Center for Independent Experts (CIE) Peer Review Report

Gulf of Alaska Flatfish Assessments - Dover sole, rex sole, and flathead sole

Prepared for the Center for Independent Experts

By

Dr. Geoff Tingley

Gingerfish Ltd
email: fishinnz@hotmail.co.uk

July 2019

Table of Contents

Executive Summary

- This document is the individual CIE Reviewer Report of the review of the stock assessments of the Gulf of Alaska stocks of Dover sole (*Microstomus pacificus*), rex sole (*Glyptocephalus zachirus*) and flathead sole (*Hippoglossoides elassodon*). The review was conducted during April and May 2019, with the review meeting held from 29th April to 3rd May. This report solely represents the views of the independent CIE reviewer Geoff Tingley.
- The current and recent historical assessment documents were provided for each species in advance of the review meeting and additional supporting documents were provided during the meeting. All documents are listed in Appendix 1.
- The assessments for each of the three stocks were clearly presented and supported by clear documentation, including detailed descriptions of the input data and an appropriate level of coverage of the uncertainties. The assessment analyst fully engaged with the review in a highly professional and constructive manner. The assessment presentations were supported by clear and informative presentations on the GOA ecosystem, GOA Bottom Trawl Survey, the North Pacific Observer Program, and the GOA Age and Growth Program, all of which were key to the understanding of the assessments being reviewed.
- During the review process, the CIE reviewers identified some shortcomings in each of the assessments. While none of the identified issues were critical, the reviewers worked with the assessment analyst and Chair to find alternative approaches and, where possible, develop model runs that explored and addressed the more important and more tractable of these issues during the review meeting.
- The assessment models and the use of data in the assessments were broadly similar for each assessment and the assessments shared the basic same strengths and weaknesses. In general, the age-structured models were appropriate for the stocks given the available biological, abundance, and composition data. A particular strength of the assessments is the availability of a consistent timeseries of biomass estimates from the GOA Bottom Trawl Survey (in particular since 1996) and for rex sole, high quality age data from the survey and fishery.
- All three stock assessments represent the best available science and are of sufficient technical quality to be used for informing management. Given the assessments, the current management approach and the scale and intensities of the three fisheries, there are no sustainability concerns for any of the three stocks. None of the three stocks are experiencing overfishing and none are overfished.
- There are a number of areas where further development work on data availability and quality may yield improved model fits and/or a reduction in uncertainty. There are also some changes to assessment modelling approaches that may yield improved assessments in future.
- Specific recommendations aimed at improving each of these stock assessments, for improving input data for these (and possible other) stocks, and for general processes are made by the reviewer as required by the reviewer Terms of Reference.
- A Summary Report of recommendations common to all three CIE reviewers is also attached (Appendix 4).

Background

This review of the 2017 Stock Assessment Reports for the Gulf of Alaska (GOA) rex and flathead soles and the 2015 Stock Assessment Report for GOA Dover sole was conducted as part of an independent review for the Center for Independent Experts (CIE).

All views expressed in this report are solely those of the independent CIE reviewer.

The key assessment reports, with supporting background documents and reports, were comprehensive, well written and clearly presented. The support provided by the local NMFS staff, particularly that provided by Carey McGilliard and Jim Ianelli, was of a high standard and much appreciated by all three CIE reviewers.

The deepwater complex is composed of three species: Dover sole (*M. pacificus*), Greenland turbot (*Reinhardtius hippoglossoides*) and deepsea sole (*Embassichthys bathybius*). Dover sole dominates the biomass of the deepwater complex in research trawl surveys and fishery catch (typically over 98%). Dover sole was the only stock within the deepwater complex under review at this meeting.

These three fisheries are all relatively small, with five-year average catch prior to the most recent assessment being about 2,600 t , 350 t and 2,400 t for the rex sole, Dover sole and flathead sole fisheries respectively.

The Chair ran the meeting, facilitated discussions within the group, and provided technical support to the stock assessment analysist as required. The various presenters provided clear and informative background on their individual areas of expertise and responsibility for the review team, covering the ecosystem, fish biology, sampling, ageing and assessment. The presenters also responded fully to any questions the review team raised. The CIE reviewers considered all of the documents and presentations, asked questions seeking clarification, and engaged to offer alternative approaches where they considered such approaches would lead to improvements in one or more of the assessments.

Description of Review Activities

This review was undertaken by Geoff Tingley between the 14th April and 22nd May 2019. The timing of the review meeting proceeded as scheduled from 29th April to 3rd May 2019.

