Center for Independent Experts (CIE) Independent Peer Review Report

Stock Assessment Review (STAR) of Cabezon in CA, OR, and WA, Panel 1

Newport, OR. May 6-10, 2019

Dr. Noel Cadigan

Centre for Fisheries Ecosystems Research
Marine Institute of Memorial University of Newfoundland
St. John's, NL. Canada

Executive Summary

The California (South and North), Oregon, and Washington assessments for cabezon constitute the best available scientific information on the current status of the stocks and the assessments provides a suitable basis for management decisions.

The major axis of uncertainty involved the natural mortality rate (M) which was adequately accounted for. M is currently higher than fishing mortality and M seems like a main driver of Cabezon productivity.

The SS3 stock assessment models for the California (South and North) and Oregon stocks were competently applied, and the model inputs were derived using best practice. The Washington assessment was a data-poor assessment using Simple Stock Synthesis (SSS) which was an improvement compared to the previously used approach.

Stock boundaries were based on pragmatic considerations of spatial differences in biogeographic shifts in life history traits and data availability. However, stock structure research suggests there may be more sub-stock structure than is reflected by the current assessment region boundaries. Cabezon appears to be fairly sedentary species so localized depletion could be a problem.

The accuracy of estimates of landings and low discards has improved over time. However, there is uncertainty in catch estimates, and moreso for historic periods, that is not measured and incorporated into the assessments.

The most important deficiencies of these assessments was the lack of reliable fishery-independent indices of abundance. Several fisheries-dependent indices were developed for the OR stock, but they were not highly consistent with each other and did not influence the stock assessment substantially. Standardization methods to derive these indices could be improved to better account for changes in management regulations. The lack of regular age composition information is another important deficiency in these assessments. Length Compositions were the most influential data in the CA and OR stock assessments.

Background

The Stock Assessment Review (STAR) Panel 1, Cabezon in California (CA), Oregon (OR), and Washington (WA) was held in Newport, OR during May 6-10, 2019. The general goals and objectives of the groundfish STAR process are to:

1) ensure that stock assessments represent the best scientific information available and facilitate the use of this information by the Council to adopt Overfishing Limits (OFLs), Acceptable Biological Catches (ABCs), Annual Catch Limits (ACLs), harvest guidelines (HGs), and annual catch targets (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 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.

The purpose of the meeting was to provide technical review of benchmark stock assessments for Cabezon in CA and OR and WA waters.

The Panel was composed of two independently appointed Center for Independent Experts (CIE) reviewers (Dr. N. Cadigan, Canada; Dr. Robin Cook, UK), an independent reviewer from Oregon State University (Dr. Will White) and an independent chair (Dr. Rishi Sharma, National Marine Fisheries Service Northwest Fisheries Science Center, NWFSC) who is also a member of the Pacific Fishery Management Council's (PFMC's) Science and Statistical Committee (SSC). The STAR Review Panel was supported and assisted by Mr. J. DeVore (PFMC), Mr. Patrick Mirick (Oregon Department of Fish and Wildlife), and Mr. Gerry Richter (Groundfish Advisory Subpanel). Assessment documents were prepared by stock assessment teams (STAT's) and presented by Dr. Jason Cope (NWFSC) for two stocks of Cabezon in CA and the WA stock, and by Dr. Aaron Berger (NWFSC) for Cabezon in OR. They were assisted by John Budrick (California Department of Fish and Wildlife), Ali Whitman (Oregon Department of Fish and Wildlife), Kate Bosely (NWFSC), Theresa Tsou (Washington Department of Fish and Wildlife), Lisa Hillier (Washington Department of Fish and Wildlife), and Kristen Hinton (Washington Department of Fish and Wildlife). The support of all of these scientists and staff to the STAR RP process is gratefully acknowledged.

CIE reviewers were tasked with conducting impartial and independent peer reviews in accordance with their SoW and ToRs. The reviewers were required to voice concerns, suggestions, and improvements while respectfully interacting with other review panel members, advisors, and stock assessment technical teams. The CIE reviewers were required to 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-structured models, use of Markov Chain Monte Carlo (MCMC) to develop confidence intervals, and use of Generalized Linear Models in stock assessment models. Each CIE reviewer's duties could not exceed a maximum of 14 days to complete all work tasks of the peer review described herein.

Role of reviewer

All assessment documents and most supporting materials were made available to the RP via an ftp server two weeks before the meeting, on April 22, 2019. These documents are listed in Appendix 1. I reviewed the background documents I was provided and compiled a list of issues to get clarification at the RP meeting. I attended the entire STAR Panel review meeting in Newport, OR. May 6-10, 2019. I reviewed presentations and reports and participated in the discussion of these documents, in accordance with the SoW and ToRs (see Appendix 2). I drafted text for the RP report. After the meeting I participated in email discussions dealing with the review panel summary report. This CIE report is structured according to my interpretation of the required format and content described in Annex 1 of Appendix 2.

Summary of findings

ToR 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.

I reviewed in detail the draft stock assessment document that combined descriptions of data and models for Cabezon in CA (2 stocks), OR, and WA. More specifically, the document described:

- a detailed Executive Summary,
- some basic information on stock structure and distribution, basic life history information, and ecosystem information,
- history of fisheries in CA, OR, and WA,
- history of fisheries management in CA, OR, and WA, and cabezon fisheries in Canada, Alaska, and Mexico
- fishery landings (directed and bycatch) and discards, for CA, OR, and WA.
- commercial length and age data,
- commercial abundance indices,
- recreational landings,
- recreational length and age data
- recreational abundance indices
- fishery independent data

- other biological information (weight-length relationship, maturity schedule, fecundity, natural mortality, stock-recruitment relationship, aging bias and precision)
- data sources evaluated, but not used in the assessment
- history of modeling approaches for CA, OR, and WA,
- response to STAR Panel Recommendations from the previous assessment,
- transition to the current sub-stock assessments
- Model Specifications for CA, OR, and WA
- Model Parameters
- Reference Model Selection and Evaluation
- Reference Model Results
- Evaluation of Uncertainty
- Reference Points
- Regional Management Considerations
- research needs,
- literature cited
- Auxiliary Files

Previous stock assessment documents

The 2009 benchmark assessment document was provided as well as the 2009 STAR Panel Meeting Report, and two CIE reports for that assessment.

Data input documents

No additional documents were provided.

Documentation on analytical models

No additional documents were provided, but I re-read:

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

and the Appendix associated with Methot and Wetzel (2013).

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

Landings Input data

CA

The historical fishery catch reconstruction back to 1916 (the first year of required reporting in the commercial fishery) and by sub-stock uses the same approach as in the 2009 assessment with new catches being added to complete the time series, and also some revision to the CA recreational catch data during 2004-2008. Various sources of information, adjustments, and interpolations are used to estimate the landings. Fishery catches were compiled for four fleet categories: commercial dead-fish, commercial live-fish, recreational boat (both charter and private boats), and recreational shore fishers. This was not reviewed in detail by the RP. The amount of dead discards is estimated (based on total discard estimates and a small discard mortality rate) and is a small part of the landings, so the dead discards were simply added to the landings for the stock assessment.

A significant change in the catch data involved the reallocation of CA catches from the SCS region to the NCS region for years 1980-1995. The reference model and other models requested by the RP used the updated catch series. The effect of this change was proportionally greater for the SCS, where there is less catch data overall than in the NCS.

OR

Cabezon are caught predominantly using hook-and-line gear by recreational fishermen and by hook-and-line or longline gear by commercial fishermen. Commercial catches were recorded back to 1979 and are thought to be negligible prior to 1979 because of the dominance of trawl landings in Oregon historically in which Cabezon would have been rarely encountered. A small dead discard rate (0.07%) was used to calculate total discarded catch by applying it to annual estimates of commercial landings by fleet. Recreational landings were derived back to 1970 using various sources of information, adjustments, and interpolations. For assessment purposes, catches were aggregated into four fleet categories (commercial dead-fish, commercial live-fish, recreational boat, and recreational shore fishers), similar to CA.

WA

Commercial landings in Washington have been low, usually less than 0.6 mt annually. Recreational catches were available for three time periods (1967, 1975-86, 1990-2018). Linear interpolation was used to fill in catches for missing years. A small dead discard rate (0.07%) was applied to catch landings estimates.

Landings – merits and deficiencies CA+OR+WA

Assessment teams have created long time-series of landings which is meritorious. The accuracy of estimates of landings and low discards has improved over time, as expected. This is also a merit. A related deficiency is that there is uncertainty in catch estimates, and moreso for historic periods, that is not measured and provided to the RP. It is also not clear that the SS3 assessment modelling approach could utilize information on catch uncertainty.

There is an important need for assessment teams to provide information on the quality of the annual landings estimates, and more specifically to quantify the uncertainty in these estimates. This is difficult to do, because landings estimates involve measurement error and some bias. More contemporary stock assessment models (e.g. SAM; https://www.ices.dk/marine-data/tools/Pages/Software.aspx) do not fit landings exactly like SS3, but a typical application of SAM does not include estimates of the precision of landings as an input. Some recent stock assessments (DFO 2017, 2018) include some basic information on the range of possible landings (i.e. bounds), although setting these bounds tends to be a subjective decision, especially for more historic time periods. Nonetheless, the DFO approaches do include available information about the quality of landings estimates.

DFO. 2018. Stock assessment of Northern cod (NAFO Divisions 2J3KL) in 2018. DFO Can. Sci. Advis. Sec. Sci. Advis. Rep. 2018/038. (Erratum: August 2018). https://waves-vagues.dfo-mpo.gc.ca/Library/4071407x.pdf.

