Center for Independent Experts (CIE) Independent Peer Review of the Lingcod STAR Panel

Matthew D. Cieri

October 2021

Executive Summary

From July 12th, 2021 to July 16th, 2021, the STAR was convened to review the assessments for Northern and Southern Lingcod *Ophiodon elongatus*. Lingcod range from Mexico to Alaska along the US Pacific coast but for this assessment they were partitioned into two stocks: a northern stock from Cape Mendocino to the Canadian border, and a southern stock from south of Cape Mendocino to the Mexican border. As is typical for the region, both northern and southern stocks were modeled using Stock Synthesis. The northern model has a longer time series, more data inputs, and better sampling than in the south. Due to a change in the boundaries between the stocks, results from this assessment are not comparable to previous work, as portions of northern California are now within the northern stock area.

Data included typical elements including, age and length compositions, fishery-dependent and independent abundance indices, catches back to 1889, and fishery-dependent sampling for ages and lengths. Parameters also included were typical for the region, including natural mortalities, selectivities, estimations of growth, maturity, and others.

Results were fairly consistent between both stocks and were robust to the main uncertainties. Both stocks were above their management targets for spawning biomass but were differentiated by key resiliency parameters. Comparing the two stocks, the northern assessment was more data-rich than the south but was also more complicated and showed the most tension in data elements. Additionally, there were some major uncertainties with both models.

Both models showed some level of retrospective pattern, and both models had tension between data elements that increased the uncertainty. Both retrospective bias and model tension was greater in the north, likely due to the south having fewer data to inform these tensions or produce those retrospectives. An additional source of uncertainty for the north was the northern border of the stock with Canada.

Both proposed assessments provide the best available information for which management decisions could be drawn. However, several key uncertainties gave this reviewer pause in making such a statement. As a result of these uncertainties, it is recommended that both stocks should be treated as Category 2 stocks when setting quotas. Additionally, given the model's behavior and the uncertainties, it was recommended that the next examination of these models should be done as a full peer review. In this reviewer's opinion, these models are missing key aspects of the population dynamics for these stocks, but the proposed model was such a vast improvement over the 2017 assessment that reverting to it would have not been the correct course of action.

This report also contains several recommendations, specifically in improving these assessments and the overall STAR process, report writing, and meeting flow. Additionally, recommendations were made work to address issues of data tension in the models, stock boundaries between US and Canadian stocks, sex-specific selectivity, and natural mortality. These should be conducted prior to the next review, as further explorations and changes to the existing models were not possible in the allotted time.

It was an enjoyable experience. The STAT did an excellent job considering what they had to work with. While there were important uncertainties, these assessments pushed the field of Assessment Science forward toward a better understanding of the population dynamics of these stocks.

Background

From July 12th, 2021 to July 16th, 2021, the STAR was convened to review the assessments for Northern and Southern Lingcod *Ophiodon elongatus.* Lingcod range from Mexico to Alaska along the US Pacific coast. They have a long history of commercial and recreational exploitation dating back to at least the 1920s. For this assessment Lingcod was separated into two stocks, north of Cape Mendocino in California and south of Cape Mendocino referred to as the northern and southern stock respectively. The break between northern and southern stocks was previously drawn at the California- Oregon border in past assessments and represents a change in part driven by recent genetic investigations. The southern stock boundary stopped at the Mexican-US border, while the northern boundary stopped at the US-Canada border.

Lingcod are a faster growing, early-maturing species compared to other groundfish in this region. They display sex-specific differences in behavior, spatial distribution, and depth. Females tend to grow faster, reach a larger size, and are distributed in deeper waters when compared to males. Likewise, there is a latitudinal cline in important biological attributes, with individuals growing faster, living longer, and reaching maturity later, as latitude increases.

Model structure

As is typical for the region, both northern and southern stocks were modeled using Stock Synthesis. The northern model has a longer time series, more data inputs, and better sampling than in the south. Due to a change in the boundaries between the stocks, results from this assessment are not comparable to previous work, as portions of northern California are now within the northern stock area.

Data elements:

Important data inputs for the northern and southern stock, as well as the length of time available, can be found in Figures 1 and 2, respectively. Landings data from both stocks were reconstructed back to 1889 and are composed of both trawl and fixed gear. Additionally, a commercial trawl dependent index for both stocks was analyzed using the VAST (Vector-Autoregressive Spatio-Temporal) approach. The northern stock was composed of five fleets: commercial trawl, commercial fixed gear, and recreational for California, Oregon, and Washington. The southern stock had three fleets: commercial trawl, commercial fixed gear, and recreational. Discards were available for only the commercial trawl and fixed gear fisheries for both north and south.

Fishery dependent and independent indices were available for both north and south. As seen in Figures 1 and 2, the northern area had five fishery-dependent indices and two independent indices. The southern area had three dependent indices and three which were independent. Length and conditional age-at-length compositions were many of these indices in the north, but the only ages available for the south were from the West Coast Groundfish Bottom Trawl (WCGBT).

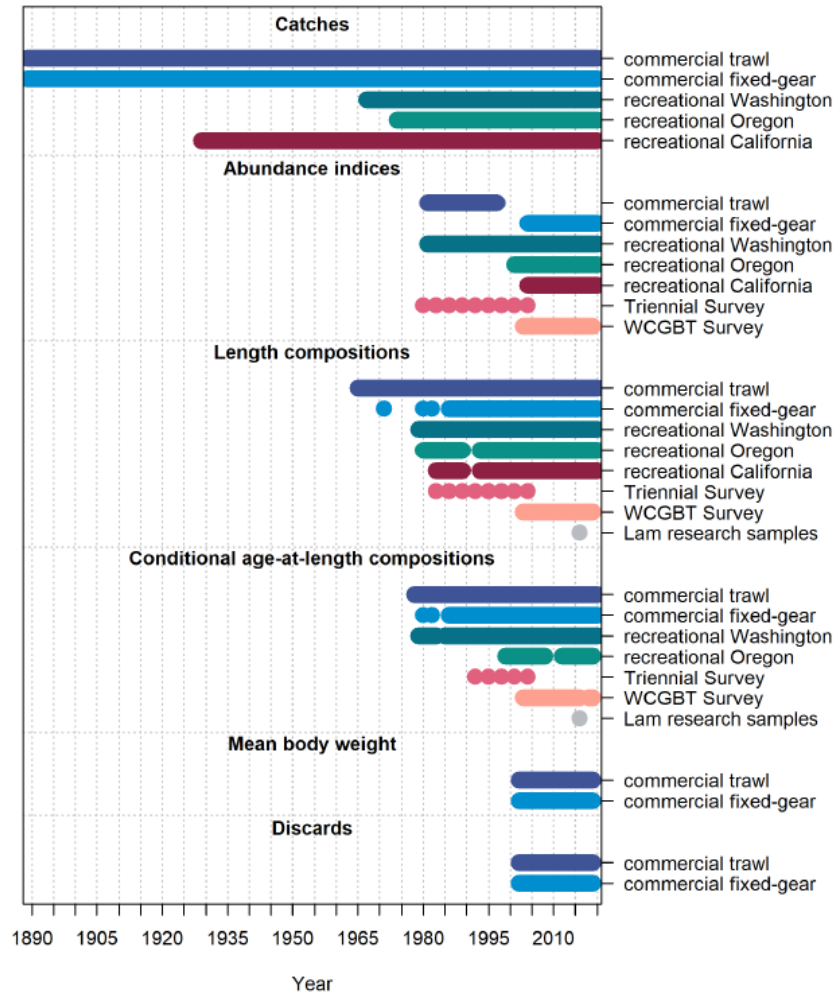

Figure 1: Data presence by year for each fleet and data type for the northern stock (Taylor et al., 2021).

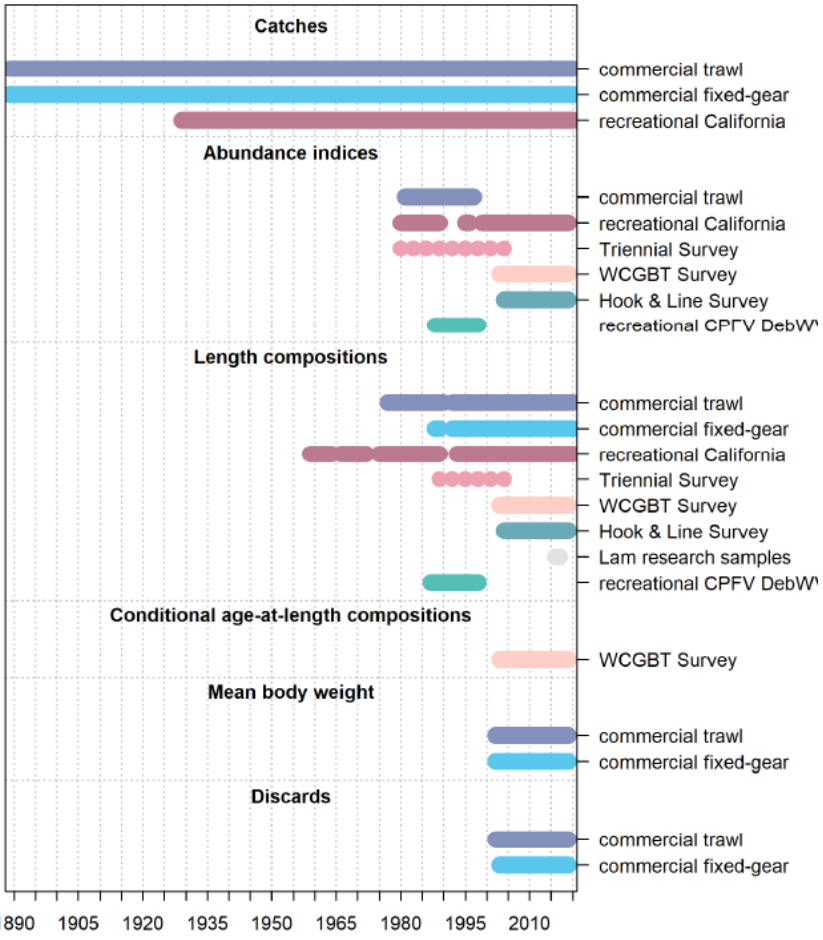

Figure 2: Data presence by year for each fleet and data type for the southern stock (Johnson et al., 2021).

