Independent Peer Review Report of the Cabezon STAR Panel

By

Robin Cook

Prepared for
Center for Independent Experts
Independent System for Peer Review

Contents

Executive Summary

i. Assessments of cabezon in Oregon, California, and Washington were reviewed during a formal, public meeting of fishery stock assessment experts from 6-10 May 2019. Two Center for Independent Experts (CIE) reviewers were included in the review panel.

ii. The assessments represent the best science available given the existing data. The limited amount of age data and lack of informative abundance indices means that it is difficult to have high confidence in the estimated stock status. The uncertainty in the California stocks spans the reference points making status determination ambiguous. For Oregon, the confidence intervals on depletion lie above the reference points, which would indicate the stock is not overfished but the intervals probably do not fully capture uncertainty.

iii. The approach to estimating CPUE abundance indices from fishery data was thorough and appears to be the best available. Disappointingly, the resulting indices did not appear to contain much population signal and tended not to contribute much to the estimated stock biomass trajectory. If possible, fishery independent surveys should be developed to calibrate estimates of biomass in the most recent year.

iv. The catch data are influential in the assessment, but are treated as exact and fixed in the model. While this is probably a necessary assumption, it is clearly unrealistic. A demanding sensitivity analysis is required where plausible alternative catch streams are generated stochastically and used to test the model.

v. Priority should be given to the collection and processing of more age samples for all stocks. The issue is greatest for California, especially the southern stock. Data collection needs to be maintained to create a coherent time series of observations.

vi. Natural mortality is the largest component of total mortality in these stocks and will drive much of the stock dynamics. Consideration should be given to modelling M by size using a relationship such at that estimated by Lorenzen and scaled to a mean value given by the Hamel or similar method. This would avoid the need to model M by gender and would capture some of its annual variation.

vii. Thought needs to be given to the appropriate level of model complexity to ensure that the final base model fitted to the data also has the appropriate forecasting properties. A procedure needs to be developed to identify the most parsimonious model using an information statistic and the parameter correlation matrix.

viii. Stock Synthesis software (SS3) provides an impressive range of diagnostics to aid model development. In its present implementation it provides asymptotic variance estimates for the parameters and quantities of interest. This is something of a limitation, as it hinders identifying problematic model fits and understanding the relative contribution of priors and data to the estimates. MCMC runs drawn from the reference models produced more realistic estimates of posterior distributions and should be a routine output of the analyses.

ix. The review meeting was constructive and productive with effective excellent co-operation from the STAT teams. Meeting facilities were good, and the local staff provided great support to the reviewers. There were no major disagreements between Panel members or the STATs.

Background

The National Marine Fisheries Service and the Pacific Fishery Management Council held a stock assessment review (STAR) panel meeting in May 2019 to evaluate and review benchmark assessments of Pacific coast Cabezon stocks.

Cabezon is a high value groundfish species for near-shore commercial and recreational fleets in California, Oregon, and Washington. This is the fourth full assessment of the population status off the west coast of the United States. The assessment includes two California sub-stocks and a sub-stock of Cabezon in the waters off Oregon (ORS). These stocks were last assessed in 2009. In addition, a data-poor assessment of cabezon in Washington was undertaken.

The technical review of pre-STAR assessments took place during a formal, public meeting of fishery stock assessment experts from 6th-9th May in Newport, Oregon. Two Center for Independent Experts (CIE) reviewers (Dr. Robin Cook and Dr. Noel Cadigan) were included in the review panel. The Terms of Reference (ToRs) of the peer review are attached in **Annex 2.**

Description of the Individual Reviewer's Role in the Review Activities

Materials for the review were made available on the 22nd April. These were studied prior to the meeting in preparation for the review. During the meeting, the reviewer took an active role in discussions. Requests for additional analyses for the STAT were noted and responses collated into a summary for the STAR panel report. The draft STAR panel report was agreed on the last day of the meeting. Comments on the final draft report were sent to the Panel chair on the 20th May.

Summary of Findings for each ToR

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.

The draft stock assessment documents were reviewed. These covered assessments of Cabezon in Washington, California and Oregon. In addition, material relating to M priors and previous STAR panel reviews were studied.