DFO. 2017. Assessment of the Atlantic Mackerel Stock for the Northwest Atlantic (Subareas 3 and 4) in 2016. DFO Can. Sci. Advis. Sec. Sci. Advis. Rep. 2017/034. https://waves-vagues.dfo-mpo.gc.ca/Library/40619576.pdf.

Length Compositions

This was the most influential data in the CA and OR stock assessments. Length compositions provide one of the few sources of data that can inform the model about variation in year class strength and total mortality rates (for fleets with asymptotic selectivity).

SCS

The combined-sex length composition samples available were: 1) commercial live fishery during 2002-2018, with 4120 measured lengths in 2118 trips for all years. 2) Recreational Shore during 1980-2018 (no samples for 7 years), with only 189 fish measured for all years. 3) Recreational Boat during 1975-2018, with 2972 measured for all years.

NCS

The combined-sex length composition samples available were: 1) commercial live fishery during 1997-2018, with 274,459 measured lengths in 9,579 trips for all years. 2) commercial dead fishery during 1993-2000, with 22, 918 measured lengths in 6,118 trips for all years. 3) Recreational Shore during 1980-2018 (no samples for 4 years), with 14,294 fish measured for all years. 4) Recreational Boat during 1980-2018, with 2972 measured for all years. 5) California Collaborative Fisheries Research Program (CCFRP) during 2007-2018, with only 347 measured lengths in 260 trips for all years.

OR

A total of 14,158 sex-specific and combined-sex (57%) length composition samples were available for the live and dead commercial fleets during 1998-2018. Only combined-sex length compositions were available from the recreational fishery during 1980-2018. The shore fleet had low sample sizes (n=787), with very few samples since 2004. Most of the recreational length frequencies were from the ocean fleet (n=23,496) with more than 500 fish sampled each year since 2001.

WA

Limited length compositions from 2002 to 2018 were used in the assessment. Summaries of the sample sizes were not provided to the RP.

Length Compositions – merits and deficiencies CA+OR

Most of the assessment model fleets had at least some length composition samples – and for some fleets and years the sample sizes seemed good. I found this hard to judge however. It would be useful to provide a plot of annual length composition sample size per ton of catch for each fleet and stock. This will provide useful information on the adequacy of coverage. A merit of the assessments is that the number of trips sampled (rather than total number of fish) was taken as the input sample size of the length compositions.

At various times during the RP it was mentioned by participants that there was some geographic separation in various fleets that might affect interpretation of the length composition samples. A deficiency involved with this data was that the RP was not provided with information on this. In future it seems useful to provide some type of annual and fleet-specific maps of the locations of trips sampled for lengths. Standardization of compositional data has been advocated by Thorson (2014) and related issues of "representative sampling" should at least be discussed for Cabezon.

A deficiency in the CA data was that no sex-specific compositions were available. Sex-based sampling was only available for the commercial fleets in OR. As the assessment models are sex-based, it seems useful to collect more sex-specific length compositions for all fleets where possible. If a simple morphometric measurement (like color, etc.) could be collected, and some accurate sex-frequency by morphometrics sampling conducted then it might be possible to infer sex-based length compositions for each fleet.

The SCS stock has limited recent removals and also size composition information. This will be a problem for future assessments of this stock.

Thorson, J.T., 2014. Standardizing compositional data for stock assessment. ICES Journal of Marine Science, 71(5), pp.1117-1128.

Age compositions – merits and deficiencies

The lack of regular age composition information is a deficiency for the CS, OR, and WA assessments. There is no aging data in the SCS assessment, and very little in the NCS assessment (578 otoliths between 1991 and 2002). More ageing information was available for the OR assessment (367 ages during 2003-2018 for the commercial dead fleet; 2328 ages during 2005-2018 for the recreational boat fleet; 112 research ages). Obviously, continued collection of aging data will improve these assessments.

Length-weight relationship

The same length-weight data were used in this assessment for the CA stocks as in the 2009 assessment. The length-weight relationship for the OR stock (combined sex) was updated based on data from the Oregon Sport Boat Survey (ORBS) biological database (recreational) and PacFIN (commercial). For the WA stock, length-weight data were collected from 929 sport caught individuals. Although the estimated relationships were plotted, the individual data were not shown so I cannot assess the adequacy of the estimation. This is a deficiency. It seems easy and useful to conduct a GLM to investigate for statistical differences in length-weight relationships across stocks, with the objective of combining information where appropriate. It is also useful to check if the relationship has varied over years.

Maturity and Fecundity

Different length-based maturity curves were used for the CA stocks and the OR+WA stocks. These were not updated since the 2009 assessment, which could be a deficiency. Contemporary maturity data should be collected to check if maturity schedules have changed.

The relationship with fecundity and size/age is uncertain, because of the possibility of multiple spawnings per year.

Indices of abundance

The most important deficiencies of the SCS and OR assessments was the lack of reliable fishery-independent indices of abundance. It seems difficult to survey this near-shore species. However, the California Collaborative Fisheries Research Program (CCFRP) nearshore survey for 2007-2018 was used to derive a fishery-independent abundance index for the NSC. The CV's for this index ranged from 20-40% which is ok for stock assessment. The CCFRP index is a merit for the NCS assessment and a recommendation is to extend this survey methodology to the SCS and OR stock areas.

Cabezon occurs in about 5% of the samples in the CCFRP survey. Although the CCFRP follows a stratified random sampling design, the NCS index was derived following a series of delta-GLM's, which is typical in west-coast assessments. This seemed to be an unnecessary complication to me, because Cabezon is not a shoaling species with infrequent large catches that the delta-GLM's are designed to accommodate. A Negative-Binomial GLM is easier to implement and may provide more reliable statistical inferences. I described this approach further in ToR 4. An advantage of this approach is straight-forward and common standard error calculations compared to the simulation procedures used in these assessments. I recommend that this approach be investigated in future assessments.

Both CA stocks used the same recreational indices based on Commercial Passenger Fishing Vessel (CPFV) logbooks as developed for the 2009 assessment. These indices were developed only for the 1960-1999 time period because of the temporal and spatial management restrictions since 2000. Having a long index time-series is a merit. The RP did not examine the analytical methods used to derive these historic indices in much detail. However, the indices seemed to vary smoothly over years, usually without large between-year variability that could not reflect variations in stock size. However, these indices were not fit very well in model runs and the STAT felt that this was because the CPFV indices could only reflect broad-scale trends in the stock (over many years). It was not clear to me why the STAT did not trust these indices. This was a deficiency, which possibly reflects that the index standardization methods need improvements.

Several fisheries dependent indices were developed for the OR stock, which is a merit for this assessment. A deficiency is that the STAT team did not seem to trust these indices, except to indicate broad-scale trends in abundance. Similar to above, I find that the application of delta-GLM models was unnecessary and that Negative-Binomial GLMs (which are more appropriate for the type of count data collected in these surveys) should be investigated in future assessments (see below).

The Oregon Logbook Index (2004-2018) standardization first involved a variety of filtering steps to extract consistent records representative of the fishery to best estimate the relative abundance trend through time. The RP did not review this aspect, although the filtered logbook dataset represented only 36.9% of recorded catch from 91 vessels. The index standardization included effects for month, port, season (two-month intervals), vessel, trip limit regulation, target species specification, and number of crew. Index standardization is clearly a merit; however, in future I suggest that the distribution of covariates over years (i.e. via boxplots, etc.) be examined. If there is a pronounced change in a covariate over years then it is possible that years effects (the abundance index) could be partially confounded with other covariate effects.

Including trip limit regulations as a covariate is a deficiency. This is not the way to address the censoring of catches because of trip limits. I illustrate a better approach using simulated Cabezon-like data in ToR 4. This should be addressed in future stock assessments.

Oregon Onboard Observer recreational Index (2001, 2003-2018) delta-GLM standardization model included factors for YEAR, DEPTH, MONTH and REEF.

The Oregon ORBS Dockside Index (2001-2018) was based on information from the Oregon Recreational Boat Survey (ORBS). Additional deficiencies with this data included more uncertain effort calculations (i.e. hours fished) due to adjustments for travel time and Stephens and MacCall filtering to identify trips that were likely to catch the species. Uncertainty in effort measurement is a type of covariate measurement error problem that can lead to "bias attenuation" in CPUE (i.e. a flattening of the stock size index compared to real stock size variation). Further filters were applied to season, bag limit, effort, and catch rate attributes; however, I did not understand the season filtering because season was a factor in the standardization model. I also did not understand the bag limit filtering. The delta-GLM standardization model included subregion, season, month, and year, and boat type. A full review of index standardization methods was beyond the scope of the STAR Panel Review; however, the above issues made me uncertain about the efficacy of the ORBS index. During the RP, a minor improvement was made to the index calculation by removing the year:boat type interaction term in the index.

The Oregon Marine Recreational Fisheries Statistical Survey (MRFSS) Dockside Index (1980-1989; 1993-2000) was derived from fish sampled in angler bags following completion of a trip. There was a discussion of uncertainty about identifying what a trip was. Stephens and MacCall filtering was used. Standardization factors were year, subregion, and various within-year timing factors. Standard errors from the delta-GLM standardization model seemed too small. This standardization had a substantial effect compared to the raw CPUE and should be investigated again in future assessments, particularly to assess if year effects are partially confounded with wave effects, Recreational bag-limits seem to have changed over time in OR, and it was not clear to me why this factor was not considered in the Onboard Observer and ORBS index standardizations. This should be considered in future assessments.

Analytical methods – merits and deficiencies

There were two main analytical methods/models used in this assessment: delta-generalized linear models to provide standardized stock size indices, and the SS3 stock assessment model (version 3.30.13-safe;_2019_03_09). I included some comments on index standardization methods in the previous section. SS3 is a flexible stock assessment modelling framework that can integrate intermittent samples of length compositions, age compositions, and various types of abundance indices. It is an appropriate choice for the CA and OR Cabezon stocks.