The supporting documentation for the review of the assessment were provided to the reviewers in electronic format adequately in advance of the review meeting as intended. These documents included historic and current stock assessments. Additional relevant documents detailing aspects of the stocks, sampling and other related science necessary for a full understanding of these fisheries and their assessments were made available electronically during the meeting as and when it became clear that these may be of use. Copies of the various presentations and additional work conducted during the review meeting were also provided as and when available. All documents provided are listed in the Bibliography (Appendix 1).

All documents provided in advance of the meeting were reviewed prior to the start of the meeting.

The assessments were reviewed against the specific Terms of Reference (ToR) provided by the CIE in the Performance Work Statement (Appendix 2).

Information relevant to this review is presented in three appendices to this review report, as required by the ToR. These are, Appendix 1: Bibliography of documents; Appendix 2: CIE Performance Work Statement (which includes its own annexes describing the (1) Peer review report requirements, (2) the ToR for the peer review, and (3) the draft agenda (as agreed at the start of the meeting) for the review meeting; and Appendix 3: Panel membership and other relevant information and decisions.

At the start of the meeting, the agenda was discussed and agreed with minor changes. More importantly, the interpretation of the coverage of the review to include data inputs for all three stock assessments was also agreed. The meeting was conducted in an open, friendly and constructive manner throughout. Presentations were made with questions asked by members of the panel. All discussions were professional and good natured, being focused on clarification and clarity around the assessments under review. Other than the reviewers and the presenters, there were no other attendees of the meeting.

Final outputs from all additional model runs requested and developed during the meeting, as well as responses to panel questions, were made available to the reviewers as soon as completed and all reviewer requests were responded to before the end of the meeting.

There was a high degree of agreement between the individual panel members on the quality, main issues and status of the three assessments.

Summary of Findings

The same stock assessment analysist was responsible for all three assessments and did an excellent job of transitioning the three assessments from stock-specific assessment code to the Stock Synthesis 3 (SS3) framework. The assessment reports were concisely written but included virtually all the required information for review.

That each assessment used the same modelling framework and that all three models were of similar construction made the review process easier and allowed for much direct comparison between the different assessments.

Areas of the assessments where one or more reviewer considered there was opportunity to improve on the assessment approach were explored. Some of these areas were common to all assessments, others were assessment specific. These are discussed in some detail and appropriate recommendations made below. There was considerable similarity between the rex and flathead assessments. The Dover sole assessment was somewhat different due to the nature of the fishery and the biology and ecology of Dover sole.

Additional model runs for each assessment were developed during the review meeting, with the majority of effort focused on the rex and Dover sole assessments. This testing of the assessments did highlight some areas where improvements could be made in future assessments. The testing also demonstrated that all three assessments are of a high quality, represent the best available science, and are appropriate for providing management advice.

The types, amount and quality of data available to assess rex and flathead sole are more than sufficient to enable assessments of very high quality to be developed, although there are some areas where improvements can be delivered. The most important issues for future assessments for all three stocks to address include one issue of early survey data suitability and one of addressing the relative weighting of the abundance index and composition data within the models. The Dover sole assessment also has specific issues associated with the survey depth coverage and a lack of age-frequency data from the fishery.

The North Pacific Bottom Trawl Survey was found to be a high-quality data source, providing fishery-independent abundance (biomass) information for each assessment. However, all three CIE reviewers agreed that the very earliest survey data included in the assessments were sufficiently different in origin (different vessel, different months, etc.) that it was inappropriate to consider these as part of the abundance time-series and recommended their removal from the information used in future assessments for these stocks. For the Dover sole assessment, the survey also only covered the deeper strata (500m to 700m and 700m to 1,000m depth) in some years, which when coupled with the depth distribution of Dover sole created additional uncertainty for this assessment.

Future assessments would benefit from improvements in the careful consideration of input data, with selection based on defined quality criteria. This is especially so for the fishery-independent abundance index from the trawl survey, which should be assured as the most influential dataset in the assessments.

Each of the most recent assessment reports included a small number of identified data gaps and research recommendations. For completeness, these are presented following the reviewer's recommendations in this report.

Recommendations for research and development work for future assessments for these stocks were considered and . These have been prioritized in this report on a personal basis. While presented as separate for each stock, most of the recommendations are the same for the three assessments, as the assessments all followed the same basic design and approach, and used similar data. There is, however, a notable difference in the recommendations for Dover sole, where there is a recommendation to address an inadequacy in the collection of age data from the fishery; this recommendation is not shared with the other two assessments here, but is probably shared with a number of other minor species assessments for GoA stocks.

Detailed findings and recommendations are presented below as required by the ToR for the review.