Parameterization:

Specifications and structure of the northern and southern models can be found in Tables 1 and 2, respectively.

Table 1: Specifications and structure of the northern model (Taylor et al., 2021)

Section	Parameterization
Population characteristics	
Maximum age	25
Sexes	Females & males
Population bins	10-140 cm by 2 cm bins
Summary biomass (mt) age	3+
Number of areas	1
Number of seasons	1
Number of growth patterns	1
Data characteristics	
Start year	1889
Catch units	mt
Data length bins	10-130 cm by 2 cm bins
Data age bins	0-20 cm by 1 year
First age with positive maturity	1
First year of main recruitment deviations	1960
Fishing characteristics	
Fishing mortality (F) method	Hybrid F
commercial trawl selectivity	double normal
commercial fixed-gear selectivity	double normal
recreational Washington selectivity	double normal
recreational Oregon selectivity	double normal
recreational California selectivity	double normal
Triennial Survey selectivity	double normal
WCGBT Survey selectivity	double normal
Lam research samples selectivity	double normal

Table 2: Specifications and structure of the southern model (Johnson et al., 2021).

Section	Parameterization
Population characteristics	
Maximum age	25
Sexes	Females & males
Population bins	10-140 cm by 2 cm bins
Summary biomass (mt) age	3+
Number of areas	1
Number of seasons	1
Number of growth patterns	1
Data characteristics	
Start year	1889
Catch units	mt
Data length bins	10-130 cm by 2 cm bins
Data age bins	0-20 cm by 1 year
First age with positive maturity	1
First year of main recruitment deviations	1965
Fishing characteristics	
Fishing mortality (F) method	Hybrid F
commercial trawl selectivity	double normal
commercial fixed-gear selectivity	double normal
recreational California selectivity	double normal
Triennial Survey selectivity	double normal
WCGBT Survey selectivity	double normal
Hook & Line Survey selectivity	double normal
Lam research samples selectivity	double normal
recreational CPFV DebWV selectivity	double normal

There wasn't much difference between the two models other than the inputs used, and the start year of the recruitment deviations.

Sensitivities and diagnostics

The STAT used three diagnostic tests for both northern and southern models, including likelihood profiles, a jitter function to determine if the model was at its best solution, and correlations to Bayesian posteriors with MCMC. Some of the selectivity parameters were strongly autocorrelated and were then fixed to the base models.

Numerous sensitivity tests were conducted for both models to examine model behavior and to explore alternative states of nature. For the north these included: using a shared M across the sexes, setting steepness to 0.07 and allowing M to be estimated, setting M to 0.3, and setting steepness to 0.7 and two levels of R0 0.8 and 0.4. Both north and south tested sensitivity to composition data by using Dirichlet-multinomial weights, dropping unsexed fish from the models, excluding all ages from the models, and combining males and females. Both north and south used a "drop one" approach to investigating sensitivity to surveys, whereupon each of the indices was dropped one at a time and the model was rerun. Additionally, a retrospective analysis was performed for both stocks, where the terminal year is dropped sequentially to examine the model for potential biases.

During the Panel discussion, a total of 20 requests were made to help further examine the models and to gauge their performance and behavior.

1) Explore freeing up the width parameter on sex-specific selectivity curves, including offsets, for both north trawl and south fixed gear.), turn off the early recruitment deviations for a 3rd run if time allows.

Perform a model run, for the status quo Base model only, for the north north model.

Results

Stock status results for both assessments can be found in their respective assessment reports. Overall, both stocks appear to be quite healthy, with stock sizes above the 40% fraction unfished the managers use as a target SSB. This status appears to be fairly robust, with few of the sensitivities pushing the stock outside its confidence intervals. That said, there are major sources of uncertainty for both models that require attention as outlined below.

Conclusions

The base models represented the best information to provide advice for each of these stocks. There were, however, a number of uncertainties that made this a difficult choice and required lengthy discussion. Even after the affirmation of the base models was made, there are many misgivings about how these models are performing. Given these issues 1) both stocks should be treated as a Category 2 stock (PFMC, 2020), and 2) that this model should go back for a full review, rather than be simply updated in the next cycle.

That is not to say that the models, as proposed by the STAT, lacked merits; quite the contrary. The STAT did an excellent job of providing multiple diagnostics and sensitivity analyses to examine model behavior. Overall, both models were data-rich, with the north being richer than the south. There are

numerous indices and generally decent data coverage for both stocks from which modeling can commence. Additionally, the treatment of the data itself appeared to be methodical, and well within the standards of this region. In general, the models fit the available data well, suggesting that future projections are well-grounded.

There were also many improvements over the 2017 STAR model.

- Incorporation of ages into the Northern model as CAAL rather than marginal
- Work done to change the boundary between northern and southern stock based on genetic analysis
- Improvement to historical catch estimates (both commercial and recreational)
- Estimation of key parameters, particularly steepness and natural mortality rates

Both models are very sensitive to both steepness and natural mortality, as expected. Also as expected the model was sensitive to changes in structure. While there are some retrospective issues in the models, they appeared to be within the bounds of the model's uncertainty when translated into depletion. They did, however, still added to the overall uncertainty.

More concerning, however, is the tension between data elements in the likelihood profiles. Length data for multiple data sources indicated a model preference for lower natural mortality, while length data, and to a smaller extent abundance data, suggesting higher natural mortalities. This can be concerning as any changes in weights to those data elements can radically change how these models see the biology of the stock; rapidly changing the productivity and resiliency of the stock with only minor changes in data weights. This tension was less apparent in the south, as there were fewer age data to inform it.

There is also the issue of differences between northern and southern stocks. Both models estimated very different natural mortalities than the 0.3 initial prior for females: with the north estimating 0.41 and the south estimating M at 0.17. Similarly, the northern model estimated steepness at 0.80 while the southern model estimated a steepness of 0.51. Even the selectivities for the continuous surveys between the two regions were vastly different. These estimates would make it seem that each model was deriving its results from different species, rather than two stocks from the same species.

The issue of stock boundaries is also a concern. While there is good information and recent studies to suggest the change in the border between northern and southern stocks, there is little in the way of information regarding the northern border for the northern stock, with Canada. From the assessment reports and presentations, it is clear that lingcod are available in Canada and beyond to Alaskan waters, the Northern border stops at the Canadian border. Efforts made to contact the Department of Fisheries and Oceans (DFO) Canada, which has assessed lingcod in Canadian waters in the past, were ongoing. That said, increased coordination with DFO is an important recommendation made elsewhere in this report.

The overall conclusion that can be drawn from this is that both models are missing key aspects of the dynamics for both areas. While there are some indications of why this might be, but all of that is mostly conjecture at this point, and it’s unlikely a short-term solution can be found. These models were an improvement over the 2017 assessments and given the level of the stocks relative to management targets. This reviewer agreed that the proposed base runs were appropriate for providing management advice in the short term, after minor alteration.

Description of the Individual Reviewer's Role

This reviewer's role was to read all provided materials on the 2021 assessment for lingcod, engage in the online discussion of the strengths and weaknesses of the proposed approach, and were possible to provide suggestions on improvement of the models. Additionally, this reviewer assisted in writing the draft Panel report for presentation to fishery managers, the public, and SSC. A full write-up, this report, is also part of the terms of reverence, where this reviewer makes an independent assessment outside of review Panel consensus, highlighting their own unique opinion of the assessment.

Summary of Findings

1. Become familiar with the draft stock assessment documents, data inputs, and analytical models along with other pertinent information (e.g., previous assessments and STAR panel report when available) prior to review panel meeting.

As seen in the background section, this reviewer received and read all of the documents for this review prior to the start of the meeting. Further reading and preparation were also required to refresh knowledge of this region's management, in particular to the Category system used by the Science and Statistical Committee to provide management advice.

Overall, the documents provided were pertinent, though it must be said that the Pre-STAR meeting assessment documents lacked key elements as more than a few analyses were only available after the documents were released (but prior to the meeting). As outlined in the Recommendations section (below) this lack of completeness of the assessment report was frustrating. It was very clear that there simply wasn't enough time prior to the reports deadline to ensure that those documents were complete and well-edited.

2. Discuss the technical merits and deficiencies of the input data and analytical methods during the open review panel meeting.

The reviewer attended and participated in the meeting, discussing many aspects of the model's behavior as well as making some requests to the STAT during the process.

This reviewer notes that both models have relatively rich data sets for both fishery-independent and dependent data available. Age information was more extensive for the northern stock than the southern, but overall, most data seemed relatively comprehensive. Likewise, data treatment by the STAT seemed proper, and well done given the amount of data involved.

However, there were some discrepancies with the recreational and commercial data noted during discussions, particularly for the Washington removals, as these didn't seem to fit with other participants' expectations. These inconsistencies were rectified, but others became apparent and were not able to be resolved. One of these issues was that it was difficult to distinguish dead discards and kept fish in the California recreational data. Additionally, there was a poor fit of the northern model dead discards to the dead discards in the trawl fishery from 2002-2010 found in the GEMM product that was also unresolved.

Large improvements were made in these most recent models when compared to 2017. Notably

- Incorporation of ages into the Northern model as CAAL rather than marginal
- Work done to change the boundary between northern and southern stock based on genetic analysis
- Improvement to historical catch estimates (both commercial and recreational)
- Estimation of key parameters, particularly steepness and natural mortality rates

Both assessments were very thorough when it came to diagnostics, sensitivities, and complementary analysis. A key strength was that the surveys appeared to be very informative.