Model convergence was checked for all models during development of a reference model by ensuring that the final gradient of the likelihood surface was less than 0.001 and produced asymptotic standard deviations (i.e., the Hessian matrix would invert). All estimated parameter values were also checked to ensure they were not hitting minimum or maximum bounds. The ability of the reference model to recover the same likelihood estimates when initialized from dispersed starting points (i.e., the jitter option in SS) was performed using 100 'jittered' starting values. Jitter magnitudes of 0.05 and 0.1 were explored. Jittering at either value did not find a lower likelihood for either of the CA models or the OR model.

I conclude from the review meeting that the model was very competently applied. The skill of the STAT with SS3 was a strong merit. I continue to be impressed with the r4ss package and the HTML outputs that allowed the STAT to quickly produce relevant plots and other output based on requests for additional runs. This greatly improved the efficiency of the review.

Growth was estimated within the NCS and OR model, which is appropriate given the various size limit regulations used in the recreational fisheries.

The CA and OR models did not include recruitment deviations in historic periods that only have catch estimates. This results in false precision about estimates of historic stock size and especially about spawning depletion. It also resulted in some lack of fit to historic indices (i.e. SCS and NCS CPFV). This is a deficiency in the application of SS3. The STAT felt that it was not reasonable to use recruitment deviations historically or to generate non-equilibrium starting population size, because this produced strange historical recruitment deviations caused by short-term variations in landings. However, this did not make sense to me. SS3 has an unpenalized annual F parameter to fit annual catches and I did not understand why the model would use penalized recruitment deviations for this. This should be further investigated in future assessments.

A related issue to the above point is that the fits to most of the indices (SCS – CPFV; NCS – CPFV; OR = Logbook, Onboard, ORBS, MRFSS) seemed to have auto-correlated errors. SS3 uses an index likelihood function that assumes independent errors across time. If errors are autocorrelated then the SS3 independence assumption will result in higher negative loglikelihoods (nll's) and perhaps too much weight given to fitting the indices, which is a deficiency. I suggest that it is reasonable to expect autocorrelated errors for fishery-dependent abundance indices because of the many possible "transitory" fishery effects in catch rates that are not related to trends in stock abundance but that standardization models do not account for. However, it is difficult to reliably estimate autocorrelation parameters and a prior on these parameters may be justified.

I also found it difficult to evaluate the adequacy of the fits to length compositions and conditional age compositions. I am never sure when fits are too bad to accept. Length composition residual diagnostics for the SCS, NSC, and OR models did not seem to conform to the underlying statistical assumptions used in model estimation. There is usually correlation in residuals across lengths, and often across years as well. However, SS3 assumes a multinomial distribution for length compositions and there should only be negative correlation between residuals, although clearly there has to be positive and negative residuals each year. The same issue exists for the Dirichlet-Multinomial distribution. The multinomial and Dirichlet-Multinomial distributions are more appropriate for nominal categorical data, where there is not a natural ordering between categories. This is why the correlation between category responses is always expected to be negative for this type of data. However, length compositions are ordinal or even interval categorical data, and we might expect different correlation patterns in this case. The impact of mis-specifying the statistical distribution of the length compositions is higher nll's than the data warrant, and perhaps too much weight given to fitting the compositions. This is a similar deficiency to autocorrelated index errors.

A potential impact of mis-specifying the correlation structure of index or length composition data is less reliable profile likelihoods and estimation of important productivity parameters. The NCS model favored lower values of female M than seemed plausible to the STAT. A low M run resulted in a substantially lower negative log-likelihood for the CPFV index. The improvement in fit to that index was concentrated on two time-periods, and the STAT felt that this was not evidence of lower M for the NCS stock. I think that if the CPFV index errors were modelled by a AR(1) multivariate normal distribution, then the M profile for this index may have provided less support for a low M.

Time-blocking of selectivity is also tedious but useful when there are important changes in management measures. I conclude that the blocking used in CA and OR models was appropriate. In other fora (e.g. Canada, ICES) this type of blocking is not commonly done. Selectivity is modelled annually but sometimes with smooth variations over time (e.g. random walk). Such an option may be useful for SS3, for diagnostic purposes at least.

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

Stock Structure

Stock boundaries were based on pragmatic considerations of spatial differences in biogeographic shifts in life history traits and data availability. Indices of abundance and length/age composition data were aggregated at the SCS, NCS, OR and WA spatial units, balancing the need for spatial resolution versus data availability and quality. However, stock structure research suggests there is more sub-stock structure than is reflected by the current assessment region boundaries. For example, genetics studies in California suggested 6 populations. The basic reproductive life history processes of Cabezon are not well understood, and there is uncertainty about larval connectivity at coastal scales, because genetic results suggest more spatial structure than may be expected given their fairly long pelagic larval duration.

Cabezon appears to be fairly sedentary species so localized depletion could be a problem (see below). However, in California there is an extensive network of no-take marine protected areas that should alleviate this concern to some extent. There is also a less-extensive set of no-take MPAs in Oregon. Nonetheless, in future assessment fine-scale (space+time) catch and effort data should be examined, where available, to investigate if local depletions occur.

Growth

The two sexes are modeled separately, as growth is very different between females and males. This seemed appropriate to me. Maturation may affect growth rates, but there is insufficient data available to investigate this for CA Cabezon.

The SCS and NCS models are basically age-structured catch at length stock assessment models. There are no ages in the SCS model and growth parameters are assumed to be known and the same as in the NCS. This was reasonable, but a deficiency is that uncertainty related to size at age for the SCS is not included in model outputs. There is only a small amount of age data for the NCS. The length at age 0 was fixed to be zero, which improved model estimation. All other growth parameters were freely estimated with no priors. The NCS Von Bertalanffy (VonB) k parameter estimate was 0.34 for males and 0.21 for females, and Linf was about 56cm for females and 42cm for males. This seemed reasonable, and I found no major evidence of lack of fit in the age at length data. However, the estimated CV_A2 for females (0.16) was much greater than for males (0.07), leading to a much wider distribution of length at older ages for females than males. Conversely, the estimated CV_A1 for females (0.16) was much lower than for males (0.43). The standard errors for these parameter estimates were small leading me to conclude that the differences in male/female CV's were statistically significant. It was not clear to me why these strange differences may exist in the NCS population.

There is much more ageing data available for the OR stock. The female Linf estimate was about 64cm and for males was about 57cm. Hence, for this stock males are estimated to have a maximum size about 13cm greater than in CA, whereas females have a maximum size only 8cm greater than in CA. The CV_A2 estimates were 0.06 and 0.08 for males and females, more consistent with the male value for CA. The CV_A1 were about 0.3 for males and females.

There are some curious differences in the estimates of size at age between the NCS and ORS that should be further investigated in future assessments. However, I do not think this is a major source of uncertainty.

Some ages (184 otoliths) were used for the WA stock to estimate a VonB growth model. We did not examine this during the RP.

Selectivity

Selectivity was modelled via fleets and time-blocks, which is typical in US stock assessments. The choice of time-blocks was decided based on knowledge of changes in management regulations and also by residual analyses and model building, aided by AIC to assessment improvement in fit relative to model complexity. Selectivity was modeled as a function of length, using 6 parameter double-normal selectivity curves. This is good, and better than modeling selectivity as a function of age.

Some fleets had asymptotic selectivity's in the CA and OR assessments, which is a merit; otherwise it is more difficult to separate mortality rates and selectivity patterns. In the SCS model the Comm.Live fleet has a strong dome shape and selectivity rapidly declined between 50 and 60cm. This fleet rarely catches Cabezon > 55 cm, whereas the recreational shore and especially the boat fleets do catch Cabezon > 55cm. Hence, the differences in selectivity between the Comm. Live fleet and the Rec boat fleet seemed appropriate. The selectivity estimated for Comm. Live fleet in the NCS was also domed, and asymptotic for the recreational shore and boat fleets; hence, there is consistency between the two models. Most fleets in the ORS had asymptotic selectivity, except the recreational shore fleet (with fairly low sample sizes) and the first time block of the commercial live fleet. This latter fleet had some curious residual patterns that seemed opposite for males and females, and this should be explored further in future assessments.

In data-rich assessments in ICES and Canada it is not common to use fixed-block parametric selectivity models. It is more common to freely estimate F's or, in more contemporary state-space models, to use a stochastic process in which F's are correlated over age/length and time. I appreciate that it is more difficult to do this with the data-moderate/poor assessments of the US west coast. However, selectivity will not be fixed in real populations like the models assume, and this is a source of uncertainty that I am not sure is adequately accounted for in general, but I also don't think this is a big issue for Cabezon because F does not seem to currently be a big source of mortality, and described below.

Natural Mortality

This is a difficult parameter to specify in most stock assessments, and Cabezon are no exception. M was assumed to be age and year invariant, but different for males and females. Given the data available, this seemed reasonable to me. The natural mortality tool was used to derive a prior on M, which facilitated a more realistic quantification of model uncertainty.

The STAT and RP closely examined specification of M for the CA and OR models. The RP and STAT agreed that the major axis of uncertainty involved M. I conclude that uncertainty about M is adequately accounted for. However, M is a much larger source of mortality for Cabezon than fishing, and M probably varies temporally, spatially, and by size/age. Hence, M seems like a main driver of Cabezon productivity and future research (e.g. data storage or telemetry tagging) should be conducted to provide more reliable values for Cabezon.

Intuitively I would not expect the CA or OR assessment models to be informative about M because of the lack of reliable abundance indices and regular age-sampling. However, profile likelihoods for the SCS, NCS, and OR assessments all indicated some information about female M, but the low value estimated in the NCS model and the high value estimated in the OR did not seem reliable and could be the consequence of some other model mis-specification. However, it was not clear what that misspecification was.