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

With the exception of Washington, all the assessments use data quantifying total catches by fleet, indices of abundance, length compositions and age compositions.

Catch data

These data are perhaps the most important input to the assessment, as they provide information on fishing mortality and help scale the assessment to the real fishable biomass. Each assessment attempts to characterize removals dating back to the beginning of the fishery, which is at least the early 20th century for California and the 1970s for Oregon. It is generally considered that "good" catch data are available from the 1980s onwards, and that the early data are subject to much uncertainty. This is in part the result of the way species were recorded historically and due to the problems in quantifying removals from the recreational fleet. A considerable amount of effort has gone into the reconstruction of the catch time series, but ultimately it is reliant on pragmatic assumptions about the development of the fishery and associations between a number of fish species in official records. The recreational catch data are estimated by surveys of participating vessels and the MRFSS program (which forms one of these) has been criticized for its design. These uncertainties mean that the catch data are subject to error and probably bias, a problem that perhaps deserves greater consideration in the assessments. These, for example, assume the catch is fixed (and by implication error free), which means errors and bias in the data are forced into the estimates of stock biomass and exploitation rate. While it is not possible to recover accurate data from historical records, a study that attempted to quantify likely uncertainty and bias would make a valuable contribution to understanding the veracity of the estimated stock trends from the assessments and provide a basis for well-designed sensitivity tests.

During the STAR panel meeting, revised catch estimates for the recreational fishery in California were made available and used in revised reference models. These re-allocated fish from the larger northern stock to the smaller southern stock, but did not make a large change to the estimated stock trends, though they did increase the scale of the SCS biomass.

Indices of abundance

There is one fishery independent survey data for the Oregon stock based on an observer program, which makes an important contribution to the assessment. For both the Oregon and California stocks, fishery dependent indices of abundance were also available. These have been derived from various components of the recreational fleets from the MRFSS survey and through official records. The main challenge in constructing such indices is in the removal of bias resulting from the way vessels target fish and to account for effort that is relevant to the species of interest in the assessments. All three assessments adopted a similar approach by identifying trips that might be expected to encounter the species concerned. An important element of this was to find the fish species in trips that were associated with the assessed species and then apply a filter that selected these trips. The STATs used a method proposed by Stephens and MacCall (2004), which uses ROC analysis to select a threshold for filtering the data based on the probability that the species of interest will occur in the catch given the species composition of the trip. Importantly, this process should help quantify the occurrence of zero catches when a trip was in an area likely to catch the target species. The AUC for the classifiers are less than 0.8, which means that the classification error rate is likely to be high. After filtering, a linear modelling approach was used to derive a CPUE index using a delta-lognormal model to extract an abundance signal. These models considered effects such as year, area and season with the “best” model being chosen on the basis of the AIC. CVs on the index derived from the best model were then calculated separately in a Bayesian modelling package, “rstan” in order to obtain more realistic CV values.

Overall, the approach to estimating CPUE abundance indices was thorough, but, given the very low occurrence of cabezon in samples, there is a very high proportion of zeros in some of the abundance index series even after filtering. Further thought should be given to the appropriateness of using delta-lognormal models in this situation. Disappointingly, the resulting indices did not appear to contain much population signal and tended not to be fitted very well. They are at best indicators of longer term trends.

For the Oregon onboard observer index, the index was revised during the meeting to remove interaction terms in the standardization. This made little change to the index, but was adopted for use in a revised reference model.

Length compositions

While some data from early years are available, the bulk of the length composition data began in the 1980s. Annual sample sizes at fleet level are generally small. Length compositions provide one of the few sources of data that can inform the model about year class strength and will be influential in estimating recruitment deviations in the model. Given that most data are for the post 1980s period, it means that there is very little information on age structure for the early period of the assessment. With the uncertainty in the early catch data, the interpretation of the estimated stock trends pre-1980 requires considerable caution.

Age compositions

A limited amount of age data is available for the NCS and Oregon assessments. For California data are available from research sampling in a few recent years. For Oregon, samples mostly from 2005 onwards are available. Since the assessment model is age structured, age data is important in providing the model with information on recruitment deviations. Age structured data is most effective when a year class is sampled regularly throughout its life time, so that an accurate picture of its survival rate can be estimated. This tends to be lacking in the data for these assessments.