There were however some difficulties and deficiencies. Chief among these was the tension in the model's sources. Age data, particularly in the north, seemed to conflict with length data when estimating natural mortality. While this conflict doesn't appear to the same degree in the southern model, this is likely more the result of a lack of age data, rather than better model behavior which drives this lack of conflict. This reviewer wonders if this conflict in data sources might be prevalent in the south but is just masked by a lack of aging data.

The inconstancies between northern and southern stocks are also an uncertainty. In addition to differences in the estimated natural mortality, the survey and fishery selectivities were also different. This result is surprising and the same survey is used for both north and south, and the fishery generally seems to operate similarly in both areas This lack of consistency, while possible biologically, suggest that each model maybe highlighting different aspects of the data inputs to arrive at different conclusions about the biology as well as fleet behavior.

A further uncertainty is the northern boundary for the northern stock. Currently, that boundary is marked at the US-Canadian border, with Lingcod north of that being assessed by the Canadian Department of Fisheries and Oceans, although not recently. However, during the Panel meeting and in the assessment report data from surveys suggested that there are Lingcod found in southern Alaska continuously. As such the connectedness between US and Canadian stocks is an uncertainty worthy of a more thorough investigation.

3. Evaluate model assumptions, estimates, and major sources of uncertainty.

Model assumptions appeared to be in line with best practices and the models used appeared appropriate. One uncertainty, as mentioned elsewhere but reiterated here is the assumption on the unit stock. While the border between Northern and Southern stocks appears robust given the information presented, the border between the US northern stock and the stock centered in Canada is less clear and a major source of uncertainty.

Both models estimated very different natural mortalities than the 0.3 initial prior for females: with the north estimating 0.41 and the south estimating M at 0.23. Similarly, the northern model estimated steepness at 0.80 while the southern model estimated a steepness of 0.51. This suggests very different biology when comparing northern and southern stock of the same species.

Stock status results appeared to be robust, with few of the sensitivities pushing the terminal results outside of the base model's confidence limits in either North or South stocks. Likewise fits to the data were reasonable in both stocks, with no major issues detected. While there was a slight retrospective uncertainty in the models, none of those were outside the confidence intervals provided by either

model when translated into depletion; and therefore, did not affect stock status. As such, both model's estimates of stock status appeared robust with any uncertainty unlikely to change the status outcome.

Overall, the STAT did the best it could with the data and models in hand, given the time constraints. However, there are numerous uncertainties and unresolved issues. Most important among those are the tension between data elements pushing the stocks to high or low productivity states of nature, the lack of consistency between two adjacent stocks of the same species in fundamental biological characteristics, and the lack of data to support the northern border for management and stock purposes.

Of those, the reviewer finds the tension between data elements the most vexing; with length data suggesting higher natural mortality when compared to aging data (see Figure 3 below). This amount of tension between elements is quite surprising, given that both age and length data are coming from the same sources, surveys, and data monitoring programs. Due to the model's overall sensitivity to natural mortality, small changes in data weightings can have a very large effect on stock productivity, status, and resiliency. In short, you get a much different picture of the stock with slight changes to those weights. Which makes for a disconcerting result.

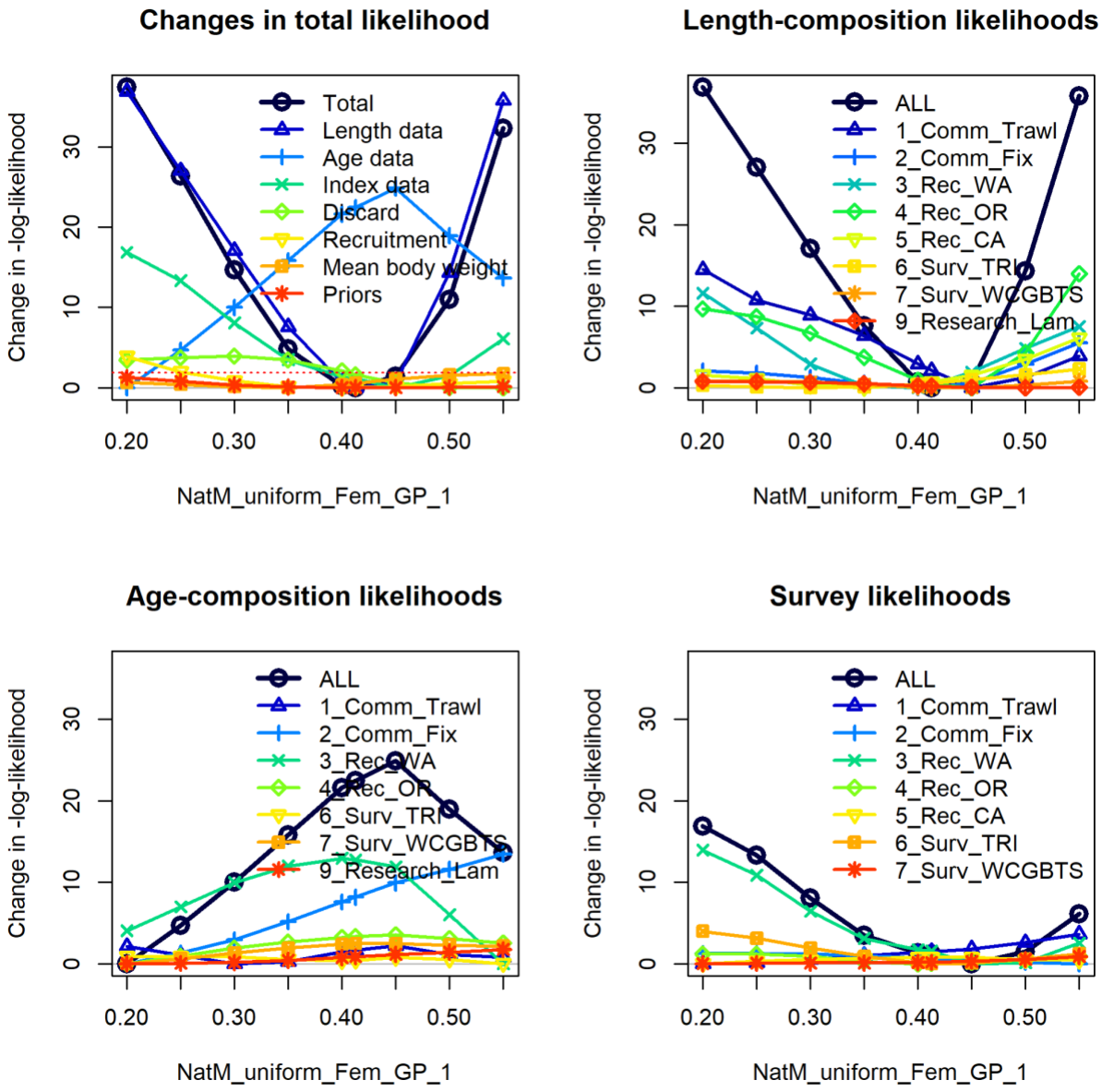

Figure 3: Change in the negative log-likelihood across a range of female M values. From the northern assessment report (Taylor, 2021; Figure 160).

Further, response to the request by the Panel highlighted that the model, at least in the north, did not support the prior for natural mortality; but rather preferred either of the two extremes (see Figure above). This gives the illusion of two stable states of nature for this stock, high productivity and resilience, and low productivity with lower resilience. A fundamental uncertainty, given that even slight differences in weight data elements can result in dramatically different pictures of the stocks productivity/resilience, and therefore the amount of appropriate removals.

4. Provide constructive suggestions for current improvements if technical deficiencies or major sources of uncertainty are identified.

During the discussion, a total of 20 requests/suggestions were made to help further examine the models and to gauge their performance and behavior. Those requests, rationale, summarized STAT response, and the Panels responses are provided below. Note that these have been summarized where responses or conclusions were rather long. Asterisks were requests/suggestions by this reviewer in whole or in part.

2) Explore freeing up the width parameter on sex-specific selectivity curves, including offsets, for both North trawl and south fixed gear. Do not have this time-varying. This is for both North and South. Provide likelihood profiles on M for females 0.18 to 0.58 in steps of 0.05.
   Rationale: The Panel noted that there is tension between age and length. There is a depth gradient by sex that would suggest sex-specific selectivity. It may help understand the tension between length and age composition data.
   Response: The revised analysis did not lead to large differences in selectivity curves. However, allowing sex-specific offsets in selectivity did somewhat relax the tension between the age and length compositions in their influence on estimating M, as reflected in a flatter likelihood profile on the age data. Yet the model also produces some biologically questionable results: a maximum selectivity for females in the commercial fleet that is only 20% that of males, and a natural mortality rate for males that was ⅔ greater than females in the North. This modeling option also led to a very strong influence of the Oregon recreational data, which could not be easily explained.
   Conclusion: It was concluded that pursuing sex-specific selectivity offsets was a valuable avenue to further explore how to resolve the tension between data sources in the model (particularly in the North) but that this option as implemented here could not be supported as an addition to the base model because of the problematic.

3) Show the model run and the fit to the DWV index, when that index has no influence on the model. Compare that to the fit that includes the index. Provide model comparison figures. This is only for the southern stock.
   Rationale: The DWV index had unusually high sensitivity without a clear reason why
   Response: Removing either the index alone or both index and associated length data had similar effects, which was to produce a current estimate of stock status that is greater than B0. With those data removed, fits to other indices did not change substantially, while the projected fit to the DWV index was much flatter and did not capture a sharp decline in the late 1980s that is present in that index.
   Conclusion: Removing the dataset led to substantial increases in the current estimate of stock status. This appears to be due to a signal of a sharply declining stock in the late 1980s/early 1990s that does not appear in other datasets from that period. The influence of that period is apparent when the DWV index is excluded from the likelihood calculation; the revised model fits the later portion of the DWV index well but does not track the late 1980s decline. the spatial coverage and representativeness of the DWV index (as described in Monk et al. 2016 technical report) was discussed and it was concluded that there was a robust dataset that should remain in the model and that its high influence on the outcome was not likely to be spurious.