Stock-recruitment

The CA and OR assessments assume a Beverton-Holt stock-recruit relationship and the steepness formulation, which is typical in US stock assessments. Steepness (0.7) and recruitment variability (sd=0.5) were fixed to the same values as in the previous assessment.

Estimating steepness was attempted through sensitivity model runs, but a lack of contrast in exploitation (among other things) led to little information about steepness and unrealistic estimates suggested by likelihood profiles, which were fairly flat in any event except for OR, where there was some conflict in data sources about steepness values.

Recruitment deviations were included only for the following periods: SCS: 1970-2016; NCS: 1962-2016; and ORS: 1980-2015. However, surely recruitment deviations also occurred outside of these periods. Other assessments have used additional 'early' deviations so that age-structure in the initial modeled year would deviate from the stable age-structure in a way that is consistent with estimated variability in recruitment. This allows recruitment variability to be included in uncertainty about B_o. The STAT suggested that current practice in the US discourages including using early recruitment deviations. As indicated above, I am not sure why this is, and I think it is an issue that should be re-visited in future assessments.

Uncertainty

The STAT had a focus of quantifying the uncertainty in stock size and status as accurately as possible given the data available.

One can argue that there are sources of uncertainty not included in model outputs. However, uncertainty estimates of SSB and SSB depletion seemed wide for the two CA stocks, and I do not feel that these assessments are overly precise. The OR assessment produced more precise estimates of SSB and SSB depletion, but that assessment has more age data and abundance indices, so more precise estimates are expected. However, profile analyses demonstrated that most of the data for the ORS favored a seemingly implausibly high value of M. This could indicate some mis-specification of some other aspect of the model, but it was not clear in the RP what this was. Nonetheless, this may suggest there is additional unaccounted uncertainty in SSB and SSB depletion for ORS.

Sensitivities

The STAT did really good work before the RP in performing a wide range of sensitivity analyses (both to data and structural model uncertainty) and documenting the results in succinct and easy to understand graphical displays.

Retrospective analysis for the CA stocks did not indicate estimation problems. There were appreciable differences from the previous assessment 10 years ago, but a substantial amount of new data (especially for NCS) has been collected since 2009, and I do not find the retrospective differences in depletion to be a problem. The current assessment uncertainties in depletion cover the 2009 estimates. Retrospective analyses for the ORS indicated problems estimating the scale of the assessment (i.e. SSB), but depletion was estimated more reliably. There was also a substantial retrospective difference between the 2009 and 2019 assessments of this stock. This seemed to be related to differences in SS3 versions that the STAT explored in considerable detail.

WA assessment

This was based on the The Simple Stock Synthesis model which seemed to be an improvement on the DBSRA formerly used for this component.

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

I think the delta-GLM standardization methods for fishery-dependent indices was not appropriate for Cabezon, nor were the effects of bag- or trip-limits appropriately accounted for. Cabezon is not a shoaling species that a delta-GLM approach would be appropriate for. As I understand it, Cabezon are rarely caught (i.e. per trip) in recreational fisheries and are only caught in small numbers. They are rarely the main target when fishing.

Bag or trip limits result in censored catch data, or catches that are less than or equal (<=) to what might have occurred with no trip limits. This results in a sort of biased response measurement and including trip limits as a factor in an index standardization model is not the appropriate way to correct for these limits. Note that if trip limits are very frequently met, then the catch rate data will be almost uninformative about trends in stock size.

I demonstrate using simplified Cabezon-like simulated data how to use censored regression methods to derive an unbiased abundance index when trips limits are meet infrequently (~5%) and when trip limits change over time. These seem like relevant issues for the Cabezon stock assessments.

I generate 25 years of data, with 2000 trips each year to locations (i.e. reefs) that have the same density of Cabezon (10 fish per reef; Fig. 1A). I assume there is a 5% chance of catching a Cabezon with a standard unit of effort. I assume the relative angler effort has a gamma distribution (Fig 1B), and that the number of Cabezon caught is Negative Binomial (NB) distributed (Fig's 1C,D). The simulation population generator R script is

```
n.year = 25
n.trips=2000
stock.den = 10 + arima.sim(list(order = c(1,0,0), ar = 0.7), n = n.year)

p = 0.05
mu = p*stock.den
k=2

#NB generator
gen = function(i){
  rel.effort = rgamma(n.trips,shape=1,scale=1)
  y = rnbinom(n.trips, size=k, mu=mu[i]*rel.effort)
return(cbind(y,rel.effort))}

y = sapply(1:n.year,gen)
```

Most trips (~70%) have no catch (Fig. 1D) and trips with >= 5 Cabezon are very infrequent (~0.1%).

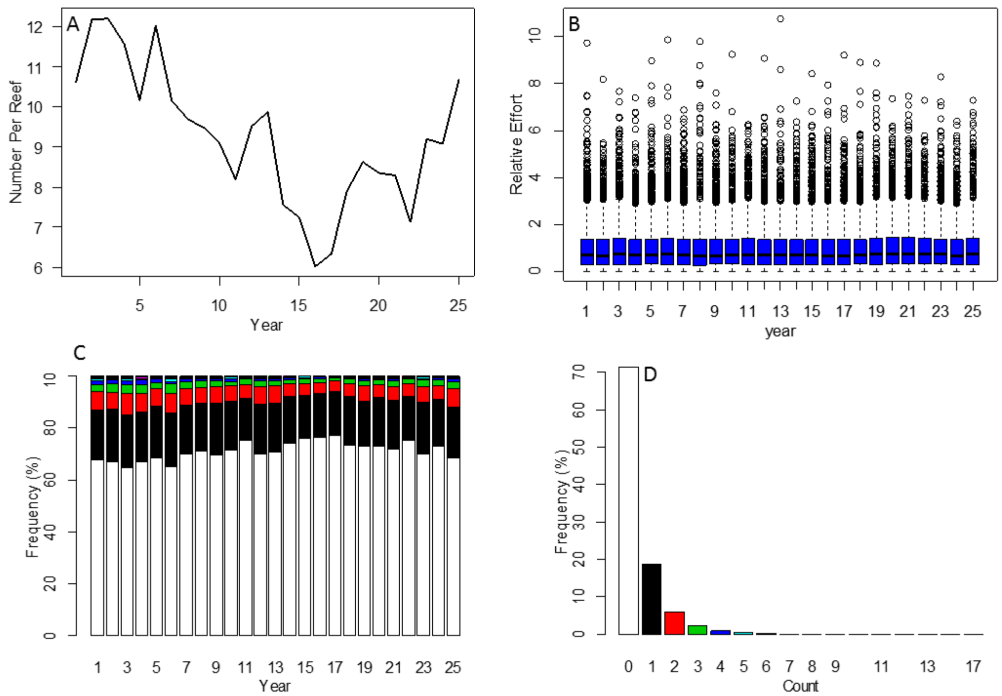

Figure 1. Panel A: Cabezon number per reef versus year. Panel B: Angler effort relative to the mean. Panel C: Frequency distribution of catch per year. Panel D: Frequency distribution of catch for all years.

If the catch counts are not otherwise affected by any management regulations then the stock size index is best estimated using a NB GLM with a year effect and log(effort) as an offset:

```
fit = glm.nb(y ~ fyear-1+offset(log.effort), data = dat)
est.mu = coef(fit)
newd = data.frame(fyear=factor(1:n.year))
newd$log.effort=0

est.mu = predict(fit,newdata=newd,se.fit=TRUE,type = "link")
est.mu$est = exp(est.mu$fit)
est.mu$L = exp(est.mu$fit + est.mu$se.fit*qnorm(0.025))
est.mu$U = exp(est.mu$fit + est.mu$se.fit*qnorm(0.975))
est.mu$true = mu
est.mu$year = 1:n.year
```

The estimated abundance index tracks the true population value, and the 95% confidence intervals (CIs) usually contain the true value (Fig 2.)

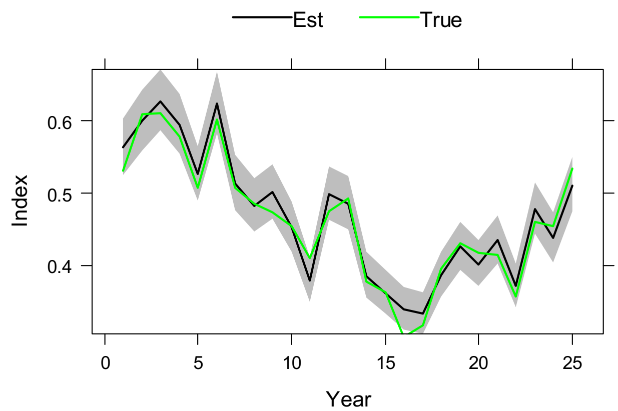

Figure 2. Estimated abundance index (black lines) with 95% CIs (shaded regions), and the true population index (green lines).

Note that it is important to treat effort as an offset and not to fit CPUE ≡ catch/effort. With the latter, you lose information about zero catches with low effort versus zero catches with high effort. These are different observations that provide different information about abundance, especially when there are many zeroes, which seems to be the case for Cabezon. Also, effort saturation is easier to examine with the offset approach.

If trip limits affect the catch, then censored regression should be used. To illustrate this, I modify the simulated data based on a trip limit of 5 fish prior to year 15, and a trip limit of 2 since year 15. Any catch >= the trip limit is set equal to the trip limit each year. The R script is

```
dat$yz=dat$y
dat$cens = 0
dat$cens[dat$year<15]=5
dat$cens[dat$year>=15]=2
ind = dat$y>=dat$cens
dat$yz[ind]=dat$cens[ind]
```

Fig. 3 shows the frequency distribution of the censored catch (i.e. dat$yz).