By the end of the review meeting, the approach to modeling the three stocks had been thoroughly explored, was considered thorough and sound, and appropriately addressed uncertainty to the principle assumptions through the range of models and sensitivities explored. The ranges of input data available and used were clearly described. The overall outcome of these assessments, as reviewed, is that they meet the description of best available science and are of sufficient quality to be used to inform management.

Addressing the Terms of Reference for the Peer Review

A: 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 2017 age-structured assessment of the GOA rex sole stock was newly transitioned to the SS3 assessment framework. There was an audit trail for this transition (i.e. the steps of the transition were tracked and reported on) and this was fully explained during the review.
   - The SS3 framework is more appropriate for delivering consistent assessments to inform management than using the bespoke assessment code used previously.
   - The data available from the fishery and from fishery-independent sources were of a high standard and provided for all of the necessary inputs to the model to deliver a high quality stock assessment, including landings data, discard information, length and age-frequencies from the fishery and the survey, biological information (growth, maturity, etc.), and a high quality, fishery-independent biomass index from the survey.
   - The design and implementation of the assessment model was appropriate to the biology of rex sole and the type, scale and extent of the fishery.
   - The assessment appropriately considered and addressed the main uncertainties in the data and the assumptions necessary to develop and implement the model.
   - The assessment, using the available data, was therefore evaluated to be able to provide sufficiently accurate estimates of the current (2017) status of the rex sole stock that will be useful, usable and informative for management.

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

   - Key strengths of the 2017 assessment of the GOA rex sole stock include changes to the assessment framework, the design of the age-structured model and the quality, quantity and range of data available.
   - The updating of the assessment framework to SS3, coupled with appropriate model design and implementation represents a strengthening of the assessment.
   - There were considerable data available to the model, most of which were of a high quality. Of specific note were the fishery-independent biomass index from the survey; high quality ageing giving informative age-frequencies from the survey and the fishery; length-frequency information from the survey and from the fishery (from the Observer Program).
   - Differences in the spatial growth rates of rex sole were accounted for in the assessment.
   - None of the observed weaknesses in the 2017 assessment were critical, but all can and should be addressed to further strengthen the future assessments and, in particular, reduce uncertainty in the results.
   - The most notable weaknesses in this assessment are:
     (i) the inclusion of a number of the Northern Pacific Bottom Trawl Survey data points from earlier years (1984, 1987; and 1990 and 1993), where there were differences in the vessel, approach and/or timing that are likely to mean that these points are not part of the main timeseries; the 1984 and 1987 surveys used different vessels, a different

approach and different timing, and as such should not be considered part of the main BTS timeseries. The 1990 and 1993 surveys occurred later in the year and had a somewhat different structure, with sufficient difference to consider these points as not part of the main biomass timeseries;

(ii) filling in gaps in the biomass index timeseries (e.g. where some areas were not surveyed in some years), the model is designed to handle gaps in the data, it is neither necessary nor desirable to create data to fill such gaps;

(iii) lack of consistency in the stratification of the fishery data;

(iv) the survey *q* was assumed to equal 1 (a common but unjustified assumption in a number of US assessments);

(v) estimation of recruitment deviates where there are no informative recruitment data; and

(vi) disproportionally high weighting on the composition data relative to the survey biomass index. Best practice supports higher weight in determining the biomass, and thus, stock status in most assessments being placed on reliable biomass indices not on composition data.

- During the review meeting different approaches to addressing the identified weaknesses were tried in additional model runs and flow through into the recommendations below. Of note:

  (i) the earliest years of the trawl survey (1984 and 1987), including biomass estimates and composition data, were dropped from the assessment. The second set of years, 1990 and 1993 were either dropped or included as a separate timeseries in a sensitivity;

  (ii) all gap-filling data were dropped from the assessments;

  (iii) consistent restratification of the fishery data was considered too big a task to attempt during the review meeting;

  (iv) approximate priors for the survey *q* were developed and applied resulting in model derived estimates of survey *q*;

  (v) where there were no age data, recruitment deviates were no longer estimated in model runs conducted during the review meeting;

  (vi) a number of attempts to improve the balance of the weighting between the biomass and composition data were made using two different approaches. While some improvement in the relative weighting and subsequent fit to data was achieved, this was not consistent or conclusive, and aspects remained poorly understood.

3. Recommendations for improvements to the assessment model.

Peer reviewer recommendations for GOA rex sole

- Drop the 1984 and 1987 survey biomass and survey composition data from all future assessments, as these are clearly not part of the longer survey timeseries due to the use of different vessels, a different survey approach and different survey timing.
- Either drop the 1990 and 1993 survey biomass and survey composition data from all future assessments or include them only as a separate timeseries in specific sensitivities,

as these do not appear to be part of the longer survey timeseries due to different timing and survey structure.