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

Model framework

The Oregon and California assessments make use of the latest version of Stock Synthesis (SS3). This is a flexible modelling framework that can make use of a variety of disparate data and is particularly useful when time series data are discontinuous or where there are intermittent observations on length or age. It is therefore an appropriate choice for the assessments considered at the meeting.

Maximum likelihood forms the basis for parameter estimation but can be modified through the use of penalty functions that SS3 refers to as priors. The model is therefore founded in maximum likelihood but leans toward a Bayesian approach by incorporating prior information. However, as currently implemented, parameter estimates are characterized by point estimates with approximate asymptotic variances rather than their posterior distributions. MCMC sampling, however, was used for the reference models to better estimate posterior parameter distributions. These suggested that the posteriors were typically asymmetric. In one example from NCS, the SB0 distribution appeared slightly bimodal. Clearly, if distributions are not unimodal, then the interpretation of the model fit is problematic. In addition, where priors are applied, especially on parameters such as natural mortality or steepness that are not well informed by data, comparing the posterior distribution to the prior is a useful tool in understanding information in the data.

For the Washington assessment, a novel approach had been developed for data poor stocks. This builds on the DBSRA approach used earlier but can be implemented within the SS framework. In particular, the productivity parameters in DBSRA can be replaced by conventional steepness used in SS making the assessment more consistent with the other cabezon assessments. It allowed the same steepness assumption (0.7) to be applied. The use of the length composition tool was used to estimate selectivity parameters from recent observed length compositions and current biomass depletion. The panel felt this was an innovative and appropriate modelling approach to compute OFLs and investigate uncertainty.

Size composition model

The underlying population model is fully age-structured, but it also models the size composition of the population. This is done by assuming growth follows a von Bertalanffy curve with dispersion around the mean. The size composition of the population is then reconstructed from the age composition using the length at age distribution. In the assessments considered here, observed length distributions were assumed to represent mid-year distributions with invariant growth rates. This inevitably raises the question as to whether this somewhat rough growth assumption is sufficiently robust in the light of real changes in growth by cohort, month and year. Fits to the length compositions were often poor and may reflect over-simplified modelling assumptions or poorly sampled length distributions. As length compositions are likely to be influential in the estimation of recruitment deviations (especially where age data are few or absent), this is an issue that merits further investigation. However, the fits to the aggregate length compositions (i.e., summed overall years) were generally good.

Model parsimony

Each stock assessed reported the parameters that were estimated. These were generally in the region of 100 though they omitted survey catchability, q, which while estimated in closed form, nevertheless are model parameters and need to be considered especially when estimating M. The number of parameters is large when considering the available data, and there are clearly correlations between them. One would expect, for example correlations between R0, q and M that may be very high, and would indicate redundancy. Effort to try to find the most parsimonious model might help in reducing complexity and in identifying a model that had better predictive properties. For SCS, for example, given the absence of recent data, a simple stock synthesis model as was applied to the Washington stock might prove more valuable than attempting to fit a full SS3 flour fleet model.

Selectivity

An important element of the SS3 approach is the need to model selectivity. The selectivity curves filter the length composition of the underlying population to explain the observed fleet specific length compositions. Selectivity is likely to evolve over time as management measures are changed. The approach adopted for these assessments was to use time blocks for fleets where such regulations are thought to have affected selectivity. Clearly, it is desirable to model changes in selectivity to avoid miss-specification, but this comes at a cost of increasing the number of parameters to be estimated. In addition, the choice of blocks may not capture the response of fleets to a variety of different factors that may go beyond management measures alone. An alternative is to allow the selection pattern to evolve over time using an auto-correlated random effect to smooth the data rather than force a rigid parametric form. In the Oregon model, the abrupt change from domed to asymptotic selectivity in the commercial live fleet implies a sudden change in the survival of older fish that at face value looks unrealistic.