4) Run the southern model with the main recruitment deviations starting in 1955. Produce model comparison results and dynamic B0 for both models. If the early recruitment deviations are unusually too high (comparable to what we see in the base model at present, turn off the early recruitment deviations for a 3rd run if time allows.
   Rationale: There is an unexplained and implausible run-up of recdevs in the pre-data period that may be a model artifact.
   Response: Starting main recruitment deviations in 1955 did not produce a large change to the model fit, nor did it eliminate the questionable early deviations. Eliminating early recruitment deviations did not produce a large change in later model fit but did produce a higher estimate of B0. As an additional analysis, completely excluding the recreational length data from 1959-1972 did eliminate the questionable recruitment deviations from the 1950s and did not lead to a large change in the current estimated stock status.
   Conclusion: Changes to the way recruitment deviations were handled did not resolve the concern about unusual early recruitment events and excursions well above B0. Removing pre-1975 length composition data did eliminate the questionable recruitment patterns in the 1950s but it is undesirable to completely exclude those data. Those pre-1975 recreational length data were entirely from the Monterey Bay area, so an alternative solution is to add a selectivity block for that time period for that fleet. This will be the rationale for Request 6, below.

5) *Provide removals (i.e., total dead discards and total retained catch) by fleet and disposition by year for both North and South. Stat discretion on display of data-figure or tables.
   Rationale: This is not currently displayed in the draft reports as a single informational graphic.
   Response: The figures have been provided. Of note in these data is a steady decline in landings in the trawl fleets in both regions in the late 1990s and early 2000s, as well as an extremely high (~90%) discard rate in the North. This likely reflects management changes during that period when the stock was in a rebuilding phase.
   Conclusion: The immediate onset of very high discards in the 1990s, coupled with the absence of discards (e.g., of smaller fish) in prior years is very suspect. The model is likely overestimating the discard mortality in that period. There is not a clear path to resolving this issue for the current assessment, but future assessments should explore a reconstruction of historical (pre-1999) discards or consider allowing a ramp-up of discard rates to the 1999 period. It would be valuable for future versions of Stock Synthesis to allow separate weighting of discarded and retained catches.

6) Update the northern model to reflect the error in the WA recreational mortality estimates. Provide a figure of WA dead discard catch and retained recreational catch.
   Rationale: An error was noted in the report regarding the WA recreational mortality estimates.
   Response: This error did increase the estimate of WA recreational landings by a considerable amount (50+ tons a year in many years), so this was an important correction. A figure of the relative landed catch compared to estimated dead discards suggested that the total dead discards represent a fairly modest fraction of the total recreational catch. The corrected catch model trajectory resulted in very little change in the overall model result.

Conclusion: There was agreement that this change was to be made and reflects the new base North model.

7) Add time-varying selectivity for the CA rec fishery in the South model. Add a block for 1959 to 1972. Provide the model results (biomass, recruitments, and recdevs) and the estimated selectivity curves. Provide information on how the fit to the data improved.
Rationale: STAR panel understands the 1959 to 1972 data encompasses a small amount of CA coastline and may not be representative of the entire area
Response: The STAT divided the early period of the fishery into two blocks, one from the model start to 1972, and a second block from 1973-1982 (there was an existing time block beginning in 1982). Adding the additional block had very little effect on the model results.
Conclusion: Removing these from the model seemed to be the most satisfying solution, recognizing that future research would be appropriate to compare the regional differences between more contemporary data with the region in which these historical data were collected, or to otherwise to discern if there could be a more appropriate way to use the data in future models. This would become the new South base model (pending request #9).

8) *Perform a model run, for the status quo Base model only, for the North without the fishery age data, and rerun with Dirichlet Multinomial (DM) weighting. Provide initial results and comparison of Francis vs Dirichlet weights.
Rationale: Dirichlet weighting seemed to resolve some of the tension between age and length data and estimated the natural mortality rate more comparable to the South model.
Response: The change in weighting did not appear to be as great as expected. The DM weighting did lead to fully weighting most age composition data, which led to a scaling down of the model spawning output and a slight decrease in the estimate of M. Without fishery-dependent age data biomass scales up (with or without DM) and the estimated M is smaller still (from 0.41 in the base to 0.37 to the base with DM, to 0.27 to the base with DM and no fishery age data). Without age data and with DM weighting, a very strong year class shows up in 1962 that appeared to be potentially spurious, suggesting that age data were not the only data creating interesting signals in recruitment. The natural mortality estimate in the northern model declined substantially without age data
Conclusion: The take-home message from this analysis seems to be that there will continue to be tension between the length and the ages at this stage in the model. The DM method puts much more weight on both length and age relative to the indices. There was general agreement that, at this stage, recommending a switch to DM weighting was infeasible.

9) A) Remove all fixed gear (FG) age composition data. B) Remove fixed gear age composition data for the years 1999-2011. C) Do A and B with and without offsets for sex-based selectivity for all fleets. Requests are only for the model. Provide the model comparisons slides and table(s) with likelihoods and key parameter outputs, selectivity outputs, etc. The STAT will include the updated results from Day 2, Request 5.
Rationale: These model runs are requested to better understand whether management changes or unbalanced periods of age sampling data in the associated fisheries are contributing disproportionally to tension in the model. There is an observed disconnect between the fixed

gear length and Conditional Age-at-Length (CAAL) in the 2000s that are more likely to reflect fleet behavior.
Response: These analyses were all undertaken with the new north base model. Removing the FG ages of the years 1999-2011 made no difference and results were similar to the base case. However, removing all the FG ages had a large impact on the results by increasing the spawning biomass (although there was not a large impact on the fraction of unfished trends relative to the new base model). These results highlight that the source of tension in the model is not from the period 1999-2011 but more likely the earlier data where there are fewer alternative sources of information.
Conclusion: The implementation of the female offset could be improved and further refined. The data suggests that sex-based selectivity may be occurring, but the model runs with the offset included do not provide results that are entirely feasible. Although not possible to suggest for this assessment, including smaller unsexed fish into the CAAL data could improve the potential conflict between growth, M, and selectivity by providing more information on the growth of the smaller length and ages.

10) Conduct additional tuning and exploration of the southern model without the 1959-1972 length composition data to see if a reasonable new base model can be developed. Provide the model comparisons slides and table(s) with likelihoods and key parameter outputs, selectivity outputs, etc.
Rationale: Conduct additional tuning and exploration of the southern model without the 1959-1972 length composition data to see if a reasonable new base model can be developed.
Response: The results of the initial test of the Southern model without the early length composition data tuned and untuned were provided. The tuned model with no early lengths required adjustment to the mean recruitment deviation period and redoing the Francis weighting (as expected). Poor behavior of the Triennial survey's descending selectivity parameter was resolved by adding additional error to the Triennial index. The tuned model affected the overall trend and the fit to the Triennial survey index. The uncertainty of the "All lengths" (e.g., including the 1959-1972 length data) model is smaller than the runs without the early lengths. The early period is quite uncertain for the new runs, but more similar to the base case in recent years. The fraction of unfished in 2020 of the re-tuned model is at the management target, but this is lower than the other two runs. Removing the early length data resolved the issue of the large autocorrelations in the recruitment deviations. In terms of likelihood components, this model fits the WCGBTS index better than the draft base case which, in terms of lingcod, is a more reliable index due to its more consistent spatial coverage. None of the models fit the last year of the WCGBTS well. All three runs have similar steepness values, but the M values are markedly different with the new tuned model being lower.
Conclusion: It was recommended that the tuned version of the model that removes the length composition data from 1959-1972 should be the new base case in light of the above findings. The STAT agreed. Additional exploration of the historical length-frequency data should be a research need for future efforts

11) *Perform runs for both North and South base models with the Lorenzen mortality function. The reference age for the naturality mortality prior should be age eight (8). Provide the model

comparisons slides and tables/figures with likelihoods and key parameter outputs, selectivity outputs, etc. for review.
Rationale: Allowing for a variable natural mortality with age for fish less than age 8 would be a more realistic estimate of natural morality for smaller and immature fish. It is well documented that natural mortality tends to be negatively correlated with body size.
Response: Runs were completed as suggested.
Conclusion: Overall, the results indicated that using the Lorenzen function did not dramatically change the trajectory of the SSB for either Northern or Southern stocks and had less effect than the use of a sex selectivity offset. It was noted the South model was very sensitive and gave unrealistic results when changing the reference age of the mortality prior.

12) Perform runs for both the North and South base models in which sex-specific selectivity offsets are implemented for fisheries but not surveys. Explore other parameterizations of the selectivity offset as time might allow. Provide the model comparisons slides and tables/figures with likelihoods and key parameter outputs, selectivity outputs, etc. for review.
Rationale: The STAT has observed that the selectivity differences appear to be greater for the fisheries data relative to the survey data and has been considering alternative forms of implementing the selectivity offsets.
Response: Overall, the model showed a lower scale in SSB after approximately 1940 and a lower stock size relative to management reference points for the North. For the South, the difference in scale was prior to the mid-1970s, but also showed a lower stock size relative to the reference points. While the selectivity pattern has improved and the fit to the data improved substantially, the STAT was not convinced that the results led to an improved base model, particularly as the fits to the indices degraded.
Conclusion: It was noted that while this change improved the overall fit in the North, it worsened it in the South. Likewise, while there was an improvement in the overall fit and the fits to both length and age comps, it came at the expense of the survey fits to the abundance. The selectivities in both the North and the South were better, but still rather implausible. While this change improved the model, it was only a partial improvement. Given this, it Request 12 (below) was suggested as a way of more fully exploring the sex selectivity difference between males and females in the model.