- Do not fill data gaps by creating, extrapolating, interpolating or modelling data as this may introduce greater uncertainty or bias into the results. The age-structured model and the SS3 framework are designed to handle gaps in datasets. This is notable for survey data for this assessment.
- A more consistent, analytical and defensible approach to the scaling and stratification of fisheries data should be applied. This should follow accepted 'best practice', using approaches, such as explicitly exploring the spatial and temporal patterns of fish length and age to inform appropriate stratification and scaling.
- Conduct further analyses and investigations in order to find the most appropriate weighting for the composition data to ensure the biomass signal from the survey index is appropriately strong.
- Models should not assume that the survey *q* is equal to 1. Stock-specific informed priors for survey *q* should be developed and then used within the models.
- Recruitment deviates should not be estimated where there is no information to inform the estimation, i.e., there have to be age data from a survey or fishery to inform the estimation process.

For completeness, there were the following research recommendations included in the 2017 rex sole assessment report:

> Updated information on maturity-at-age for GOA rex sole would reduce uncertainty in the maturity curve relative to the fishery selectivity curve, as this is important for the determination of F_{OFL} and F_{ABC} for this stock. The ADF&G small mesh survey could be included as well, and an ageing error matrix could be developed. Further exploration of natural mortality rates and catchability for GOA rex sole could be conducted.
>
> This assessment showed that growth curves in the Eastern GOA differ from those in the Western and Central GOA. The age and growth laboratory previously noted that GOA rex sole otoliths appear to show two different patterns for the same age and year of fish, corroborating the results of this assessment. Further research could be conducted to determine whether the two growth patterns represent two genetic sub-stocks or one genetic sub-stock where environmental conditions or other ecosystem dynamics contribute to different growth rates in the two regions modeled in this assessment.

Of these two identified research areas, the reviewer suggests that developing an appropriate way to include ageing error and updating maturity-at-age data may be worth pursuing.

B: 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 2015 assessment of the GOA Dover sole stock was newly transitioned to the SS3 assessment framework. There was an audit trail for this transition (i.e., the steps of the transition were tracked and reported on) and this was fully explained during the review.
- The SS3 framework is more appropriate for delivering consistent assessments to inform management than using the bespoke assessment code used previously.
- The data available from the fishery and from fishery-independent sources were of a high standard and provided for the necessary inputs to the model to deliver a high quality stock assessment, including landings data, discard information, length and age-frequencies from the survey, biological information (growth, maturity, etc.), and an adequately high quality fishery-independent biomass index from the survey. Fishery age-frequency data were not available.
- The design and implementation of the assessment model was broadly appropriate to the biology of Dover sole and the type, scale and extent of the fishery.
- The assessment appropriately considered and addressed most of the main uncertainties in the data and the assumptions necessary to develop and implement the model.
- The assessment, using the available data, was therefore evaluated to be able to provide sufficiently accurate estimates of the current (2015) status of the Dover sole stock that will be useful, usable and informative for management.

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

- Key strengths of the 2015 assessment of the GOA Dover sole stock include changes to the assessment framework, the design of the age-structured model and the quality, quantity and range of data available.
- The updating of the assessment framework to SS3, coupled with appropriate model design and implementation represents a strengthening of the assessment.
- There were considerable data available to the model, most of which were of a high quality. Of specific note were the fishery-independent biomass index from the survey (although this has some stock-specific issues considered in the weakness section below); high quality ageing giving informative age-frequencies from the survey; and length-frequency information from the survey.
- None of the observed weaknesses in the 2015 assessment were critical, but all can and should be addressed to further strengthen future assessments and, in particular, reduce uncertainty in the results. Overall, this was probably the weakest of the three assessments.
- The most notable weaknesses in this assessment are:
  (i) the inclusion of a number of the Northern Pacific Bottom Trawl Survey data points from earlier years (1984, 1987; and 1990 and 1993), where there were differences in the vessel, approach and/or timing that are likely to mean that these points are not part of

the main timeseries; the 1984 and 1987 surveys used different vessels, a different approach and different timing, and as such should not be considered part of the main BTS timeseries. The 1990 and 1993 surveys occurred later in the year and had a somewhat different structure, with sufficient difference to consider these points as not part of the main biomass timeseries;

(ii) filling in gaps in the biomass index timeseries (e.g., where some areas were not surveyed in some years), the model is designed to handle gaps in the data, it is neither necessary nor desirable to create data to fill such gaps;

(iii) lack of consistency in the stratification of the fishery data;

(iv) the survey *q* was assumed to equal 1 (a common but unjustified assumption in a number of US assessments) and only a single q was used across all depth strata;

(v) estimation of recruitment deviates where there are no informative recruitment data; and

(vi) disproportionally high weighting on the composition data relative to the survey biomass index. Best practice supports higher weight in determining the biomass and thus stock status in most assessments being placed on reliable biomass indices not on composition data.

