flathead Sole Stock in the Gulf of Alaska. In Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Gulf of Alaska. pp. 841-912. North Pacific Fishery Management Council, P.O. Box 103136, Anchorage AK 99510.
https://www.afsc.noaa.gov/REFM/Docs/2017/GOAflathead.pdf

McGilliard, C.R. and Palsson, W., J. 2015. Assessment of the Flathead Sole Stock in the Gulf of Alaska. In Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Gulf of Alaska. pp. 751-808. North Pacific Fishery Management Council, P.O. Box 103136, Anchorage AK 99510. https://www.afsc.noaa.gov/REFM/Docs/2015/GOAflathead.pdf

McGilliard, C.R., Palsson, W., Stockhausen, W., and Ianelli, J. 2013. Assessment of the Flathead Sole Stock in the Gulf of Alaska. In Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Gulf of Alaska. pp. 612-756. North Pacific Fishery Management Council, P.O. Box 103136, Anchorage AK 99510.
https://www.afsc.noaa.gov/REFM/Docs/2013/GOAflathead.pdf

Stockhausen, W., Wilkins, M.E., and Martin, M.H. 2011. Assessment of the Flathead Sole Stock in the Gulf of Alaska. In Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Gulf of Alaska. pp. 753-820. North Pacific Fishery Management Council, P.O. Box 103136, Anchorage AK 99510.
https://www.afsc.noaa.gov/REFM/Docs/2011/GOAflathead.pdf

Gulf of Alaska Rex Sole

McGilliard, C.R. and Palsson, W., J. 2017. Assessment of the Rex Sole Stock in the Gulf of Alaska. In Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Gulf of Alaska. pp. 657-742. North Pacific Fishery Management Council, P.O. Box 103136, Anchorage AK 99510. https://www.afsc.noaa.gov/REFM/Docs/2017/GOArex.pdf

McGilliard, C.R., Palsson, W., and Stockhausen, W. 2015. Assessment of the Rex Sole Stock in the Gulf of Alaska. In Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Gulf of Alaska. pp. 625-674. North Pacific Fishery Management Council, P.O. Box 103136, Anchorage AK 99510.
https://www.afsc.noaa.gov/REFM/Docs/2015/GOArex.pdfStockhausen, W., Wilkins, M.E., Martin, M.H. 2011. Assessment of the Rex Sole Stock in the Gulf of Alaska. In Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Gulf of Alaska. pp. 629-690. North Pacific Fishery Management Council, P.O. Box 103136, Anchorage AK 99510. https://www.afsc.noaa.gov/REFM/Docs/2011/GOArex.pdf

Gulf of Alaska Dover Sole (Deepwater flatfish)

McGilliard, C.R. and Palsson, W. 2015. Gulf of Alaska Deepwater Flatfish. In Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Gulf of Alaska. pp. 563-624. North Pacific Fishery Management Council, P.O. Box 103136, Anchorage AK 99510.
https://www.afsc.noaa.gov/REFM/Docs/2015/GOAdeepflat.pdf

McGilliard, C.R., Palsson, W., Stockhausen, W., and Ianelli, J. 2013. Gulf of Alaska Deepwater Flatfish. In Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Gulf of Alaska. pp. 403-536. North Pacific Fishery Management Council, P.O. Box 103136, Anchorage AK 99510. https://www.afsc.noaa.gov/REFM/Docs/2013/GOAdeepflat.pdf
Stockhausen, W., Wilkins, M.E., Martin, M.H. 2011. Gulf of Alaska Deepwater Flatfish. In Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Gulf of Alaska. pp. 547-628. North Pacific Fishery Management Council, P.O. Box 103136, Anchorage AK 99510. https://www.afsc.noaa.gov/REFM/Docs/2011/GOAdeepflat.pdf

2) Attend and participate in the panel review meeting. The meeting will consist of presentations by NOAA scientists, including the stock assessment authors and survey team members to facilitate the review, provide any additional information and answer questions from the reviewers.
3) After the review meeting, reviewers shall conduct an independent peer review report in accordance with the requirements specified in this PWS, OMB guidelines, and TORs, in adherence with the required formatting and content guidelines; reviewers are not required to reach a consensus.
4) Each reviewer should assist the Chair of the meeting with contributions to the summary report, if required in the terms of reference.
5) Deliver their reports to the Government according to the specified milestones dates.

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

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

Period of Performance
The period of performance shall be from the time of award through June 2019. 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
April 29 - May 3, 2019	Panel review meeting
May 17, 2019	Contractor receives draft reports
May 31, 2019	Contractor submits final reports to the Government

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

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

Travel
All travel expenses shall be reimbursable in accordance with Federal Travel Regulations (http://www.gsa.gov/portal/content/104790). International travel is authorized for this contract. Travel is not to exceed $7,000.