The shape of the selection curve at older ages or larger lengths can be significant in determining the scale of the estimated biomass. Inevitably, these ages and lengths tend to be poorly sampled, because they are less abundant. As a result, in fitting a selection curve, occasional observations at the largest size/oldest age may have undue influence. Choosing accumulator length bins or plus groups requires some thought. In the Oregon and California assessments, the accumulator bin had been reduced from 35 years to 20 years compared to the 2009 assessment, and this would seem a sensible change

Natural Mortality, M

Natural mortality is included in the models either as a predetermined quantity, or estimated but informed by a prior based on a range of empirical estimation methods (e.g. Hamel, 2015). This, in effect, provides an estimate of the average annual non-fishing mortality experienced by an individual over its lifetime. M is generally size dependent, while the models applied in these assessments assume it is fixed (except for gender differences). Using a size dependent relationship would imply, for example, that M for males is higher than females. Hence estimating M as a size/age invariant quantity will result in bias in the estimate of other model parameters and could be significant in the estimation of recruitment deviations. A possible way to address this issue is to use a relationship to characterize M by size (e.g. Lorenzen 1996) and then scale the relationship to an overall mean informed by the M prior. During discussion at the meeting, the STAT team suggested that M on older fish was high, since, although cabezon are large, they do not appear to live longer than about 17 years. This would imply that age specific mortality is U-shaped and there may be some merit in resorting to life history theory to derive a survival vector that could be scaled to an estimated mean value.

The ability to estimate M within the model will depend on contrast in the data and constraints or assumptions on other parameters. It is usually difficult to estimate within the model, because it is confounded with other parameters such as catchability. In view of the necessary simplifying assumptions on M, it is worth reflecting on whether trying to estimate its value is very useful. Inspection of the likelihood profiles over M for these stocks did not suggest that the various data sets provided consistent information. In the case of NCS for example, the length data imply lower M than the prior, while for Oregon recreational age data suggested a higher value. The belief of the STAT team was that these low or high values were unrealistic, which means that there is a problem with the data or the model specification. There is very little contrast in the Oregon stock trajectory, which would imply the estimation of M within the model is likely to be very unreliable.

A very useful sensitivity analysis to M was conducted for all three stocks, and this perhaps is the most informative insight into the interpretation of stock status.

It is worth remembering that for these stocks the estimated fishing mortality rate is much lower than M; so much of the stock dynamics will be driven by factors external to the fishery. Whatever the true level of M, it is likely to vary over time, and since M cannot be included in the model dynamically (as there are no data to support it), the interpretation of stock trends is extremely difficult.

Weighting multinomial data

Length and conditional age compositions are modelled as multinomial distributions where sample size is a critical weighting factor in the likelihood. The problem of identifying the correct effective sample size is well known. It will be most pronounced when the actual number of samples is small, because the variability in the observations will be greatest. In all three assessments, sensitivity to the choice of weighting was investigated using Francis and harmonic mean weighting and shown to be a significant source of uncertainty, i.e. the sensitivity run produced estimates of the quantities of interest outside the range of the reference model. Perhaps the lesson here is the need to increase the number of samples both to provide the assessment with better data and to reduce the sensitivity to choice of weights.

Beverton-Holt steepness

All the assessment models use the Beverton-Holt stock-recruitment function parameterized in terms of steepness, h, and virgin biomass. For all these stocks steepness was fixed at 0.7 and probably represents a sensible choice. Plots of estimated stock recruit data indicate there are no data points in the left hand portion of the plot that would enable steepness to be estimated.

In the case of the Washington stock, earlier assessments using the DBSRA approach used productivity parameters that translate into steepness values about half the values used for California and Oregon. For consistency, however, this low value was not used in the simple stock synthesis model.

Sensitivity testing

Systematic sensitivity analysis that considers the principal sources of uncertainty is presented for the three full assessments. The analysis considers the influence of data components (indices, length compositions and conditional age) and model specification (M, growth, data weighting and recruitment assumptions) in the principal stock metrics. The results of these sensitivities are plotted to show where the estimates lie in the range of uncertainty as derived from the reference model. This provides a very clear indication of where the main issues lie. For Oregon, these analyses tend to show changes of scale rather than changes of overall trend or shape. For the NCS model, stock depletion in 2019 is sensitive to the length compositions from the commercial live fleet. Similarly, the SCS depletion is sensitive to the recreational boat length frequencies.