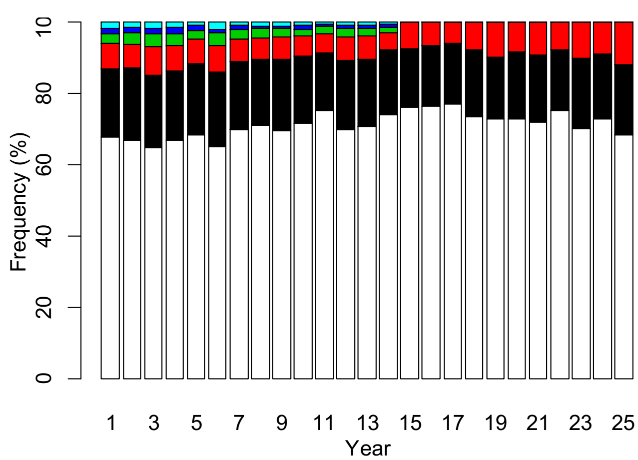

Figure 3. Frequency distribution of censored catch per year because of trip limits.

R does not have a NB censored GLM package as far as I can tell, but it is not difficult to create one. Some basic R script for this is:

```
dat$iyear = as.numeric(dat$fyear)
ind = dat$y>=dat$cens; # select censored catches

# create parameter starting values
start.parm = c(log(tapply(dat$y/dat$effort,dat$year,mean)),log(1))

# function that returns the negative log-likelihood (nll) for uncensored and
# censored catches
cens.nll = function(parm){
 parm1 = parm[1:(length(parm)-1)]
 kp = exp(parm[length(parm)]); #the NB k parameter
 mup = exp(parm[dat$iyear] + dat$log.effort); # the GLM
 #nll for uncensored observations;
 nll = -sum(dnbinom(dat$y[!ind], size=kp, mu=mup[!ind], log = TRUE))
 #nll for censored observations;
 nll = nll - sum(pnbinom(dat$yz[ind]-1, size=kp, mu=mup[ind],lower.tail =
FALSE,log = TRUE))
 return(nll)
}

opt = nlminb(start.parm,cens.nll); #get the MLE's

parm1 = opt$par[1:(length(opt$par)-1)]
est.mu$cens.est = exp(parm1)

hess = hessian(cens.nll, opt$par)
parm.se = sqrt(diag(solve(hess)));#hessian-based sd.err for log year effects;
est.mu$cens.est.L = exp(parm1 + parm.se[1:(length(opt$par)-1)]*qnorm(0.025))
est.mu$cens.est.U = exp(parm1 + parm.se[1:(length(opt$par)-1)]*qnorm(0.975))
```

The censored regression model results (Fig. 4) in an abundance index that tracks population abundance well, whereas the GLM that does not account for the censored effects of trip limits results in a biased index.

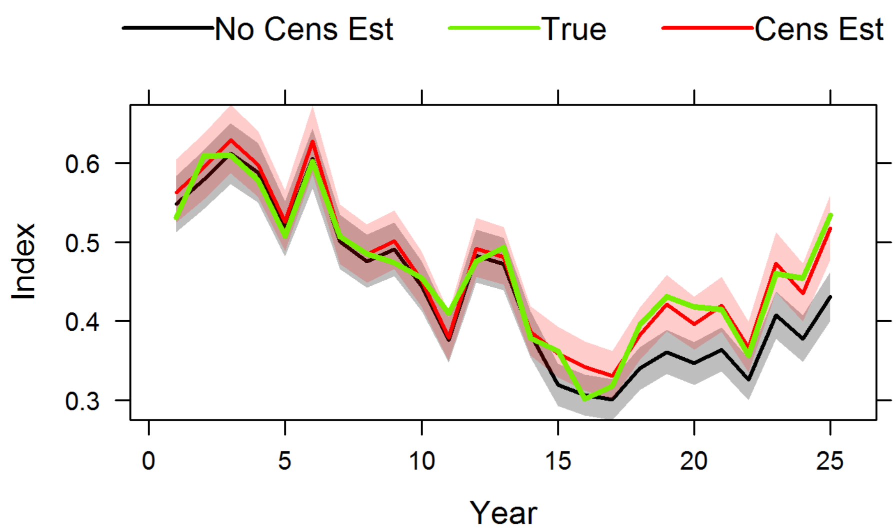

Figure 4. NB GLM (without censoring) estimates of trends in Cabezon abundance (black lines) with 95% Cis (grey shaded regions), censored NB GLM (red lines and shaded regions), and the true population index (green lines)

I have run this code about 20 times and the censored index estimation performed well in all cases. The code took just a few minutes to run on a laptop. This code could be modified to include additional factors (e.g. month, gear) in the GLM. The simulated data included a change in trip limits. However, even if there is no change in trip limits, censoring will still be a problem especially when the percent of trips hitting the limit changes with time.

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

I concur with the STAR RP that CA, OR, and WA assessments for cabezon constitute the best available scientific information on the current status of the stock(s) and that the assessments provides a suitable basis for management decisions.

ToR 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.

I have provided suggestions for improvements under ToR 2 and 3. In this section I summarize these. I first provide RP recommendations for future research and data collection, and expand on these where relevant. I think these recommendations are all for longer-term time frames or the next assessment.

All areas

1. All assessments need to deal with catch uncertainty. *I suggest* that there is an important need for assessment teams to provide information on the quality of the annual landings estimates, and more specifically to quantify the uncertainty in these estimates. Also, this information needs to be included in the stock assessment models.

2. Evaluate the way in which the likelihood is calculated for recruitment deviations to properly capture the nested random effect structure that should be present.

3. Stock structure issues need to be addressed as current spatial scales seem too large given the life history of the species. Issues on localized depletion cannot be addressed as currently formulated. *My additional comments*: Cabezon appears to be fairly sedentary species so localized depletion could be a problem. My sense from the assessments is that Cabezon have a broad-scale and somewhat homogenous spatial distribution at low densities. They are not a shoaling species that forms dense aggregations at various times and places. Their near-shore distribution makes them difficult to survey. They have low mobility apart from larval stages, and therefore localized depletion may be a fisheries management problem, but this may also be a stock assessment opportunity. It may be possible to estimate local stock size based on depletion information that may be available from high-resolution (time and space) catch and effort monitoring. Depletion may only occur at fairly fine spatial scales, and may be obscured by fleet movements at broader spatial scales. Therefore, a spatiotemporal depletion model may be required (e.g. see Cadigan et al., 2017). This can also be coupled with spatial survey information and also censored regression methods to deal with trip limits.

    - Cadigan, N.G., Wade, E. and Nielsen, A., 2017. A spatiotemporal model for snow crab (*Chionoecetes opilio*) stock size in the southern Gulf of St. Lawrence. Canadian journal of fisheries and aquatic sciences, 74(11), pp.1808-1820.

4. Small scale tagging studies may inform range of their movement and appropriate scale for substock structure. *My additional comments*: M seems like a main driver of Cabezon productivity and future research (e.g. data storage or telemetry tagging) should be conducted to provide more reliable values for Cabezon. If acoustic arrays have been established around marine protected areas then telemetry tagging combined with survival analysis may be an inexpensive way to collect direct information on M, at least for the tagged population.

5. Recruitment models describing nesting/guarding/territorial behavior should be attempted.

6. Examine catch and effort data at finer spatial scales as time/data permits in the future to examine localized depletion issues. *See my comments for 3 above*.

7. Develop fishery independent surveys to better understand stock dynamics to use as an independent data source in model fitting. *My additional comments*: The most important deficiencies of the SCS and OR assessments was the lack of reliable fishery-independent indices of abundance. It seems difficult to survey this near-shore species. The CCFRP index is a merit for the NCS assessment and a recommendation is to consider extending this survey methodology to the SCS and OR stock areas.

8. Develop methods to include recruitment variability in uncertainty intervals for historic stock size and SBo. *My additional comments*: The CA and OR models did not include recruitment deviations in historic periods that only have catch estimates. This results in false precision about estimates of historic stock size and especially about spawning depletion. It also resulted in some lack of fit to historic indices (i.e. SCS and NCS

CPFV). This is a deficiency in the application of SS3. This should be further investigated in future assessments.

9. Consider developing a tagging program to understand the spatial extent of the localized populations at different stages of their life cycle. *See my comments for 3 above*.

10. Future assessments should re-visit standardization of the California Collaborative Fisheries Research Program Nearshore Survey and provide more diagnostics. *My additional comments*: This was a stratified random survey and yet strata was not taken into account in the index standardization. I also suggest that the delta-GLM standardization model needs additional investigation (see ToR 4).

California

1. California Collaborative Fisheries Research Program Nearshore Survey: Future assessments should re-visit standardization and provide more diagnostics. But for this assessment re-analyzing this does not seem to be high priority. *See my comments for 10 above*.

2. SCS needs to collect some data on indices of abundance and life history to update their series. At the very least start collecting some better length composition data as well. *My additional comments*: There is little recent data to base an assessment of this stock on.

Oregon

1. Finer spatial scales could be explored with the indices currently being developed and used. If one area is used, an effort to get a single signal analyzing different set/trip level data across all fleets should be attempted.

2. A 2-3 area spatial assessment should be explored to address the differences in indices and length composition data being collected at different locations. Perhaps a finer resolution time step could also be developed.

3. A better understanding of the MRFSS index and the appropriate level of error to represent uncertainty should be examined. *My additional comments*: The standard errors for the MRFSS index were really small. The standardization resulted in an index that differed substantially from trends in raw CPUE, which required further investigation to check if year effects have been partially confounded with some other covariate effects.