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

Project Contact(s):

Carey McGilliard
Resource Ecology & Fisheries Management Division
NMFS| Alaska Fisheries Science Center
7600 Sand Point Way NE, Bldg. 4, Seattle, WA 98115-6349
Phone: 206-526-4696
carey.mcgilliard@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.

Gulf of Alaska Rex Sole

1. Evaluation of the ability of the stock assessment model for GOA rex sole, with the available data, to provide parameter estimates to assess the current status of rex sole in the Gulf of Alaska
2. Evaluation of the strengths and weaknesses in the stock assessment model for GOA rex sole
3. Recommendations for improvements to the assessment model.

Gulf of Alaska Dover Sole (Deepwater flatfish)

1. Evaluation of the ability of the stock assessment model for GOA Dover sole, with the available data, provide science advice to inform the management of Dover sole in the Gulf of Alaska
2. Evaluation of the strengths and weaknesses in the stock assessment model for GOA Dover sole
3. Recommendations for improvements to the assessment model.

Gulf of Alaska Flathead Sole

1. Evaluation of the ability of the stock assessment model for GOA flathead sole, with the available data, to provide parameter estimates to assess the current status of flathead sole in the Gulf of Alaska.
2. Evaluation of the strengths and weaknesses in the stock assessment model for GOA flathead sole.
3. Recommendations for improvements to the assessment model.

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

CIE Reviewers
Patrick Cordue, Center for Independent Experts (CIE)
Geoff Tingley, Center for Independent Experts (CIE)
Kurt Trzcinski, Center for Independent Experts (CIE)

Participants

Jim Ianelli, NMFS, chair
Carey McGilliard, NMFS, stock assessment scientist
Wayne Palsson, NMFS
Craig Faunce, NMFS
Jennifer Calahan, NMFS
Beth Matta, NMFS

Agenda

CIE Review of Gulf of Alaska rex, Dover, and flathead sole

Alaska Fisheries Science Center
7600 Sand Point Way NE
Seattle, WA 98115

April 29-May 3, 2019
Alaska Fisheries Science Center, Building 4, Room 2039
Conference Line: 1-877-953-3919 (PP:5944500)

Monday, April 29, 2019

9:00-9:30 Arrive, sort out any issues with getting into the building, etc.

9:30-9:45 Introductions, settling in

9:45-10:00 Adopt agenda, review and agree upon the goals of the review (Jim Ianelli, chair)

10:00-10:20 Introduction to Gulf of Alaska flatfish fisheries, ecosystem, and management (Carey McGilliard)

10:20-10:40 Break

10:40-11:40 GOA trawl survey (Wayne Palsson)

11:40-12:20 Observer program and data (Craig Faunce and Jennifer Calahan)

12:20-1:30 Lunch

1:30-2:00 Age and growth program and data

2:00-3:30 Rex sole assessment (Carey McGilliard)

3:30-3:50 Break

3:50-5:00 Rex sole assessment

5:00 Adjourn for the day

The review from here forward involves only 1 assessment author, the CIE chair, and the CIE reviewers and can be adjusted as needed. It is ok to move through the species ahead of schedule, but we should aim to move on from species to species no later than suggested on the agenda to ensure that all three species are reviewed.

Tuesday, April 30, 2019

9:00-10:40 Continue with rex sole assessment, as needed (Carey McGilliard)

10:40-11:00 Break

11:00-12:00 Rex sole assessment, as needed (potentially writing time)

12:00-13:30 Lunch

1:30-3:30 Rex sole assessment (potentially writing time)

3:30-3:50 Break

3:50-5:00 Rex sole assessment (potentially writing time)

Wednesday, May 1, 2019

9:00-10:40 Dover sole assessment

10:45-11:00 Break

11:00-12:30 Dover sole assessment, continued

12:30-2:00 Lunch

1:30-2:00 Dover sole assessment, continued

2:00-4:00 Break, writing time: *Carey and Jim will attend another meeting at this time*

3:45-5:00 Dover sole assessment (potentially writing time)

5:00 Adjourn

Thursday, May 2, 2019

9:00-10:40 Dover sole assessment (potentially writing time)

10:40-11:00 Break

11:00-12:30 Dover sole assessment (potentially writing time)

12:30-2:00 Lunch

2:00-3:30 Flathead sole assessment (Carey McGilliard)

3:30-3:50 Break

3:50-5:00	Flathead sole assessment, continued
5:00	Adjourn

Friday, May 3, 2019

9:00-10:30	Flathead sole assessment, as needed
10:30-10:50	Break
10:50-12:30	Flathead sole assessment (potentially writing time)
12:30-2:00	Lunch
2:00-3:30	Flathead sole assessment (potentially writing time)
3:30-3:50	Break
3:50-5:00	Flathead sole assessment (potentially writing time)
5:00	Adjourn