Specific to the Dover sole assessment only, were two other weaknesses:

(vii) as the depth distribution of Dover sole is recorded to below 1,000m depth, the survey does not cover the full distribution of the stock as the survey does not go deeper than 1,000m in any year. The survey also has a variable depth coverage by year, with coverage to a depth of 500m in some years, and to 700m or 1,000m in other years. These issues create problems for appropriately applying the swept area biomass estimates and for estimating survey *q*.

(viii) As a result of the scale of the fishery, which only has a catch of about 350 t per year, combined with the protocol for sampling minor species by the Observer Program, there are no recent years for which usable age-frequency data are available for the fishery. This creates some important limitations with respect to estimating selectivity in the fishery and is an important issue to address for future stock assessments.

- During the review meeting, different approaches to addressing the identified weaknesses were tried in additional model runs and flow through into the recommendations below. Of note:

  (i) the earliest years of the trawl survey (1984 and 1987), including biomass estimates and composition data, were dropped from the assessment. The second set of years, 1990 and 1993 were either dropped or included as a separate timeseries in a sensitivity analysis;

  (ii) all gap-filling data were dropped from the assessments; the 2001 survey was dropped altogether as the southeast Gulf was not sampled in this year;

  (iii) consistent restratification of the fishery data was considered too big a task to attempt during the review meeting;

  (iv and vii) approximate priors for the survey *q* were developed and applied resulting in model derived estimates of survey *q*. This was more complex for this stock as three survey timeseries needed to be developed (to 500m, to 700m, and to 1,000m) each with an informed prior for *q*. Information from the West Coast BT survey, which goes deeper than the Northern Pacific BT survey, was used to help develop the *q* priors which were

roughly estimated as 1.2 for years when the survey went to 1,000m, 1.17 for survey years to 700m, and 1.08 for survey years to 500m only;

(v) where there were no age data, recruitment deviates were no longer estimated in model runs conducted during the review meeting;

(vi) a number of attempts to improve the balance of the weighting between the biomass and composition data were made using two different approaches. While some improvement in the relative weighting and subsequent fit to data was achieved, this was not consistent or conclusive, and aspects remained poorly understood;

(vii) see (iv) above;

(viii) correcting for the lack of fishery age data was not possible during the review meeting.

3.Recommendations for improvements to the assessment model.

Peer reviewer recommendations for GOA Dover sole

- The Observer Program delivers information to support stock assessments for a large number of groundfish stocks. On the whole, the approach used works well but this is not the case for all stocks. Age sampling of Dover sole from the fishery, due to the scale of the fishery and the sampling prioritisation approach of the Observer Program, provides insufficient samples to create any usable, recent age frequencies for use in the assessment. This is unlikely to change in the future while the catch of Dover sole remains low and without change in sampling priorities with the Observer Program. The simplest approach to address this would be not to collect any Dover sole samples in most years, but to have a specific year where observers collect Dover sole length and age data from all fishery tows sampled to ensure that (i) sufficient age and length data are collected to enable high quality age and length frequencies from the fishery to be developed, and (ii) ensure that the AF and LF sampling are representative of the fishery. The specific years selected for sampling Dover sole should precede the stock assessment and be determined by the timing of stock assessment and the time required for reading the otoliths and developing the AFs and LFs in advance of the stock assessment. The same basic approach could be followed for any stock which has the same issue of low catches that prevent adequate age sampling from the fishery under the standard Observer Program sampling protocols. It is recommended to initiate a full-scale trial of this recommended approach prior to the next assessment of the Dover sole stock, even if that means that the assessment needs to be delayed by a year.
- An associated issue with low level sampling of what are essentially bycatch species (including Dover sole) by the Observer Program is that there will be a real prospect of sampling being unrepresentative of the fishery and so of limited value in an assessment. The development of alternative Observer Program sampling strategies to ensure representative sampling for low catch and bycatch fisheries to provide the required data to support stock assessments should be conducted as a matter of priority.
- Given the low level of catch, the current good stock status and trend, and the longevity of Dover sole, the requirement for a stock assessment every four years could be relaxed to every five or six years with little additional sustainability risk. It is recommended that those responsible for stock assessment planning consider these options in order to

review the potential to free up scarce resources for other stocks in need of more active monitoring. If such a change were implemented, a review trigger that responds to an increase in the catch of Dover sole and that would increase the frequency of assessments should the catch rise above specific trigger amounts should also be implemented.