4. There is some evidence of different year-classes evident in commercial live length compositions recently compared to Recreational boat compositions. This is something to keep in mind, but at present there is no way to address this.

Washington

Length at age data should be examined for growth and a more integrated model developing an index of abundance could be developed for WA in the future. Tagging studies to inform movement could also be conducted and genetic data examining stock structure could be addressed.

My Additional Research Recommendations

All areas

1. Investigate improved methodologies for accounting for the impacts of bag- and trip-limits when deriving abundance indices from fishery-dependent sampling programs. I outlined a censored-regression possibility under ToR 4.

2. Delta-GLM's were used to provide abundance indices from fishery-dependent and independent (i.e. CCFRP) sampling programs. This seemed to be an unnecessary complication to me, because Cabezon is not a shoaling species with infrequent large catches that the delta-GLM's are designed to accommodate. Alternative and simpler approaches should be explored, including the Negative-Binomial approach I outlined under ToR 4. Effort should be treated as an offset in a catch GLM.

3. Index standardization is clearly a merit; however, in future I suggest that the distribution of covariates over years (i.e. via boxplots, etc.) be examined. If there is a pronounced change in a covariate distribution over years, then it is possible that years effects (the abundance index) could be partially confounded with other covariate effects.

4. CPUE standardizations when there are no onboard observers to determine if trips were targeting the species of interest used an approach published by MacCall and Stevens to determine which trips were targeted. It is useful to examine the proportion of trips deleted to examine if there is a time trend in zero-trips removed that could be confounded with changes in abundance.

5. Provide plots of annual length composition sample size per ton of catch for each fleet and stock. This will provide useful information on the adequacy of coverage.

6. Provide some type of annual and fleet-specific maps of the locations of trips sampled for lengths.

7. Collect more sex-specific length compositions for all fleets where possible. If a simple morphometric measurement (like color, etc.) could be collected, and some accurate sex-frequency by morphometric sampling conducted, then it might be possible to infer sex-based length compositions for each fleet.

8. Continued collection of aging data will improve these assessments.

9. Conduct a GLM to investigate for statistical differences in length-weight relationships across stocks, with the objective of combining information where appropriate. It is also useful to check if the relationship has varied over years.

10. Contemporary maturity data should be collected to check if maturity schedules have changed.

11. There are some curious differences in the SS3 estimates of size at age between the NCS and ORS that should be further investigated in future assessments.

12. There is borrowing of information between the SCS and NCS assessments (i.e. VonB k). There may be other parameters that we may also expect to be similar between models, including the OCS model. In the long-term, explore assessment models that facilitate this more directly (e.g. Punt et al., 2011).

Punt, A. E., Smith, D. C., and Smith, A. D. M. 2011. Among-stock comparisons for improving stock assessments of data-poor stocks: the "Robin Hood" approach. – ICES Journal of Marine Science, 68: 972–981.

Oregon

1. Recreational bag-limits seem to have changed over time, and it was not clear to me why this factor was not considered in the Onboard Observer and ORBS index standardizations. This should be considered in future assessments.

2. The commercial live fleet had some curious residual patterns that seemed opposite for males and females, and this should be explored further in future assessments.

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

The RP report first presented an overview of the proceedings and recommendations of the RP. In particular, the RP recommended that "all four substock assessments for cabezon constitute the best available scientific information on the current status of the stock(s) and that the assessments provides a suitable basis for management decisions. The panel further recommends that the SCS, NCS, and ORS assessments be considered category 1 (full integrated assessments), and that the data-poor WAS assessment be considered category 3".

The RP report provided a summary of data and assessment models, and this text was provided by the STAT. This included a description of notable updates to data sources and descriptions of assessment model details including changes from the previous 2009 assessment model and parameter choices for reference models.

Most of the RP report involved RP requests and STAT responses. In summary, these were:

North California Substock (NCS)

1. Update reference models by re-weighting the alternative catch histories in the SCS and NCS models. *Conclusion*: No need to revise weighting at this time.
2. Provide a run with uninformed priors on growth curves. *Conclusion*: Marginal differences in assessments as priors on k have little influence. The reference model should proceed without priors on k.
3. Produce r4ss plots for the no time blocking run on selectivity and calculate AIC for both models. *Conclusion*: The blocking of selectivity is appropriate.
4. Provide comparative plots of prior and posterior values of M. Also provide posterior plots for five derived sensitivity quantities (B0, B2019, depletion, F SPR, MSY at SPR). *Conclusion*: The posteriors are approximately symmetrically distributed around the

median (except MSY SPR0.45). The data is shifting the posterior distribution of M lower than the prior, but posterior mean value appears to be biologically reasonable.

5. Provide a low M run (M = 0.1) and provide diagnostics. *Conclusion*: The low M run resulted in a substantially lower negative log-likelihood for the CPFV index. The improvement in fit to that index was concentrated on two time-periods.
6. Investigate alternative model formulations that address overfitting of the CPFV survey with low M. *Conclusion*: All reductions in weights for the CPFV index resulted in higher M estimates. However, these models did not result in much difference in SSB or status estimates. The down-weighting seemed too subjective and the RP recommended using the reference model weights; however, we should document that the apparent assessment information about M in the CPFV index does not seem real.
7. Provide a plot of yearly length composition likelihood values for the recreational boat fleet in the NCS model. *Conclusion*: Length composition likelihoods get worse over time.
8. Provide a plot by year of the age selectivity 2 vector. *Conclusion*: Selectivity at age is a function of selectivity at length multiplied by length at age for selectivity at age. This changes as selectivity is time blocked for 3 periods.
9. Re-calculate the M prior as a lognormal and present the model fits and diagnostics. Rationale: To explore the possibility of a traditional decision table that does not require an MCMC approach. Conclusion: The use of the log normal M fixes the issues with unrealistic stock status trajectories with unrealistically low M's creating low stock status conditions.
10. Run two models: fix female M's at 0.2 and 0.33 (these values come from the Bayesian posterior distribution shown to the panel). Alternatively, use the male M offset from the reference model and estimate the male M offset. Provide model fits and diagnostics. Response: Unnecessary given results of request #9.

South California Substock (SCS)

1. Assume the alternative catch histories for the SCS (and NCS) models and then: a) force the rec. shore fleet to have logistic selectivity; b) fix the selectivity to the outcome in 1a; and c) remove the rec. shore length compositions and redo the M profile. *Conclusion*: While rec shore length compositions are influential, they do not affect the overall M profile.
2. a) Cap the data weighting for the rec. shore fleet to 1; b) modify effective N in the rec. shore fleet by removing samples less than 5 or only exclude samples with very tight intervals; c) redo the M profile; and d) provide the comp. plots and diagnostics. *Conclusion*: While rec shore fleet has an effect it is quite marginal on estimates of M.
3. Recalculate the M prior as a lognormal and present the model fits and diagnostics. Rationale: To explore the possibility of a traditional decision table that does not require an MCMC approach. Conclusion: The use of the log normal M fixes the issues with unrealistic stock status trajectories with unrealistically low M's creating low stock status conditions.

Oregon cabezon Substock (OCS)

1. Remove the year:boat type interaction term in the ORBS index and re-compute the index. In the short term, assume the CVs from the current configuration (recalculate CVs for the post-STAR version of the base model). Compare the resulting indices and synthesis

results. *Conclusion*: The effect of the index is marginal at best and changes in standardization do not affect the overall trajectory much.

2. Provide comparative biomass and depletion trends for jittered models within 2 negative log likelihoods. *Conclusion*: The new versions of the model are not creating any divergent dynamics.
3. Provide a model on the commercial live-fish fleet with another time block in 2015. *Conclusion*: While the apparent misfit for the later periods is fixed with time varying selectivity, the AIC values do not suggest adding more parameters.
4. a) Provide the fleet-specific age and length likelihood profiles over M. b) Down-weight age components leading to high M; re-run the model while freely estimating M (a model with both sexes freely estimated and a model with only the male offset freely estimated) and growth parameters. c) If there are age components identified leading to high M that does not fix the aberrant growth patterns seen, fix the CVs at older ages (using the NCS model), to see if that makes the growth patterns more reasonable. *Conclusion*: None of these issues fixed either the estimate of M or the CV estimated at older ages.
5. a) Estimate female and male M with a more informed Hamel prior (female M = 0.318; CV = 0.219). Provide diagnostics and profiles over M. b) If the results for request 5a prove unsatisfactory, make M more consistent with the NCS model (0.25 females, 0.318 males at Hamel prior). *Conclusion*: None of these fixed the estimation of M.
6. Increase the CV on the MRFSS index to a level consistent with the other rec. indices. *Conclusion*: Freeing up the CV's gave the model more flexibility to capture dynamics that are more in line with what field data suggests.
7. Provide plots of lengths and associated ages by fleet and sex. *Conclusion*: Uncertain whether the age length data represents the live capture fleet. Also not sure whether these growth curves differ by fleet.
8. Present an updated reference model that includes the fixed ORBS index, down-weighting the MRFSS index (added CV), and fix female M at 0.240 and male M at 0.280 consistent with the NCS model). Provide plots and diagnostics. *Conclusion*: Represents the realistic dynamics as suggested by STAT. However the next analysis makes it more evenly distributed.
9. Explore runs for the low and high states of nature by deriving the 12.5% and 87.5% quantiles of 2019 biomass to estimate ranging M values in the ORS model. If that does not provide adequate contrast, use the low and high M's from the NCS model. *Conclusion*: This set of graphs is more appropriate given the CI's overlap.

Washington Substock (WAS)

1. Aggregate length in years 2014-2018 and use that one length composition in LB-SPR to estimate depletion and the associated uncertainty. *Conclusion*: This does not change the state of depletion for the stock. And in addition selectivity is constant for the time period examined.
2. Use male growth values to compute LB-SPR to estimate depletion and the associated uncertainty. *Conclusion*: This does not work as selectivity and F is related to females and not males.