Appendix 4: Summary Report

CIE Review of assessments for Gulf of Alaska rex, Dover, and flathead soles

April 29-May 3, 2019

Alaska Fisheries Science Center, Building 4, Room 2039, Seattle

Patrick Cordue, Center for Independent Experts (CIE)
Geoff Tingley, Center for Independent Experts (CIE)
Kurt Trzcinski, Center for Independent Experts (CIE)

Participants

Jim Ianelli, NMFS, chair
Carey McGilliard, NMFS, stock assessment scientist
Wayne Palsson, NMFS
Craig Faunce, NMFS
Jennifer Calahan, NMFS
Beth Matta, NMFS

Summary

A CIE review of three stock assessments of Gulf of Alaska (GOA) flatfish stocks was conducted at the Alaska Fisheries Science Center from April 29 to May 3 2019. The participants included three CIE reviewers, the primary assessment author, the chair of the meeting and NMFS staff who presented on relevant topics.

On the first day, an introductory presentation was given on the GOA ecosystem and flatfish fisheries. Presentations on the GOA trawl survey, the observer program, and the ageing of flatfish were also given. Stock assessment presentations for the three species were given over the following days.

The stock assessments were primarily conducted by the same author who transitioned the assessments from purpose written code to Stock Synthesis 3 (SS3) in 2013 (Dover and flathead) and 2015 (rex). Subsequent assessments have primarily been refinements of the models developed in 2013 and 2015.

The assessment models and the use of data in the assessments were similar across the three assessments. Therefore, the assessments broadly shared the same strengths and weaknesses. In general, the age-structured models were appropriate given the available biological, abundance, and composition data. A particular strength of the assessments is the availability of a consistent time series of biomass estimates from the GOA trawl surveys (in particular since 1996).

The preparation of the input data can be improved in some respects. More exploratory and formal analysis of the composition data is required so that length, age, and age-at-length data can be appropriately post-stratified (if necessary) and scaled. The trawl biomass time series also needs to

be treated carefully, especially for species which have a distribution below 500 m (the maximum depth of the survey in some years).

The assumption that the trawl survey biomass indices are estimates of absolute biomass ($q = 1$) is inappropriate for most stock assessments. It is better to estimate the "catchability" (q) and support the estimation with an informed prior (which contains the currently available information on the value of q). A first attempt at producing an informed q prior for each stock was performed during the meeting and model runs were performed with the informed priors. Although the point estimates of spawning biomass and stock status were similar to the original models the results reflected a greater and much more appropriate level of uncertainty.

The reviewers appreciated the excellent presentations by the NMFS staff, the hard work of the assessment author, and the collegial and constructive atmosphere under which the review meeting was conducted.

Main Recommendations

These recommendations address common issues found in each of the three assessments reviewed, and that may also be relevant for other assessments. These were agreed by the three CIE reviewers.

Gulf of Alaska Bottom Trawl Survey (BTS)

1. The surveys conducted in 1984 and 1987 used different vessels, a different approach and with different timing. These surveys should not be considered as part of the same timeseries as the subsequent BTS timeseries. Specifically, the biomass estimates and the composition data from these two surveys should be dropped from each of these assessments, and probably from all other assessments also.
2. The surveys in 1990 and 1993 had a different timing (later) and somewhat different survey structure. While clearly not as 'different' as the 1984 and 1987 surveys, there is sufficient difference that model sensitivities should be run on a species-by-species (stock-by-stock) basis that include and exclude the biomass and composition data from these two surveys.
3. Where there are gaps in survey data due to, for example, not surveying some areas in some years, these should be left as data gaps. The model structures used are more than capable of dealing with such data gaps. Data should not be created by extrapolation, interpolation or modelling to fill such gaps.

Fishery sampling

4. A more consistent, analytical and defensible approach to the scaling and stratification of fisheries data should be followed. This should meet accepted 'best practice' approaches, including, for example, studying the spatial and temporal patterns of length and age followed by appropriate stratification and scaling.

Modelling

5. Models should not assume that the survey q is equal to 1. Informed priors should be developed on a stock-by-stock basis.
6. Recruitment deviates should not be estimated where there is no information to inform the estimation i.e. there has to be age data from a survey or fishery to inform the estimation process.

Observer data to support the stock assessments

7. The Observer Program delivers information to support stock assessments for a large number of groundfish stocks. On the whole this works very well but is not the case for all stocks. With respect to this review, age data for Dover sole from the fishery are, due to the scale of the fishery and the sampling prioritisation approach of the Observer Program, insufficient to provide any recent age frequencies for use in the assessment. In addition, for some bycatch species there will be a real prospect of sampling being unrepresentative. The development of alternative Observer Program sampling strategies for low catch and bycatch fisheries to provide the required data to support the assessments should be conducted as a matter of priority.