- Drop the 1984 and 1987 survey biomass and survey composition data from all future assessments as these are clearly not part of the longer survey timeseries due to the use of different vessels, a different survey approach and different survey timing.
- Either drop the 1990 and 1993 survey biomass and survey composition data from all future assessments or include them only as a separate timeseries in specific sensitivities, do not appear to be part of the longer survey timeseries due to different timing and survey structure.
- Do not fill data gaps by creating, extrapolating, interpolating or modelling data as this may introduce greater uncertainty or bias into the results. The age-structured model and the SS3 framework are designed to handle gaps in datasets. This is notable for survey data for this assessment.
- A more consistent, analytical and defensible approach to the scaling and stratification of fisheries data should be applied. This should follow accepted 'best practice', using approaches, such as explicitly exploring the spatial and temporal patterns of fish length and age to inform appropriate stratification and scaling.
- Conduct further analyses and investigations in order to find the most appropriate weighting for the composition data to ensure the biomass signal from the survey index is appropriately strong.
- Models should not assume that the survey *q* is equal to 1. Stock-specific informed priors for survey *q* should be developed and then used within the models. The '*back of the envelope*' estimation of *q* priors for the different survey components conducted during the meeting should be redone in a more rigorous way and fully documented.
- Recruitment deviates should not be estimated where there is no information to inform the estimation, i.e., there have to be age data from a survey or fishery to inform the estimation process.

For completeness, there were the following research recommendations included in the 2015 Dover sole assessment report:

> The 2013 and 2015 stock assessment incorporated ageing error by using an existing ageing error matrix for West Coast Dover sole. A priority for future assessments is to analyze ageing error data for GOA Dover sole using methods described in Punt *et al*. (2008) and to incorporate a resulting ageing error matrix into the assessment. The assessment would benefit from an exploration of ways to better account for scientific uncertainty, especially uncertainty associated with parameters that are currently fixed in the model, including an exploration of natural mortality and catchability. The full coverage survey selectivity estimates indicate that males are selected at younger ages than females, which is counterintuitive. Future research could be done to explore this phenomenon.

The reviewer suggests that developing an appropriate way to include ageing error is worth pursuing.

C: 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 2017 assessment of the GOA flathead sole stock was newly transitioned to the SS3 assessment framework. There was an audit trail for this transition (i.e., the steps of the transition were tracked and reported on) and this was fully explained during the review.
- The SS3 framework is more appropriate for delivering consistent assessments to inform management than using the bespoke assessment code used previously.
- The data available from the fishery and from fishery-independent sources were of a high standard and provided for all of the necessary inputs to the model to deliver a high quality stock assessment, including landings data, discard information, length and age-frequencies from the fishery and the survey, biological information (growth, maturity, etc.), and a high quality fishery-independent biomass index from the survey.
- The design and implementation of the assessment model was appropriate to the biology of flathead sole and the type, scale and extent of the fishery.
- The assessment appropriately considered and addressed the main uncertainties in the data and the assumptions necessary to develop and implement the model.
- The assessment, using the available data, was therefore evaluated to be able to provide sufficiently accurate estimates of the current (2017) status of the flathead sole stock that will be useful, usable and informative for management.

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

- Key strengths of the 2017 assessment of the GOA flathead sole stock include changes to the assessment framework, the design of the age-structured model and the quality, quantity and range of data available.
- The updating of the assessment framework to SS3, coupled with appropriate model design and implementation represents a strengthening of the assessment.
- There were considerable data available to the model, most of which were of a high quality. Of specific note were the fishery-independent biomass index from the survey; relatively high-quality ageing giving informative age-frequencies from the survey and the fishery; length-frequency information from the survey and from the fishery (from the Observer Program).
- None of the observed weaknesses in the 2017 assessment were critical, but all can and should be addressed to further strengthen future assessments and, in particular, reduce uncertainty in the results.
- The most notable weaknesses in this assessment are:
  (i) the inclusion of a number of the Northern Pacific Bottom Trawl Survey data points from earlier years (1984, 1987; and 1990 and 1993), where there were differences in the vessel, approach and/or timing that are likely to mean that these points are not part of the main timeseries; the 1984 and 1987 surveys used different vessels, a different approach and different timing, and as such should not be considered part of the main BTS timeseries. The 1990 and 1993 surveys occurred later in the year and had a

somewhat different structure, with sufficient difference to consider these points as not part of the main biomass timeseries;