As noted earlier, the catch data in all assessments are assumed to be exact or estimated with high precision. This is likely to be a necessary assumption for model convergence, though it is clearly unrealistic. A sensitivity test that examined this assumption would be highly desirable. In some Pacific coast assessments, sensitivity to doubling or halving the historical catch has been investigated. However, this simply changes the scale, and the more important question is whether annual changes in the catch are well estimated, since these may alter the perceived stock trend. Random draws of possible catch streams from likely ranges of uncertainty would be a more demanding sensitivity test. Preserving the annual autocorrelation would be necessary and could be achieved by adding a random error to the nominal values.

Retrospective analysis

Retrospective runs did not reveal any major problems as data are sequentially removed from the assessments. However, the analyses illustrate the dependence of the assessment on the catch data that is assumed to be known without error.

Bridge models

The assessments used the latest version of SS3, which is an update on the version used in the last assessments in 2009. The STAT had configured bridge models that tested the 2009 models in the current version of SS3. The NCS model reproduced the 2009 results, but the Oregon model gave substantially different results. When the parameters were fixed at the 2009 values the stock trends were accurately reproduced, but when estimated within the model, a lower log likelihood (two units) was found with very different parameter values. This was investigated, but no explanation has yet been found. It suggests that either the objective function is differently specified in the software or that the minimization algorithm has changed. In any event, it is illustrative of the flat likelihood surface the models tend to exhibit and the sensitivity to small changes in the minimum found. The panel discussed this problem and accepted that given the 10 years of new data since the last assessment, the new

Oregon model was the best available. However, this problem is a concern and to some degree undermines confidence in the results.

Jitter analyses

Jitter analyses suggest that the models converged on the lowest negative log-likelihood. In the case of Oregon, due to the problem with the bridge model, jittered results that showed two or less likelihood units of difference were investigated at the meeting to see if these produced major changes in the parameter estimates. These investigations confirmed that similar parameter values were recovered.

Model estimates of CVs

SS3 estimates CVs on recruitment and there is a process of correcting for bias in this estimation. For the earliest time period, no recruitment deviations are estimated, yet the estimated CV on these recruitment values is lower than the period when recruitment is allowed to vary and when there is information in the data about year class strength. This gives a false impression of precision and knowledge about recruitment in the early period. Presumably the CV on these early values represents the precision of mean recruitment rather than distribution of recruitment itself. The problem extends to the CVs on estimated biomass, which appear to show that the earliest SSB estimates are as precisely known as more recent values that are informed by much more data. The early SSB estimates will be the least precise, not only because there is almost no data to inform the model, but also because the catch data during this period are the most uncertain. The CVs shown on the recruitment and SSB time series plots therefore need to be interpreted with considerable caution.

Stock status

The limited amount of age data and lack of informative abundance indices means that it is difficult to have high confidence in the estimated stock status. The uncertainty in the California stocks spans the reference points making status determination ambiguous. For Oregon, the confidence intervals on depletion lie above the reference points, which would indicate the stock is not overfished but the intervals probably do not fully capture uncertainty.

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

A number of minor improvements were made to the reference models based on discussions at the review meeting. This included revised catch data for the California assessments and removal of the recreational boat samples from the SCS assessment. The prior on k in the von Bertalanffy growth parameter was removed in the NCS assessment. For the Oregon assessment, a revised ORBS index was used and the MFRSS index was downweighted to reduce the influence of uncertain data.

Natural mortality was identified as a major source of uncertainty to be addressed in decision tables.

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

The principal limitation in these assessments is the available data. Catch data pre-1980 are regarded as uncertain and there is a shortage of age data. The absence of a good quality fishery independent survey is also a major weakness for California and Washington. With these limitations in mind, the analyses are of a very high standard making use of state-of-the-art analytical methods. I would judge the science to be the best available.

Stock Synthesis is now a well-established modelling framework and is well suited to the type and quantity of data available for assessment. It is, however, very complex both in the form of the objective function and the multiplicity of configuration options, which can obscure what it actually is doing. By their nature stock assessment models are over-parameterized and SS3 is no exception. With relatively uninformative data as in these assessments, the model is not well anchored and a wide variety of possible interpretations of the data are possible. Thus, while the science is of a high standard, the results of the assessments are not robust. At this point in time much more could be gained by collecting more informative data.