13) Explore freeing up the peak and the descending limbs of the offset selectivity curves for lingcod in both the northern and southern models, including the base model and the model from #11. Try this parameter by parameter, or in a combination deemed reasonable by the STAT. Provide model comparison slides.
Rationale: Adding sex-based selectivity offsets to the model improved some aspects of model fit but still produced implausible selectivity curves (among other issues). It is possible that allowing more latitude in fitting the selectivity curves would improve model behavior.
Response: The general outcome was that models allowing sex-specific differences in selectivity in one or more fleets led to much lower estimates of M than the base model, while non-sex-specific models produced M estimates more similar to the base model. Some of the estimated selectivities had a knife-edge right-hand side, which seems implausible. The most complex model considered had sex-dependent variation in both the peak and descending limb of the

selectivity curves, and it did produce selectivity curves that could be consistent with a fixed-gear targeting of smaller inshore males. No alternative model produced large differences in estimated stock status relative to management targets.
Conclusion: These exploratory analyses confirm that the data support some form of sex-dependent selectivity in the fisheries. Yet the attempts to model that feature lead to a trade-off between estimating selectivity and M, such that either both M and selectivity are estimated to be low (fish live long but are not caught) or high (fish are highly selected but die before they are caught). Without further research and likely a different approach to handling fleet structure and age data (see research recommendations) there is not an acceptable way to capture these dynamics with the current model structure.

14) Provide a likelihood profile for female M for the model that the STAT deems most plausible based on the result from request 12. M values should range from 0.1 to 0.5.
Rationale: As both base models have changed, and both continue to be highly sensitive to changes in M, this will provide the opportunity to re-evaluate sources of tension in the model.
Response: The attempt to produce this profile was not entirely successful because the model had difficulty converging at low values on M. There is some evidence for a bimodal likelihood surface with negative-log-likelihood minima near both 0.2 and 0.5.
Conclusion: This analysis appears to confirm the suspicion based on the results from Request 12 that the data as currently structured favor two alternative modes in parameter space; one with low M and one with high M. Unfortunately, neither M falls within the range that would be expected for lingcod life history.

15) Investigate the discrepancies in catches between the Groundfish Expanded Multiyear Mortality (GEMM) reports and the northern model estimates. Explore modeling methods to attempt to improve the fit to the dead discard rates in the model to produce dead discard estimates that are more in line with the GEMM values. This adjustment would consider these earlier years (1998-2001) as well. This investigation should focus on the trawl fleet.
Rationale: There are large discrepancies in catches between the GEMM reports and the model estimates. WCGOP discard rates in the model are from the same data that produce the GEMM estimates, hence we would expect the model estimates to have coherence with the GEMM estimates.
Response: Several options to address this issue were considered: two different adjustments to reduce the standard error associated with the discard rate data (to encourage closer fits to that data stream) and an artificial deflation of the discard sample size to reduce its weighting in the likelihood calculation.
Conclusion: Because the changes to the overall model fit are small and the considered alternative solutions are very ad-hoc, the base model should not be changed to include these options. However, this issue is a future research need.

16) For the North model, provide a run in which the standard deviation on the prior on female natural mortality is divided by 2, and do this for both the base model and for the model with sex-specific selectivity (model 417). Present model comparisons and associated informative tables/figures as in prior requests. Provide the whole suite of r4ss files for review.

Rationale: There continues to be considerable tension between model structure and the dynamics of the population, with model estimates of M seeming to vary widely in response to relatively modest structural changes in the model.
Response: The results clearly indicated that a more informative prior resulted in very little change to the model result, as the improvement in fit with a substantially higher M is far greater than the constraints on the prior. In other words, the data are pushing the model to a higher value of M despite the prior, and the result was only a trivial change in M.
Conclusion: There was also general agreement that there seem to be two alternative local minima in the negative log-likelihood surface suggested by the data, associated with low and high natural mortality rates. The results also indicate that the management quantities do not change that much among those two states. There was general agreement that although these explorations have allowed a better understand the model, it was mostly learned that there is a lot more to do in order to better fit the data and come up with a model that appropriately captures the complex sex-specific dynamics of this population.

17) *For the north model, provide a run in which female M is fixed at 0.3. Present model comparisons and associated informative tables/figures as in prior requests.
Rationale: The tighter uncertainty bounds on the prior did not influence the model result, but rather reinforced the notion that the model may have two stable states.
Response: The model result was considerably more pessimistic regarding stock status than the base model or the models with alternative selectivity functions, as had been initially suggested by the likelihood profiles in the draft document, despite a relatively modest degradation in overall fit. Fits to age data improved fits to both index and length data degraded. There was a general discomfort with this approach
Conclusion: There was general agreement that what is happening here is a bit more than just sensitivity to M and that fixing M at the prior estimate is forcing the model to a solution more contrary to the data than the base model or other alternative models. The core challenge seemed to be that there are conflicts between two types of errors- observation error in the data (leading to high tension between length and age data) as well as structural error in the model dynamics. Specifically, the inability to implement a reasonable sex-based selectivity function in the model when there is strong evidence for differences within and among fleets in selectivity by sex suggested in the data presents a key challenge in adequately capturing the dynamics of the population.

18) *Provide a retrospective analysis that goes back 5 years for the northern and southern models. Report the Mohn's rho values (Woods Hole and Alaskan/Hurtado-Ferro).
Rationale: There are significant scaling issues associated with the retrospectives that merit a closer investigation
Response: During this process, it was noticed that the approach of removing a single year's data is too simplistic given that these models have different time blocks and fixed parameters that could differentially interact with the retrospective analyses. As an example, the WA recreational fishery had a 22-inch min size limit lifted in 2017 resulting in a selectivity block in the model that would impact the N model retrospective analyses. Furthermore, typical of many assessments, the estimate of the terminal year recruitment is the most uncertain in these assessments. It

would be more appropriate to address the time blocks and therefore slightly change the model in each peel. The results produced were a first step in this direction of better characterizing the effect of 1 year's less data in the model, but this approach is only a first step rather than a final solution to the problem. It is also important to note that the retrospectives are not re-tuned models. Not retuning the model may overemphasize model sensitivity to the peels.
Conclusion: In general, the retrospectives of both assessments are showing that the recent years' data are providing information about scaling, but that there is great uncertainty in scaling the resource when these data are removed. This has already been shown through the different sensitivity tests. However, there is also some degree of stability in the fraction unfished reflecting that there is more information in the relative change in biomass and recruitment over time.

19) Provide the r4ss files for the revised northern and southern Base models.
Rationale: This is a simple diagnostic check.
Response: These were provided.
Conclusion: There was some discussion about the differences in the selectivity in the surveys between the North and South final models. This reflects that the size composition data are very different between the two regions. In the Triennial survey in the N, there are not as many small animals as in the South. It was noted that the Triennial surveys do not cover the full coast of CA. However, there is a difference in the bathymetry between the N and S regions that could be contributing to the difference in the survey data and therefore causing different interactions between the surveys, the fisheries, and the distribution of the resource. This reflects that there are spatial scale processes that are not fully captured in these assessments.

20) For the southern model, develop the runs that would fill in a decision table based on the high and low quantiles of M as inferred by the likelihood profile. Provide diagnostic outputs as appropriate.
Rationale: This is likely to be a reasonable basis for the decision table for the southern model.
Response: There were convergence issues when diagnostics (such as the jitters and profile likelihood) were run in the base model as agreed on day 4 of the meeting. This meant that further changes had to be made to the southern model. This included changing some of the bounds and estimating parameters that were fixed for the triennial selectivity parameters. This resulted in minor changes to the model but avoided these convergence issues. This was proposed as the new base case.
Conclusion: This change was supported as the final base case

21) *For the northern model, develop model runs that might encompass the different types of both observational and structural uncertainty by 1) excluding fishery-dependent age data from a model and, 2) running a model with sex-specific selectivity (as in model 420) to capture the "process" uncertainty. Include in the comparison plots and tables the model run in which M is fixed at 0.3 for females.
Rationale: Model sensitivities without the age data had the best fit to length data and indices, while the model with sex-specific selectivity implemented as freeing up descending limbs of the

selectivity curve had the best overall fit to the data. The results of each also bracket a wide range of possible model outcomes.

Response: Four options were provided for the northern model decision table to describe uncertainty – the final model, the final model without ages (representing the high stock state), the model with sex-specific selectivity (which represents the low productivity alternative state and structural uncertainty) and the fixed M=0.3 model (the low stock state). There was some concern expressed regarding the M=0.3 run in that the likelihood of this model compared to the base model is almost 17 units above the base model and does not fit the data well.

Conclusion: These are two axes of uncertainty reflected therein. The first emphasizes data uncertainty by reflecting the tension within the composition data that produces uncertainty with regard to the scale and productivity of the resource. When the age data are excluded, the virgin SSB is increased, but the natural mortality is considerably lower than the base model (0.33 for females, 0.29 for males). The results associated with excluding the age data from the final model are a good reflection of this axis of uncertainty. The second source of uncertainty is structural uncertainty which here is demonstrated by the inclusion of sex-specific selectivity. The possibility of sex-specific selectivity is likely related to processes associated with different movement patterns of males and females, or of the different habitat types that are more or less available to the trawl fleet, for example. In this configuration, the SSB is higher than the final model, but natural mortality is very low – similar to the south model.

After a multitude of requests and changes to the base models examined during discussions, the base models were selected as the best performing after some minor alterations. This reviewer provided a number of suggestions during the course of the Panel discussions. However, it is noted that almost none of the suggestions made addressed the core issues given the model's behavior. Given the level of effort required to implement changes, and the time available, further alteration of the models to address these issues was just not possible. While the STAT did an excellent job of fulfilling those requests, further change to the model had to be done outside this peer review, as they require further in-depth exploration and in some cases data collection.

As such no further improvements are recommended for this review. Rather, work on issues of data tension in the models, stock boundaries between US and Canadian stocks, sex-specific selectivity, and natural mortality should be conducted prior to the next review, as stated in TOR 6.

5. Determine whether the science reviewed is considered to be the best scientific information available.

The base models as formulated for both Northern and Southern lingcod represented the best available information and analysis to provide managers with reasonable catch advice.