(ii) filling in gaps in the biomass index timeseries (e.g., where some areas were not surveyed in some years), the model is designed to handle gaps in the data, it is neither necessary nor desirable to create data to fill such gaps;

(iii) lack of consistency in the stratification of the fishery data;

(iv) the survey *q* was assumed to equal 1 (a common but unjustified assumption in a number of US assessments);

(v) estimation of recruitment deviates where there are no informative recruitment data; and

(vi) disproportionally high weighting on the composition data relative to the survey biomass index. Best practice supports higher weight in determining the biomass, and thus stock status in most assessments being placed on reliable biomass indices not on composition data.

- During the review meeting different approaches to addressing some of the identified weaknesses were tried in additional model runs and flow through into the recommendations below. Due to time constraints, the additional development and testing of this assessment was less fulsome than for the other two stocks, but the approach should broadly follow that seen for rex sole. Of note:

  One model run was also undertaken where the biomass index was removed and the model fitted just to the composition data to test the relative importance and informativeness of these data.

3. Recommendations for improvements to the assessment model.

Peer reviewer recommendations for GOA flathead sole

- Drop the 1984 and 1987 survey biomass and survey composition data from all future assessments, as these are clearly not part of the longer survey timeseries due to the use of different vessels, a different survey approach and different survey timing.
- Either drop the 1990 and 1993 survey biomass and survey composition data from all future assessments or include them only as a separate timeseries in specific sensitivities, do not appear to be part of the longer survey timeseries due to different timing and survey structure.
- Do not fill data gaps by creating, extrapolating, interpolating or modelling data, as this may introduce greater uncertainty or bias into the results. The age-structured model and the SS3 framework are designed to handle gaps in datasets. This is notable for survey data for this assessment.
- A more consistent, analytical and defensible approach to the scaling and stratification of fisheries data should be applied. This should follow accepted 'best practice', using approaches, such as explicitly exploring the spatial and temporal patterns of fish length and age to inform appropriate stratification and scaling.
- Conduct further analyses and investigations in order to find the most appropriate weighting for the composition data to ensure the biomass signal from the survey index is appropriately strong.

- Models should not assume that the survey q is equal to 1. Stock-specific informed priors for survey q should be developed and then used within the models.
- Recruitment deviates should not be estimated where there is no information to inform the estimation, i.e., there have to be age data from a survey or fishery to inform the estimation process.
- An appropriate prior for natural mortality, M, should be developed and used in the model.

For completeness, there were the following research recommendations included in the 2017 flathead sole assessment report:

> The 2015 and 2017 stock assessments incorporated ageing error by using an existing ageing error matrix for BSAI flathead sole. A priority for future assessments is to analyze ageing error data for GOA flathead sole using methods described in Punt *et al*. (2008) and to incorporate a resulting ageing error matrix into the assessment. Future analyses should explore the relationship between natural mortality and catchability in the model, alternative parameter values, and the effects of these parameters on estimation of selectivity and other parameters. The assessment would benefit from an exploration of ways to better account for scientific uncertainty, especially uncertainty associated with parameters that are currently fixed in the model. Examination of genetic stock structure of flathead sole throughout its range and within the Gulf of Alaska and the Bering Sea is important for understanding whether spatial management units are properly allocated.

The reviewer suggests that developing an appropriate way to include ageing error is worth pursuing.

General Recommendations

- Some GOA fish stocks are already showing signs of climate change impacts. While there is good information on, for example, sea surface temperate, data for temperature-at-depth are much sparser but also highly relevant to many demersal fish stocks. Setting up additional fishery-dependent data collection could add considerably to future understanding in this area, especially given the wide geographic and temporal coverage of the fishery. It is, therefore, recommended that the collecting seabed temperature data from all tows by the groundfish fleet should be explored as a matter of urgency.
- Consider using detail, fine-scale commercial spatial location data to assist in defining the relationship between trawlable and untrawlable ground.
- Estimates of bycatch are based on simple ratios. More informative, statistical model-based methods exist to estimate bycatch, and these should be explored. See for example Anderson, O.F.; Edwards, C.T.T. (2018). Fish and invertebrate bycatch and discards in New Zealand arrow squid and scampi trawl fisheries from 2002–03 until 2015–16. New Zealand Aquatic Environment and Biodiversity Report No. 199. 135 p. https://fs.fish.govt.nz/Doc/24589/AEBR-2018-199-Bycatch-and-discards-Squid-and-Scampi.pdf.ashx

Appendix 1: Bibliography

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

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

Additional Papers

Anon. (2013). Report of the working group on methods for averageing surveys: Updated through 2013. Plan Team Draft, September 2013. 18p NPFMC Bering Sea, Aleutian Islands and Gulf of Alaska SAFE.