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.

Stock identity

The current stock structure was not updated from the 2009 assessment as no new data were available. Discussion at the meeting suggested that the very wide spatial scale of the stocks masked local effects and that more fine grained models may perform better. The value of such an approach has to be balanced against the amount of data available. Further research into stock boundaries may be worthwhile before a more elaborate assessment is attempted.

Data

At present there is a large investment of analytic effort into somewhat limited data. More resources devoted to data collection would be highly beneficial. Priority should be given to the collection and processing of more age samples from all four stocks. This needs to be maintained to create a long time series of observations. While much has been done to derive abundance indices from fishery data, these do not appear to be very informative and a dedicated survey would help overcome this problem. If possible, a fishery independent survey should be developed for California to calibrate estimates of biomass in the most recent year.

Modelling approach

The use of SS3 allows highly complex and parameter rich models to be developed and the assessment models used in the assessments reviewed fall into this category. In general, while exploring complex models is undoubtedly useful, there should a systematic attempt to reduce complexity by critically examining the precision and posterior distributions of the parameters as well as their correlations. This would help in identifying redundancy and may help in improving model stability and predicative power. Time blocking of selectivity curves may help reduce residuals, but the question of whether the parameters of the curves were significantly different needs to be explored. A non-parametric time series approach may be a better way to capture time varying selection without over-parameterizing the model.

The assessment models chosen are likely to be the best way of incorporating a variety of different data into a comprehensive analysis. However, there may be some merit in applying other simpler approaches to gain insights into the information content of the data and identify conflicting signals. The SS3 model applied to Washington, for example, might prove a useful comparator if applied to the Oregon and California models if only to see how projected OFLs differ.

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

The review was conducted in a constructive manner and the STAT teams were responsive to the requests from the Panel for additional analyses with all the essential runs being completed during the meeting.

Many of the issues discussed have been referred to in earlier sections of this report. These included:

- Identifying selectivity assumptions that better explained the observed data
- Appropriate assumptions on natural mortality, particularly the best approach to applying a prior for M

Towards the end of the meeting there were discussions on the states of nature for decision tables.

Overall, there was effective engagement from all members of the Panel, the STATs and the Panel advisors. This led to improvements in the configuration of the base models.

Recommendations for future assessments are discussed in the next section.

Conclusions and Recommendations

The assessments of cabezon represent the best science available given the existing data. The analyses were thorough and considerable work had gone into making good use of data from a variety of sources. The limited amount of age data and lack of informative fishery independent abundance indices means that despite the elegance of the assessments, it is difficult to have high confidence in the estimated stock trends.

Should managers attach importance to these stocks, then I would **recommend that priority be given to the collection and processing of more age samples from all four stocks. This needs to be maintained to create a coherent time series of observations. If possible, fishery independent surveys should be developed to calibrate estimates of biomass in the most recent year.**

In common with many other assessments in this region, early catch estimates are subject to considerable uncertainty. The assumption that catches are exact and treated as fixed in the model is probably necessary but clearly unrealistic. Sensitivity to this problem needs to be adequately investigated as the catch data are influential in the assessment. I **recommend that a demanding sensitivity analysis is performed where plausible alternative catch streams are generated stochastically**. The practice of halving or doubling the catch as a sensitivity test is not very demanding and is unlikely to probe the nature of the uncertainty in the data.

Natural mortality is the largest component of total mortality in these stocks and will drive much of the stock dynamics. I was not entirely convinced that modelling M as a constant value by gender was the best approach, or that trying to estimate its value within the model was reliable. I **recommend that the way M is modelled and estimated is reviewed. Consideration should be given to modelling M by size and scaling it to a mean value given by an empirical method.** This would avoid the need to model M by gender and would capture some of its annual variation.