The STAR panel and STAT agreed that the major axis of uncertainty for the NCS, SCS, and OSC assessments was in natural mortality (M). Uncertainty in spawning biomass was used to determine natural mortality values to bracket the uncertainty around the reference model. Quantile values of 12.5% (low state) and 87.5% (high state) for 2019 spawning biomass were calculated from the asymptotic estimates of error from the reference model. For the WAS assessment, three catch scenarios and five relative stock status values were explored for OFL calculation using SSS.

The RP report included a description of the technical merits of the assessment, itemized by substock. There was also a description of the technical deficiencies of the assessment, itemized by issues of data inputs and models common to the CA and OR assessments, and then issues specific to substocks.

There were no disagreements among the GAP, GMT, and PFMC representatives with respect to STAR panel recommendations. There were no disagreements among the stock assessment teams with respect to STAR panel recommendations.

The RP report included a description of critical uncertainties, which involved a) incomplete understanding of the basic reproductive life history processes of Cabezon, b) it is sedentary species so localized depletion could be a problem, c) uncertainty in interpreting fisheries catch rate and size composition information, d) the lack of a dedicated age-sampling program, and e) the potential for more sub-stock structure than is reflected by the current assessment region boundaries.

The RP report finished with a list of research recommendation relevant for all sub-stocks, and recommendations that were specific to each of the four substocks. These recommendations are reviewed (and expanded on) under ToR6 above.

Conclusions and Recommendations

Recommendations are provided under ToR 6.

ToR 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.

I found overall that the documentation of the stock assessment inputs, methods, and results were very helpful. As usual for US stock assessments, I found the description of Cabezon fisheries and their management, including implications for stock assessments, to be very good. Sections that described the implications of changes made from the last assessment and particularly the sensitivity analyses were very helpful.

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

The historical fishery catch reconstruction back to 1916 (the first year of required reporting in the commercial fishery) and by sub-stock uses the same approach as in the 2009 assessment with new catches being added to complete the time series, and also some revision to the CA recreational catch data during 2004-2008. A significant change in the catch data involved the reallocation of CA catches from the SCS region to the NCS region for years 1980-1995. The reference model and other models requested by the RP used the updated catch series. The effect on this change was proportionally greater for the SCS, where there is less catch data overall than in the NCS.

The STAT created long time-series of Cabezon landings which is meritorious. The accuracy of estimates of landings and low discards has improved over time, as expected. This is also a merit. A related deficiency is that there is uncertainty in catch estimates, and moreso for historic periods, that is not measured and provided to the RP. It is also not clear that the SS3 assessment modelling approach could utilize information on catch uncertainty. There is an important need for assessment teams to provide information on the quality of the annual landings estimates, and more specifically to quantify the uncertainty in these estimates. This is difficult to do because landings estimates involve measurement error and some bias.

Length Compositions were the most influential data in the CA and OR stock assessments. Most of the assessment model fleets had at least some length composition samples – and for some fleets and years the sample sizes seemed good. At various times during the RP it was mentioned by participants that there was some geographic separation in various fleets that might affect interpretation of the length composition samples. A deficiency involved with this data was that the RP was not provided with information on this.

A deficiency in the CA data was that no sex-specific compositions were available. Sex-based sampling was only available for the commercial fleets in OR. The SCS stock has limited recent removals and also size composition information. This will be a problem for future assessments of this stock.

The lack of regular age composition information is a deficiency for the CS, OR, and WA assessments. There is no aging data in the SCS assessment, and very little in the NCS assessment.

The most important deficiencies of the SCS and OR assessments was the lack of reliable fishery-independent indices of abundance. It seems difficult to survey this near-shore species. Several fisheries dependent indices were developed for the OR stock, which is a merit for this assessment. A deficiency is that the STAT team did not seem to trust these indices, except to indicate broad-scale trends in abundance.

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

Stock boundaries were based on pragmatic considerations of spatial differences in biogeographic shifts in life history traits and data availability. Indices of abundance and length/age composition data were aggregated at the SCS, NCS, OR and WA spatial units, balancing the need for spatial resolution versus data availability and quality. However, stock structure research suggests there is more sub-stock structure than is reflected by the current assessment region boundaries.

Cabezon appears to be fairly sedentary species so localized depletion could be a problem.

The two sexes are modeled separately, as growth is very different between females and males. This seemed appropriate to me.

I conclude from the review meeting that the SS3 model was very competently applied. The skill of the STAT with SS3 was a strong merit. The CA and OR models did not include recruitment deviations in historic periods that only have catch estimates. This results in false precision about estimates of historic stock size and especially about spawning depletion.

The natural mortality rate (M) was assumed to be age and year invariant, but different for males and females. Given the data available, this seemed reasonable to me. The natural mortality tool was used to derive a prior on M which facilitated a more realistic quantification of model uncertainty. The STAT and RP closely examined specification of M for the CA and OR models. The RP and STAT agreed that the major axis of uncertainty involved M. I conclude that uncertainty about M is adequately accounted for. M seems like a main driver of Cabezon productivity and future research (e.g. data storage or telemetry tagging) should be conducted to provide more reliable values for Cabezon.

The CA and OR assessments assume a Beverton-Holt stock-recruit relationship and the steepness formulation, which is typical in US stock assessments. Steepness (0.7) and recruitment variability (sd=0.5) were fixed to the same values as in the previous assessment. This approach was reasonable.

The STAT did a really good job before the RP in performing a wide range of sensitivity analyses (both to data and structural model uncertainty) and documenting the results in succinct and easy to understand graphical displays.

Retrospective analysis for the CA stocks did not indicate estimation problems. There were appreciable differences from the previous assessment 10 years ago but a substantial amount of new data (especially for NCS) has been collected since 2009 and I do not find the retrospective differences in depletion to be a problem.

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

I have summarized suggestions for improvements under ToR 6.

I think the delta-GLM standardization methods for fishery-dependent indices was not appropriate for Cabezon, nor were the effects of bag- or trip-limits appropriately accounted for. Cabezon is not a shoaling species that a delta-GLM approach would be appropriate for. I demonstrate using simplified Cabezon-like simulated data how to use censored regression methods to derive an unbiased abundance index when trips limits are meet infrequently (~5%)

Including trip limit regulations as a covariate is a deficiency. This is not the way to address the censoring of catches because of trip limits. I illustrated a better approach using simulated Cabezon-like data.

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

I concur with the STAR RP that CA, OR, and WA assessments for cabezon constitute the best available scientific information on the current status of the stock(s) and that the assessments provides a suitable basis for management decisions.

The pre-panel efforts by the STAT in documenting the assessments and sensitivity analyses, and the STAT's remarkable progress in responding to requests from the RP, were the primary factor that led to a smooth and complete RP meeting.

ToR 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.

I have summarized suggestions for improvements under ToR 6. The main ones are

- Provide information on the quality of the annual landings estimates, and more specifically to quantify the uncertainty in these estimates
- Stock structure issues need to be addressed as current spatial scales seem too large given the life history of the species.
- Develop fishery independent surveys to better understand stock dynamics to use as an independent data source in model fitting.
- Continued collection of aging data will improve these assessments.
- SCS needs to collect some data on indices of abundance and life history to update their series.
- Investigate improved methodologies for accounting for the impacts of bag- and trip-limits when deriving abundance indices from fishery-dependent sampling programs.

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

The Review Panel (RP) report first presented an overview of the proceedings and recommendations of the RP. The report provided a summary of data and assessment models, and this text was provided by the STAT. Most of the report involved RP requests and STAT responses. The STAR panel and STAT agreed that the major axis of uncertainty for the NCS, SCS, and OSC assessments was in natural mortality (M). The RP report included a description of the technical merits and deficiencies of the assessments. The report included a description of critical uncertainties. The RP report concluded with a list of research recommendations.

Appendix 1: Bibliography of materials provided for review

Draft and Background Documents
Stock Assessment Review (STAR) Panel 2 for Bocaccio and China rockfish

Meeting Materials:

STAR Panel Meeting Proposed Agenda

Terms of Reference for the Groundfish and Coastal Pelagic Species Stock Assessment and Review Process for 2019-2020. Pacific Fishery Management Council. APRIL 2019.

Draft Stock Assessment Documents

Cope, J.M., Berger, A.M., Whitman, A.D., Budrick, J., Bosley, K.M., Tsou, T., Niles, C.B., Privitera-Johnson, K., Hillier, L.K., Hinton, K.E., and Wilson, M.N. 2019. Assessing Cabezon (*Scorpaenichthys marmoratus*) stocks in waters off of California and Oregon, with catch limit estimation for Washington State. Pacific Fishery Management Council, Portland, OR. Available from http://www.pcouncil.org/groundfish/stock-assessments/

Background Materials

Cope, J.M., Key, M., 2009. Status of Cabezon (*Scorpaenichthys marmoratus*) in California and Oregon Waters as Assessed in 2009. Pacific Fisheries Management Council, Portland, OR.

CABEZON STAR Panel Report. July 27-30, 2009

Maguire, J-J. Report on the cabezon and lingcod Stock Assessment Review (STAR) Panel July 27 - 31, 2009. Seattle, WA

Smith, S. Report for the Center of Independent Experts on the Stock Assessment Review (STAR) Panel for Cabezon and Lingcod (July 27 to 31, 2009).

Appendix 2: CIE Statement of Work

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

Stock Assessment Review (STAR) Panel 1

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 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 scientific information available and facilitate the use of this information by the Council to adopt Overfishing Limits (OFLs), Acceptable Biological Catches (ABCs), Annual Catch Limits (ACLs), harvest guidelines (HGs), and annual catch targets (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 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.