There are, however, two caveats. First, despite the level of modeling and data, both northern and southern models should be as Category 2 stocks. This was in part due to the retrospective pattern seen, as well as the general uncertainty in how the models scale the stock. Second both assessments should not be updated but rather go through a full peer review when considered again. This was because of the uncertainty surrounding this model, in particular; the bounds around the stock size to the north, the

retrospective pattern, and the overall tension in the model with regards to the data elements, as mentioned elsewhere in this report.

This is not to say that there was an overall approval of either model reviewed here without some trepidation. It might have been preferable to state that this model was not appropriate for management use, given the uncertainties and, in particular, the tension between length and age data in the model. However, during such considerations, available staff suggested that rejection of these models would result in either referring both models to the "mop-up" review held at the end of this year's STAR process or revert to the 2017 reviewed assessment.

These were both untenable options in this reviewer's opinion. Referring this model to a "mop-up" panel was likely not possible given the time constraints. Also, the current model was a dramatic improvement over the one reviewed in 2017. As such, this reviewer agreed that this model should be considered the best available science and that it was appropriate for management use in the interim until the next full review. This discussion resulted in a recommendation on process.

As such both modes and data used in these assessments should be considered best available and are recommended to provide management advice. This is due in part to the fact that the models were relatively robust to sensitivities, and the overall stock status for both Northern and southern stocks was very positive relative to management targets. However, there is enough uncertainty for both stocks to merit advice being based on a Category 2 rather than a Category 1 stock (PFMC, 2020), despite the level of data and analysis. Additionally, both stocks should likely go through full peer reviews rather than be updated for the next examination of stock status.

Recommendation: Allow for alternatives when a proposed model may not be appropriate for management use but is better than the most recent peer-reviewed model. During this STAR process, many were uncomfortable approving the proposed base models for management use. After discussions with the staff it was clear that a mop-up panel for this assessment was not feasible, give the work that needed to be done, as well as the schedule already being full for that mop-up. The only alternative was to have lingcod advice revert to the 2017 peer-reviewed assessment. This was untenable, as there had been vast improvements in the model since 2017. Finding an alternative to this issue is recommended. Whether that be a short-term "Plan B" like is used in the Northeast US or an expedited re-review the following year, some process mechanism is needed to address instances where the models are improvements, but still may have large uncertainties.

6. When possible, provide specific suggestions for future improvements in any relevant aspects of data collection and treatment, modeling approaches, and technical issues, differentiating between the short-term and longer-term time frame.

A multitude of research recommendations were made during discussions, including:

High priority

- Consider alternative means of modeling sex-varying selectivity curves.

- Better information and exploration of the differences in fishing strategies throughout the fixed-gear fisheries, particularly the live fish fishery in the northern model.
- Further exploration of natural mortality for both Northern and Southern stocks.
- The ability to weight discard length composition data vs. retained length composition in SS.
- Including the CCFRP dataset in future assessments as a fishery-independent data source.
- Explore consequences of modeling recreational catch and discards as numbers rather than converted biomass.
- Explore possible approaches to improve the estimation of trawl discard mortality rates, given the mismatch between model estimates and GEMM estimates.
- Additional and expanded age structure collection for the southern model.

Medium Priority

- Explore time-varying growth.
- Investigate the Northern stock border and connectedness with Canadian/Alaskan stocks, including the Salish Sea.
- Support ongoing efforts to compare fin ray and otolith aging comparisons as well as further comparisons in aging between NWFSC, DFO, and ADFG.
- An in-depth analysis of the biology of the species relative to alternative approaches to standardizing the composition data is warranted, in the north model particularly.
- Re-evaluate separate sigma R between north and south in future assessments.
- More consideration needs to be made on how best to implement retrospective analyses in models that include time blocks and address parameter inefficiencies.
- Analyze the trawl fishery relative abundance index data by individual stock rather than both combined.
- Continue to investigate the potential sources of bias to the nature of fishery-dependent age-data sampling.
- Consider improvements to both sampling approaches and the preparation and treatment of the age data in assessments.

While there were no low priority recommendations made during discussions, this reviewer would like to put forth recommendations (in order) that should be addressed. These have been marked as either likely to be completed prior to the next assessment (short term) and those which should be done as soon as possible thereafter (long term).

Recommendation: Further work on stock boundary definitions, particularly with Canada. As mentioned elsewhere in this report, the boundary between the northern stock and the stock assessed and managed by DFO is uncertain. Lingcod are clear found, based on trawl survey and other data, continuously from Washington, through British Columbia, into southern Alaska. Further work, including tagging and genetic/morphometric studies, might shed light on the connectedness of this northern US stock with the adjacent Canadian stock (short or long term).

Recommendation: Alternative means of modeling sex-specific selectivity curves. One important aspect, given the spatial differences in the distribution of males vs females, would be a more thorough analysis on varying selectivity by sex in both models. It has been hypothesized by this reviewer as well as others, that this might be one of the causes for the tension in data elements for both models. It is

possible, that the differences in sex-selectivity might be different latitudinally, possibly explaining some of the North-South, differences observed in these assessments (short term).

Recommendation: Further explorations including tag-based studies and/or simulation modeling to inform natural mortality. Natural mortality is the largest source of uncertainty in both the northern and southern models. As such, any possible studies, either tagged base, meta-analysis, or life history, which can be done to resolve this parameter would be very important (long term).

Recommendation: Additional and expanded age structure collection for the southern model. This is an important data need for this stock, from both fishery-dependent (e.g., hook and line survey) and fishery-independent sources. Completing more ages may reveal similar tensions between data sources in the south, as appear in the north (long term).

Recommendation: Allow SS the ability to weight discard length composition data vs retained length composition data. It was noted that discard length composition data are better sampled than retained length composition data, yet when the data are treated within the same fleet the tuning and weighting approaches treat the two types of data equally. The ability to weight these two components of the length composition data separately would be very helpful in disentangling sources of tension in this (and other) models (short term).

Recommendation: A better process for compiling recreational catches, including landings and discards. Concerns were expressed during the review workshop on estimates of discards as well as recreational removal, particularly in Washington. Generally, by the time a model is being reviewed most, if not all of the data issues are more or less worked out. However, it was clear that there were still some outstanding issues relative to recreational removals, particularly in the northern stock. While this is understandable given the recent border changes and the multitude of fleets. A more formalized process of vetting those removals, prior to the STAR meeting is recommended. One suggestion would be to produce a formalized report on the recreational catch which is then signed off on the major contributors prior to inclusion in the models (short term)

Recommendation: Examination of the live fish fishery. The live fish fishery is a growing source of removals in the northern stock area. It is not unfathomable that this may represent a completely separate fleet, with its own exploitation pattern and selectivity, in the near future. As such, sensitivity analysis in the next assessment should be conducted to determine if the live fish fishery has a different selectivity (short term).

7. Provide a brief description of panel review proceedings highlighting pertinent discussions, issues, effectiveness, and recommendations.

In short, a five-day meeting was held remotely to examine both Northern and Southern lingcod assessment as part of the STAR process. The first day centered around the presentations of each of the stock's models, their structure, inputs, and parameterizations. The second through fourth days focused on requests to the STAT, to allow Panel members to explore model behavior, and to test the robustness to uncertainty. The fifth day was mostly for writing the STAR Panel report, with some follow-up questions and results from the STAT on the axis of uncertainty, as well as lingering requests.

Conclusions and Further Recommendations

Overall, this was an enjoyable review. Though there was some frustration among Panel members, including this reviewer, regarding model performance and the short timelines, as always. It was intriguing to review such complex models for two adjacent stocks. It is noted that online-only meetings of this type have some significant drawbacks. In this reviewer's opinion, this would have been an easier process had this meeting occurred face-to-face.

The STAT did an excellent job in preparing for this meeting. While there were some shortcomings with the report, it was clear that the STAT had done their due diligence to fully explore both models, their behavior, and their uncertainty. The STAT should be commended for their work, in many cases, they have pushed the envelope of assessment science, and dramatically improved the model for these species on the West coast of the US. These are complicated models, and the STAT did well in applying them to these stocks.

There is good data to inform modeling approaches here, in this reviewer's opinion. However, the phrase "data-rich, but information poor" was used quite often, and this reviewer agrees with that sentiment. While there was good information from the data on length, ages (at least for the north), and surveys, there was not very good data on natural mortality stock boundaries and movements in the north, the differences in sex selectivity, and the potential influences of the live fish fishery. There is lots of data, but there isn't the right type of data to completely support models this complex, especially given the productivity differences between north and south assessments. As an example, the southern model with fewer data available had better diagnostics and more reliable results than the more data-rich northern model, in part due to a lack of age data in the south which prevented conflicts between the length and age data seen in the north.

This reviewer was satisfied with the results. Approving both models for management use, listed as Category 2 stocks, and with a recommendation to move forward with a benchmark for the next assessment was a tenable outcome: but barely. This reviewer left the final day of the meeting with the feeling that something is being missed here but was unable to define what or why. Regardless, there is little doubt about the status of the stock relative to reference points.

That said several additional recommendations became apparent both during the meeting as well as during this report's write-up.

Recommendation: Completing most of the report prior to the start of the meeting. The pre-STAR report for Lingcod was pretty rough, lacking key write-ups, with some errors, and lacking a few of the analyses presented to the Panel. In the future, more time should be allotted to the STAT to more fully and completely write up the report. This is rather crucial, particularly for reviews that are conducted remotely.

Recommendation: Send presentations for day one ahead of time. Presentations on introductory material, model formulations, data, inputs, sensitivities, and initial model behavior should be provided at least a few days prior to the meeting. Again, this is crucial for a remote meeting. While the reports were made available prior to the deadline of this Panel, they were not as complete as hoped for. Further having this information in one place ahead of time allows the reviewers to see each of the areas side-by-

side, to compare and contrast. This would simply make the first couple of days of the meeting more productive, with less time spent on this introductory material.