I recognize that SS3 is a powerful, useful and appropriate tool for the assessment of these stocks. However, thought needs to be given to the appropriate level of model complexity to ensure that the final base model fitted to the data also has the appropriate forecasting properties. I would **recommend that a procedure is developed to identify the most parsimonious model using an information statistic and the parameter correlation matrix.**

SS3 provides an impressive range of diagnostics to aid model development. In its present implementation, it does not appear to provide realistic posterior distributions of the estimated parameters. This is something of a limitation, as it hinders identifying problematic model fits and understanding the relative contribution of priors and data to the estimates. I **recommend that SS3 be updated to provide full parameter posterior distributions.**

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

Draft assessment documents and supporting material were made available on the Pacific Fisheries Management Council FTP site two weeks in advance of the meeting. This is realistically the minimum advance time to review the assessments adequately. The principal documents are voluminous and take time to digest. As always, more time would be appreciated and would lead to more considered interventions at the review meeting. Understandably, there is a compromise to be struck between the completion of assessment documents and time available for review. Perhaps the two-week period is the best that can be achieved.

The meeting itself was constructive and productive with effective and excellent co-operation from the STAT teams. Meeting facilities were good, and the local staff provided great support to the reviewers.

References

Hamel, O. 2015. A method for calculating a meta-analytical prior for the natural mortality rate using multiple life history correlates. ICES Journal of Marine Science 72: 62-69.

Lorenzen, K. (1996). The relationship between body weight and natural mortality in juvenile and adult fish: a comparison of natural ecosystems and aquaculture. Journal of Fish Biology, 49, 627–647.

Stephens, A., and MacCall, A. 2004. A multispecies approach to subsetting logbook data for purposes of estimating CPUE. Fisheries Research 70: 299-310.

Annex 1: Bibliography of materials provided for review

The following materials were made available in the PFMC ftp site before and during the meeting. They can be found at: ftp://ftp.pcouncil.org/pub/!2019%20GF%20STAR%20Panels/STAR%20Panel%201%20-%20Cabezon/

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. Title. Pacific Fishery Management Council, Portland, OR. Available from http://www.pcouncil.org/groundfish/stock-assessments/

Background

Appendix F. Reef Delineation and Drift Selection.

CABEZON STAR Panel Report, July 27-30, 2009.

Cope, J. M. and M. Key.2009. Status of Cabezon (Scorpaenichthys marmoratus) in California and Oregon Waters as Assessed in 2009.

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

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

Presentations

Cope, J.M. and Berger, A.M. Assessing Cabezon (Scorpaenichthys marmoratus) stocks in waters off of California and Oregon, with catch limit estimation for Washington State.

Annex 2: 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.

<u>Contract Deliverables - Independent CIE Peer Review Reports</u>: 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**.

<u>Other Tasks – Contribution to Summary Report</u>: 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-nationalregistration- 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: Tentative Agenda

Final Agenda to be provided two weeks prior to the meeting with draft assessments and background materials.

Stock Assessment Review (STAR) Panel 1

NMFS Northwest Fisheries Science Center

2032 SE OSU Drive

Newport, Oregon 97365

May 6-10, 2019

TBD

Annex 3: Panel membership and participation

STAR Panel Members

Rishi Sharma, National Marine Fisheries Service Northwest Fisheries Science Center, Chair
Noel Cadigan, Center for Independent Experts
Robin Cook, Center for Independent Experts
Will White, Oregon State University

Stock Assessment Team (STAT) Members

Jason Cope, National Marine Fisheries Service Northwest Fisheries Science Center
Aaron Berger, National Marine Fisheries Service Northwest Fisheries Science Center
John Budrick, California Department of Fish and Wildlife
Ali Whitman, Oregon Department of Fish and Wildlife
Katelyn Bosely, National Marine Fisheries Service Northwest Fisheries Science Center
Theresa Tsou, Washington Department of Fish and Wildlife
Lisa Hillier, Washington Department of Fish and Wildlife
Kristen Hinton, Washington Department of Fish and Wildlife
Megan Wilson, Oregon State University

STAR Panel Advisors

Mr. Patrick Mirick, Oregon Department of Fish and Wildlife, Groundfish Management Team
Mr. Gerry Richter, B&G Seafoods, Inc., Groundfish Advisory Subpanel
Mr. John DeVore, Pacific Fishery Management Council