Benchmark stock assessments will be conducted and reviewed for cabezon in California and Oregon waters, and a data-limited approach may be considered for Washington waters. cabezon stocks were identified as strong candidates for assessment during the Pacific coast groundfish regional stock assessment prioritization process, which was based on the national stock assessment prioritization framework (http://www.st.nmfs.noaa.gov/Assets/stock/documents/PrioritizingFishStockAssessments_FinalWeb.pdf.

Previous assessments were conducted for cabezon stocks in California waters and Oregon waters in 2009. The assessments estimated stock depletion of 48.3 percent of unfished biomass at the start of 2009 for the combined California substocks, and 52.4 percent depletion for the Oregon stock (Cope and Key 2009). These 2009 assessments are now outdated and no longer provide accurate forecasts for future management. Assessments for these stocks are needed to provide the basis for the management of the groundfish fisheries off the West Coast of the U.S. including providing scientific basis for setting Overfishing Limits (OFLs) and Acceptable Biological Catches (ABCs) as mandated by the Magnuson-Stevens Act. The technical review will take place during a formal, public, multiple-day meeting of fishery stock assessment experts. Participation of external, independent reviewer is an essential part of the review process. The specified format and contents of the individual peer review reports are found in **Annex 1**. The Terms of Reference (ToRs) of the peer review are 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 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 "consistent" CIE reviewer will participate in all STAR panels held in 2019 and the PWS and ToRs for the "consistent" 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-structured models, use of *Markov Chain Monte Carlo (*MCMC) to develop confidence intervals, and use of Generalized Linear Models in stock assessment models. Each 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.

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 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 1 meeting include:

- The current draft stock assessment reports;
- 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 (including previous stock assessments and STAR panel reports).
- An electronic copy of the data, the parameters, and the model used for the assessments (if requested by reviewer).

Panel Review Meeting: The CIE reviewers 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. Each CIE reviewer shall actively participate in a professional and respectful manner as a member of the meeting review panel, and their peer review tasks shall be focused on the ToRs as specified herein. The NMFS Project Contact is responsible for any facility arrangements (e.g., conference room for panel review meetings or teleconference arrangements). The NMFS Project Contact is responsible for ensuring that the Chair understands the contractual role of the CIE reviewers as specified herein. The CIE 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 reviewers 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**. Each CIE reviewer shall complete the independent peer review addressing each ToR as described in **Annex 2**.

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

Timeline for CIE Reviewers

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

1) Conduct necessary pre-review preparations, including the review of background material and reports provided by the NMFS Project Contact in advance of the peer review.
2) Participate during the STAR Panel 1 review meeting in scheduled in Newport, Oregon during the dates of May 6 -10, 2019 as specified herein, and conduct an independent peer review in accordance with the ToRs.
3) No later than May 24, 2019, each CIE reviewer shall submit their draft independent peer review report to the contractor. Each CIE report shall be written using the format and content requirements specified in **Annex 1**, and address each ToR in **Annex 2**

Foreign National Security Clearance

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

Place of Performance

The place of performance shall be at the contractor's facilities, and in Newport, OR.

Period of Performance

The period of performance shall be from the time of award through July 2019. The CIE reviewers' duties shall not exceed 14 days to complete all required tasks.

Schedule of Milestones and Deliverables

The contractor shall complete the tasks and deliverables in accordance with the following schedule.

Within two weeks of award	Contractor selects and confirms reviewers
At least two weeks prior to the panel review meeting	Contractor provides the pre-review documents to the reviewers
May 6-10, 2019	Each reviewer participates and conducts an independent peer review during the panel review meeting
May 24, 2019	Contractor receives draft reports
June 7, 2019	Contractor submits final reports to the Government

Applicable Performance Standards

The acceptance of the contract deliverables shall be based on three performance standards: (1) The reports shall be completed in accordance with the required formatting and content in **Annex 1**; (2) The reports shall address each ToR as specified **Annex 2**; and (3) The reports shall be delivered as specified in the schedule of milestones and deliverables.

Travel

All travel expenses shall be reimbursable in accordance with Federal Travel Regulations (http://www.gsa.gov/portal/content/104790). International travel is authorized for this contract. Travel is not to exceed $7,600.

Restricted or Limited Use of Data

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

NMFS Project Contacts:

Stacey Miller, NMFS Project Contact
National Marine Fisheries Service,
2032 SE OSU Drive
Newport, OR 97365
Stacey.Miller@noaa.gov
Phone: 541-867-0535

Jim Hastie
National Marine Fisheries Service,
2725 Montlake Blvd. E,
Seattle WA 98112
Jim.Hastie@noaa.gov
Phone: 206-860-341

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 1

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.
2. Discuss the technical merits and deficiencies of the input data and analytical methods during the open review panel meeting.
3. Evaluate model assumptions, estimates, and major sources of uncertainty.
4. Provide constructive suggestions for current improvements if technical deficiencies or major sources of uncertainty are identified.
5. Determine whether the science reviewed is considered to be the best scientific information available.
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.
7. Provide a brief description on panel review proceedings highlighting pertinent discussions, issues, effectiveness, and recommendations.

Annex 3: Agenda

Proposed Agenda

Stock Assessment Review (STAR) of Cabezon in CA, OR, and WA

NOAA Fisheries, Northwest Fisheries Science Center
Barry Fisher Building, Conference Room 101
2032 SE OSU Drive
Newport, OR 97365

May 6-10, 2019

The STAR panel meeting will also be streamed online for those who want to follow the proceedings remotely. *To join the webinar in listen-only mode, visit this link:* *https://nwfscfram.webex.com/nwfscfram*

1. *Enter the Webinar Meeting Number: 628-714-189*
2. *Please enter your name and email address (required)*
3. *AFTER logging in to the webinar, dial the TOLL number 1-650-479-3208.*

NOTE: You will be muted upon entry to the webinar. For technical assistance, please contact Stacey Miller via email (Stacey.Miller@noaa.gov) or by phone: (541-867-0535).

Monday, May 6

8:30 a.m.	Welcome and Introductions
8:45 a.m.	Review the Draft Agenda and Discuss Meeting Format (Chair)
	- Review the Terms of Reference (TOR) for assessments and STAR panel responsibilities
	- Assign reporting duties
	- Agree on time and method for accepting public comments
9:15 a.m.	Overview of Cabezon Assessments
10:30 a.m.	Presentation of the CA Cabezon Assessment - Data Inputs
	- STAR Panel Q & A
11:30 a.m.	Presentation of the CA Cabezon Assessment – Modeling Approach
	- STAR Panel Q & A
12:30 p.m.	Lunch
1:30 p.m.	STAR Panel Discussion of CA Cabezon Assessment
	- Panel develops written request for additional model runs / analyses
3:00 p.m.	Presentation of the OR Cabezon Assessment – Data Inputs
	- STAR Panel Q & A
4:00 p.m.	Presentation of OR Cabezon Assessment - Modeling Approach
	- STAR Panel Q& A

5:30 p.m. Adjourn for day

Tuesday, May 7

8:30 a.m. Review of Agenda Topics for the Day

8:45 a.m. STAR Panel Discussion of OR Cabezon Assessment
- Panel develops written request for additional model runs / analyses

10:30 a.m. Presentation of the WA Cabezon Assessment

12:30 p.m. Lunch

1:30 p.m. STAR Panel Discussion of the WA Cabezon Assessment

3:00 p.m. Presentation of the First Set of Requested Model Runs for the CA Cabezon Assessments
- Q&A session with the CA Cabezon STAT & Panel discussion
- Panel develops request for second round of model runs / analyses

5:30 p.m. Adjourn for day

Wednesday, May 8

8:30 a.m. Review Agenda for the Day

8:45 a.m. Presentation of the First Set of Requested Model Runs for the OR Cabezon Assessment
- Q&A session with the OR Cabezon STAT & Panel discussion
- Panel develops request for second round of model runs/analyses for the OR Cabezon Assessment

11:00 a.m. Presentation of the Second Set of Requested Model Runs for the CA Cabezon Assessments

12:30 p.m. Lunch

1:30 p.m. Continue Presentation of Second Set of Request Model Runs for CA Cabezon Assessments
- Q&A session with the CA Cabezon STAT & Panel discussion
- Panel develops request for third round of model runs / analyses

4:30 p.m. STAR Panel Discussion / Begin Drafting Report

5:30 p.m. Adjourn for day

Thursday, May 9

8:30 a.m. Review Agenda for the Day

8:45 a.m. Presentation of the Second Set of Model Runs for the OR Cabezon Assessment
- Q&A session with the OR Cabezon STAT & panel discussion
- Panel develops request for third round of model runs / analyses

10:30 a.m. Presentation of the Third Set of Model Runs for the CA Cabezon Assessments
- Q&A session with the CA Cabezon STAT & panel discussion
- Agreement of the preferred model and model runs for the decision table
- Panel continues drafting the STAR report

12:30 p.m. Lunch

Thursday, May 9 (Continued)

1:30 p.m. STAR Panel Discussion / Continue Drafting STAR Report

2:30 p.m.	Presentation of the Third Set of Model Runs by the OR Cabezon STAT - Q&A session with the OR Cabezon STAT & panel discussion - Agreement of the preferred model and model runs for the decision table - Panel continues drafting the STAR report.
5:30 p.m.	Adjourn for day

Friday, May 10

8:30 a.m.	Consideration of Remaining Issues - Review decision tables for all assessments
12:30 p.m.	Lunch
1:30 p.m.	Review First Draft of the STAR Panel Report
4:00 p.m.	Panel Agrees to Process for Completing the Final STAR Report for Council's September Meeting Briefing Book (Requested by August 15th)
5:30 p.m.	Review Panel Adjourns