Recommendation: STAT should relay key findings on the slides when responding to requests. During the request and response process between STAT and Panel, Panel members, including this reviewer found that when key points and findings were listed on the slides used, it made the report writing phase easier. Often this reviewer, and others, found themselves trying to discern what the STAT was alluding to when writing up the Panel report. Having these findings in written form on the slides presented would bypass any misinterpretations during report writing.

Generally, by the time a model is being reviewed, most (if not all) of the data issues are more or less worked out. This is understandable given the recent border changes and the multitude of fleets. **Recommendation: An overarching workshop on retrospective patterns, what they mean, and what to do about them is needed.** Both models, to some degree, have retrospective patterns associated with them. Moreover, retrospective patterns have started to emerge in this and other North Western US fishery stock assessments. While much work on this topic has been done in the North Eastern US and ICES, an examination in this region is warranted. Main region-specific causes, how this uncertainty should be captured and accounted for when setting OFLs, and other topics should be discussed. As the length of time data are available increases, it is likely that more, and potentially worse, retrospective patterns will emerge. This will require dedicated effort.

References

Johnson, K.F., I.G.Taylor, B.J. Langseth, A. Stephens, L.S. Lam, M.H. Monk, J.E. Budrick, M.A. Haltuch. 2021. Status of lingcod (Ophiodon elongatus) along the southern U.S. west coast in 2021. Pacific Fisheries Management Council, Portland, Oregon. 187 p.

Pacific Fishery Management Council (PFMC). 2020. Terms of reference for the Groundfish and Coastal Pelagic Species Stock Assessments Review Process for 2021-2022. Pacific Fishery Management Council.

Taylor, I.G., K.F. Johnson, B.J. Langseth, A. Stephens, L.S. Lam, M.H. Monk, A.D. Whitman, M.A. Haltuch. 2021. Status of lingcod (Ophiodon elongatus) along the northern U.S. westcoast in 2021. Pacific Fisheries Management Council, Portland, Oregon. 293 p.

APPENDICES

Appendix 1: Bibliography of materials provided for review

Budrick, J. 2016. Evolutionary processes contributing to population structure in the rockfishes of the subgenus genus Rosicola: implications for fishery management, stock assessment and prioritization of future analyses of structure in the genus Sebastes. PhD thesis, University of California, Berkeley.

Haltuch, M.A., Wallace, J., Akselrud, C.A., Nowlis, J., Barnett, L.A.K., Valero, J.L., Tsou, T., Lam, L. 2018. 2017 Lingcod Stock Assessment. Pacific Fishery Management Council, Portland, OR. Available from http://www.pcouncil.org/groundfish/stock-assessments/

Hyde, J.R.; Kimbrell, C. A.; Budrick, J. E.; Lynn, E. A.; Vetter, R.D. 2008. Cryptic speciation in the vermilion rockfish (Sebastes miniatus) and the role of bathymetry in the speciation process. Molecular Ecology 17: 1122–1136. doi: 10.1111/j.1365-294X.2007.03653.x.

Hyde, J.R., and Vetter, R.D. 2009. Population genetic structure in the redefined vermilion rockfish (Sebastes miniatus) indicates limited larval dispersal and reveals natural management units. Canadian Journal of Fisheries and Aquatic Sciences 66(9): 1569–1581. doi: 10.1139/F09-104.

Longo, G.C., Lam, L., Basnett, B., Samhouri, J., Hamilton, S., Andrews, K., Williams, G., Goetz, G., McClure, M., and Nichols, K.M., 2020. Strong population differentiation in lingcod (Ophiodon elongatus) is driven by a small portion of the genome. Evolutionary applications, 13(10), pp.2536-2554.

Methot, R. and Wetzel, C. 2013. Stock synthesis: A biological and statistical framework for fish stock assessment and fishery management. Fisheries Research 142 (2013) 86– 99.

Monk, M., Miller, R., Field, J., Dick, E.J., Wilson-Vandenberg, D., Reilly, P. 2016. Documentation for California Department of Fish and Wildlife's Onboard Sampling of the Rockfish and Lingcod Commercial Passenger Fishing Vessel Industry in Northern and Central California (1987- 1998) as a Relational Database. NOAA Technical Memorandum NMFS. NOAA-TM-NMFS-SWFSC-558. doi:10.7289/V5/TM-SWFSC-558.

Pearson, D.E.D., and Erwin, B. 1997. Documentation of California's Commercial Market Sampling Data Entry and Expansion Programs. National Marine Fisheries Service.

Pearson, D.E., Erwin, B., and Key, M. 2008. Reliability of California's groundfish landing estimates from 1969-2006. Tational Oceanic; Atmospheric Administration.

Thorson, J.T. and Barnett, L.A., 2017. Comparing estimates of abundance trends and distribution shifts using single-and multispecies models of fishes and biogenic habitat. ICES Journal of Marine Science, 74(5), pp.1311-1321.

Appendix 2: A copy of the CIE Performance Work Statement

Performance Work Statement

External Independent Peer Review by the Center for Independent Experts

Stock Assessment Review (STAR) Panel 2 - Virtual

Lingcod (Northern and Southern stocks)

July 12-16, 2021

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 National Marine Fisheries Service and the Pacific Fishery Management Council will hold three stock assessment review (STAR) panels and potentially one mop-up panel if needed, to evaluate and review benchmark assessments of Pacific coast groundfish stocks. The goals and objectives of the groundfish STAR process are to:

1) ensure that stock assessments represent the best available scientific information and

facilitate the use of this information by the Council to adopt OFLs, ABCs, ACLs, (HGs), and ACTs;
2) meet the mandates of the Magnuson-Stevens Fisheries Conservation and Management Act (MSA) and other legal requirements;
3) follow a detailed calendar and fulfill explicit responsibilities for all participants to produce required reports and outcomes;
4) provide an independent external review of stock assessments;
5) increase understanding and acceptance of stock assessments and peer reviews by all members of the Council family;
6) identify research needed to improve assessments, reviews, and fishery management in the future; and
7) use assessment and review resources effectively and efficiently.

Lingcod was last assessed in 2017 in two separate single-stock assessment models, addressing the stock in Washington and Oregon in the north, and the California stock in the south. That assessment estimated relative depletion of the northern stock at 57.9% of unfished biomass. The southern stock depletion was estimated at 32.1%, which falls within the precautionary zone for Pacific Fishery Management Council stocks. Lingcod are large opportunistic top predators in the nearshore demersal ecosystem of the northeast Pacific Ocean and are valued both commercially and recreationally in the U.S. groundfish fishery. They range from Kodiak Island, Alaska to Baja California, Mexico. The historical center of abundance is in the waters off British Columbia and Washington State. Male and female lingcod exhibit different life-histories in that the males guard nests in shallow water for 5-7 weeks in the fall, while the females remain in deeper water; accordingly, the sexes are represented independently in the model. Lingcod are harvested commercially by trawl and longline gear, and recreationally by hook-and-line and spear. In California, the recreational fishery accounts for more than 90% of the landings; in Washington and Oregon the landings are more evenly divided between the recreational and commercial fisheries.

Assessments for these stocks will provide the basis for the management of the groundfish fisheries off the West Coast of the U.S., including providing scientific basis for setting OFLs and ABCs as mandated by the Magnuson-Stevens Act. The technical review will take place during a formal, public, multiple-day virtual meeting of fishery stock assessment experts. Participation of an external, independent reviewer is an essential part of the review process. The Terms of Reference (ToRs) of the peer review are attached in **Annex 2**. The tentative agenda of the panel review meeting is attached in **Annex 3**.

Requirements:
Two CIE reviewers will participate in the stock assessment review panel. One CIE reviewer, requested herein, shall conduct an impartial and independent peer review of the assessments described above and in accordance with the Performance Work Statement (PWS) and ToRs herein. Additionally, one "common" CIE reviewer will participate in all STAR panels held in 2021 and the PWS and ToRs for the "common" CIE reviewer are included in **Attachment A**.

The CIE reviewers shall be active and engaged participants throughout panel discussions and able to voice concerns, suggestions, and improvements, while respectfully interacting with other review panel

members, advisors, and stock assessment technical teams. The CIE reviewers shall have excellent communication skills in addition to working knowledge and recent experience in fish population dynamics; with experience in the integrated-analysis modeling approach, using age- and size- (and possibly spatially-) structured models, and methods for quantifying uncertainty. Familiarity with environmental, ecosystem and climatic effects on population dynamics and distribution may also be beneficial. The CIE reviewer's duties shall not exceed a maximum of 14 days to complete all work tasks of the peer review described herein.

Tasks for Reviewers:
The CIE reviewer shall complete the following tasks in accordance with the PWS and Schedule of Milestones and Deliverables herein.

Prior to the Peer Review: Upon completion of the CIE reviewer selection by the CIE Steering Committee, the CIE shall provide the CIE reviewer information (full name, title, affiliation, country, address, email) to the NMFS Contracting Officer Representative (COR), who forwards this information to the NMFS Project Contact no later than the date specified in the Schedule of Milestones and Deliverables. The CIE is responsible for providing the PWS and ToRs to the CIE reviewer. The NMFS Project Contact is responsible for providing the CIE reviewer with the background documents, reports, and other information concerning pertinent meeting arrangements. The NMFS Project Contact is also responsible for providing the Chair a copy of the PWS in advance of the panel review meeting. Any changes to the PWS or ToRs must be made through the COR prior to the commencement of the peer review.

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

Documents to be provided to the CIE reviewers prior to the STAR Panel meeting include:

- The current draft stock assessment reports;
- Previous stock assessments and STAR Panel reports for the assessments to be reviewed;
- The Pacific Fishery Management Council's Scientific and Statistical Committee's Terms of Reference for Stock Assessments and STAR Panel Reviews;
- Stock Synthesis (SS) Documentation;
- Additional supporting documents as available;
- An electronic copy of the data, the parameters, and the model used for the assessments (if requested by reviewer).