Cordue, P. L. (2007). A note on non-random error structure in trawl survey abundance indices. ICES Journal of Marine Science, 64: 1333 – 1337.

Helser, T.E., Benson, I., Erickson, J., Healy, J., Kastelle, C. and Short, J.A. (2019). A transformative approach to ageing fish otoliths using Fourier transform near infrared spectroscopy: a case study of eastern Bering Sea walleye pollock (*Gadus chalcogrammus*). Can. J. Fish. Aquat. Sci. 76: 780–789 (2019) dx.doi.org/10.1139/cjfas-2018-0112.

Kastelle, C.R., Anderl, D.M., Kimura, D.K. and Johnston, C.G. (2008). Age validation of Dover sole (*Microstomus pacificus*) by means of bomb radiocarbon. Fish. Bull. 106:375–385.

Mary Elizabeth Matta and Daniel K. Kimura (editors). Age determination manual of the Alaska Fisheries Science Center Age and Growth Program. NOAA Professional Paper NMFS 13 14p.

Orr, J. W., D. T. Drumm, E. A. Laman, D. E. Stevenson, and G. R. Hoff. (2014). Species identification confidence in the Gulf of Alaska and Aleutian Islands surveys (1980-2011). AFSC Processed Rep. 2014-01, 258 p. Alaska Fish. Sci. Cent., NOAA, Natl. Mar. Fish. Serv., 7600 Sand Point Way NE, Seattle WA 98115.

Punt, A E. (2015). Some insights into data weighting in integrated stock assessments. *Fisheries Research*. http://dx.doi.org/10.1016/j.fishres.2015.12.006.

Rooney, S., C. N. Rooper, E. Laman, K. Turner, D. Cooper, and Zimmermann, M. (2018). Model-based essential fish habitat definitions for Gulf of Alaska groundfish species. U.S. Dep. Commer., NOAA Tech. Memo. NMFS-AFSC-373, 370 p.

Somerton, D.A., Munro, P.T. and Weinberg, K.L. (2007). Whole-gear efficiency of a benthic survey trawl for flatfish. *Fish. Bull*. 105:278–291.

Stauffer, G. (compiler). (2004). NOAA Protocols for Groundfish Bottom Trawl Surveys of the Nation's Fishery Resources. U.S. Dep. Commerce, NOAA Tech. Memo. NMFS-F/SPO-65, 205 p.

von Szalay, P. G. (2015). The impact of sample size reduction on the precision of biomass estimates in the Gulf of Alaska. U.S. Dep. Commer., NOAA Tech. Memo. NMFS-AFSC-297, 17 p. doi:10.7289/V5FX77CX.

von Szalay, P. G., and Raring, N.W. (2018). Data Report: 2017 Gulf of Alaska bottom trawl survey. U.S. Dep. Commer., NOAA Tech. Memo. NMFS-AFSC-374, 260 p.

Appendix 2: Performance Work Statement

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

Gulf of Alaska 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.

Annex 2: Terms of Reference for the Peer Review

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.

Annex 3. Draft Agenda (agreed)

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

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

9:00-10:30	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

Friday, May 3, 2019

9:00-5:00	Address any remaining questions, writing time for reviewers, adjourn meeting

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

Panel Membership

Name	Role and Affiliation	
Jim Ianelli	Chair	NMFS, NOAA
Carey McGillliard	Stock Assessor	NMFS, NOAA
Wayne Palsson	Presenter – GOA BT Survey	NMFS, NOAA
Craig Faunce	Presenter – NP Observer Program	NMFS, NOAA
Jennifer Calahan	Presenter – NP Observer Program	NMFS, NOAA
Beth Matta	Presenter – Age & Growth Program	NMFS, NOAA
Patrick Cordue	CIE Reviewer	
Geoff Tingley	CIE Reviewer	
Kurt Trzcinski	CIE Reviewer	

Other pertinent information from the panel review meeting

Following discussions with the meeting Chair, the panel agreed to:

1) consider the Terms of Reference to include reviewing data inputs for all stocks; and

2) where the CIE Reviewers had common recommendations to draw these together into a brief Summary Panel Report. This summary report is presented as Appendix 4 (below).

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 timeseries 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 timeseries 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 prioritization 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.