Test: Additionally, two weeks prior to the peer review, the CIE reviewers will participate in a test to confirm that they have the necessary technical specifications provided in advance of the panel review meeting.

Panel Review Meeting: The CIE reviewer shall conduct the independent peer review in accordance with the PWS and ToRs, and shall not serve in any other role unless specified herein. **Modifications to the PWS and ToRs cannot be made during the peer review, and any PWS or ToRs modifications prior to the peer review shall be approved by the COR and CIE Lead Coordinator.** Each CIE reviewer shall actively participate in a professional and respectful manner as a member of the review panel's virtual meeting, and their peer review tasks shall be focused on the ToRs as specified herein. The NMFS Project Contact is responsible for any facility arrangements (e.g., video or teleconference arrangements). The NMFS Project Contact is responsible for ensuring that the Chair understands the contractual role of the CIE reviewers as specified herein. The CIE Lead Coordinator can contact the Project Contact to confirm any peer review arrangements, including the meeting facility arrangements.

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

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

Place of Performance:
The CIE reviewers shall conduct an independent peer review during the panel review meeting scheduled for the dates of July 12-16, 2021. Due to current uncertainties in the state of the COVID-19 pandemic at that time, this meeting will be conducted as a virtual meeting, with technical assistance provided by staff from the Pacific Fishery Management Council.

Period of Performance:
The period of performance shall be from the time of award through **August 2021**. The CIE reviewers' duties shall not exceed 14 days to complete all required tasks.

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

Schedule	**Milestones and Deliverables**
Within two weeks of the award	Contractor selects and confirms reviewers. This information is sent to the COR, who then transmits this to the NMFS Project Contact

Approximately two weeks later	Contractor provides the pre-review documents to the CIE reviewers
July 12-16, 2021	**Virtual Panel Review Meeting**
Approximately two weeks later	Contractor receives draft reports
Within two weeks of receiving draft reports	Contractor submits final CIE independent peer review reports to the COR

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

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

Travel:
No travel is necessary, as this meeting is being held remotely.

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

NMFS Project Contact:

Andi Stephens, NMFS Project Contact
National Marine Fisheries Service,
Newport, OR 97365
Andi.Stephens@noaa.gov
Phone: 843-709-9094

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

1. The CIE independent report shall be prefaced with an Executive Summary providing a concise summary of the findings and recommendations, and specify whether the science reviewed is the best scientific information available.

2. The main body of the reviewer report shall consist of a Background, Description of the Individual Reviewer's Role in the Review Activities, Summary of Findings for each ToR in which the weaknesses and strengths are described, and Conclusions and Recommendations in accordance with the ToRs.

   a. Reviewers should describe in their own words the review activities completed during the panel review meeting, including providing a brief summary of findings, of the science, conclusions, and recommendations.

   b. Reviewers should discuss their independent views on each ToR even if these were consistent with those of other panelists, and especially where there were divergent views.

   c. Reviewers should elaborate on any points raised in the Summary Report that they feel might require further clarification.

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

   e. The CIE independent report shall be a stand-alone document for others to understand the weaknesses and strengths of the science reviewed, regardless of whether or not they read the summary report. The CIE independent report shall be an independent peer review of each ToRs, and shall not simply repeat the contents of the summary report.

3. The reviewer report shall include the following appendices:

   Appendix 1: Bibliography of materials provided for review
   Appendix 2: A copy of the CIE 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

Stock Assessment Review (STAR) Panel 2

The specific responsibilities of the STAR panel are to:

8. Become familiar with the draft stock assessment documents, data inputs, and analytical models along with other pertinent information (e.g., previous assessments and STAR panel report when available) prior to review panel meeting.
9. Discuss the technical merits and deficiencies of the input data and analytical methods during the open review panel meeting.
10. Evaluate model assumptions, estimates, and major sources of uncertainty.
11. Provide constructive suggestions for current improvements if technical deficiencies or major sources of uncertainty are identified.
12. Determine whether the science reviewed is considered to be the best scientific information available.
13. When possible, provide specific suggestions for future improvements in any relevant aspects of data collection and treatment, modeling approaches and technical issues, differentiating between the short-term and longer-term time frame.
14. Provide a brief description on panel review proceedings highlighting pertinent discussions, issues, effectiveness, and recommendations.

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

Participants

STAR Panel Members

John Field, National Marine Fisheries Service Southwest Fisheries Science Center (Chair)
Will White, Oregon State University
Matt Cieri, Center for Independent Experts
Cathy Dichmont, Center for Independent Experts

Stock Assessment Team (STAT) Members

Kelli Johnson, National Marine Fisheries Service Northwest Fisheries Science Center
Ian Taylor, National Marine Fisheries Service Northwest Fisheries Science Center
Brian J. Langseth, National Marine Fisheries Service Northwest Fisheries Science Center
Andi Stephens, National Marine Fisheries Service Northwest Fisheries Science Center
Laurel S. Lam, Pacific States Marine Fisheries Commission, Northwest Fisheries Science Center
Melissa H. Monk, National Marine Fisheries Service Northwest Fisheries Science Center
John E. Budrick, California Department of Fish and Wildlife
Melissa A. Haltuch, National Marine Fisheries Service Northwest Fisheries Science Center

STAR Panel Advisors

Lynn Mattes, Oregon Department of Fish and Wildlife, Groundfish Management Team representative
Gerry Richter, B&G Seafoods, Groundfish Advisory Subpanel representative
Todd Phillips, Pacific Fishery Management Council representative

AGENDA:

Monday, July 12, 2021 – 8:30 AM

Early Log-In to Resolve Connection Issues
(8:30 a.m.)
Welcome and Introductions
(8:45 a.m.)
• Roll Call, Introductions, Announcements, etc. John Field, Chair
• Review Terms of Reference
• Review and Approve Agenda
• Review Virtual Formal Operational Guideline Todd Phillips
• Assign Writing Duties John Field

Overview and Inputs to the Lingcod Stock Assessments
(9:15 a.m.)
1. Overview of Lingcod North and South Stock Status Ian Taylor
2. Lingcod Biology and Stock Structure Ian Taylor
– Break (10:15) –
Overview continued
(10:30)
3. Catch History and Fleet Structure Kelli Johnson
4. Fishery-Independent and Fishery-Dependent Data Kelli Johnson
– Lunch (12:00 – 1:00 p.m.) –
Modeling and Results of the Lingcod Stock Assessments
(1:00 p.m.)
5. Lingcod North: Assessment Modeling, Performance, & Results Ian Taylor

2

Monday, July 12, 2021 – continued

6. Lingcod South: Assessment Modeling, Performance, & Results Kelli Johnson
– Break (3:00 pm) –
Public Comment Period
(3:15 pm)
STAR Panel Discussion and Requests to Stock Assessment Team
(3:30 p.m.)
Adjourn for day
(4:00 p.m.)

Tuesday, July 13, 2021 – 8:30 AM

Early Log-In to Resolve Connection Issues
(8:30 a.m.)
1. Review Day 1 and STAR Panel Discussion

(8:45 a.m.)
2. Stock Assessment Team (STAT) presentations of Model Runs and

Analyses Ian Taylor/Kelli Johnson
a. Lingcod North
b. Lingcod South

(9:30 a.m.)
– Break (10:15) –
– Lunch (12:00 – 1:30 p.m.) –
3. STAR Panel Discussion and Requests for Model Runs / Analyses

(1:30 p.m.)
– Break (3:00 pm) –
Public Comment Period
(3:15 p.m.)
STAR Panel Discussion and Planning (continued)
(3:30 p.m.)
Adjourn for day
(4:00 p.m.)

Wednesday, July 14, 2021 — 8:30 AM

Early Log-In to Resolve Connection Issues
(8:30 a.m.)
1. Review Day 2 and STAR Panel Discussion

(8:45 a.m.)

3

Wednesday, July 14, 2021 — continued

2. STAT presentations of Model Runs and Analyses Ian Taylor/Kelli Johnson
a. Lingcod North
b. Lingcod South

(9:00 a.m.)
– Break (10:15) –
– Lunch (12:00 – 1:30 p.m.) –
3. STAR Panel Discussion and Requests for Model Runs / Analyses

(1:30 p.m.)
– Break (3:00 pm) –
Public Comment Period
(3:15 p.m.)
STAR Panel Discussion and Planning (continued)
(3:30 p.m.)
Adjourn for day
(4:00 p.m.)

Thursday, July 15, 2021 —8:30 AM

Early Log-In to Resolve Connection Issues
(8:30 a.m.)
1. Review Day 3 and STAR Panel Discussion

(8:45 a.m.)
2. STAT presentations of Model Runs and Analyses Ian Taylor/Kelli Johnson a. Lingcod North
b. Lingcod South

(9:00 a.m.)
– Break (10:15) –
– Lunch (12:00 – 1:30 p.m.) –
3. STAR Panel Discussion a. Additional model runs/analyses (as needed)
b. Panel/STAT Agree on Final Base Model
c. Decision Table Developed

(1:30 p.m.)
– Break (3:00 pm) –
Public Comment Period
(3:15 p.m.)

Thursday, July 15, 2021 —continued

STAR Panel Discussion (continued)
(3:30 p.m.)
Adjourn for day
(4:00 p.m.)

Friday, July 16, 2020 — 8:30 AM

Early Log-In to Resolve Connection Issues
(8:30 a.m.)
1. Review Decision Tables a. Lingcod North
b. Lingcod South

(8:45 a.m.)
– Break (10:15) –
2. Discuss STAR Panel Report Draft a. Review as appropriate
b. Agree on Process for Completion of Report (due by Aug 15)

(11:00 a.m.)
– Lunch (12:00 – 1:30 p.m.) –
3. Continue STAR Panel Report Drafting (as needed)

– Break (3:00 pm) –
4. STAR Panel Adjourns

(4:00 p.m.)
PFMC 06/17/21