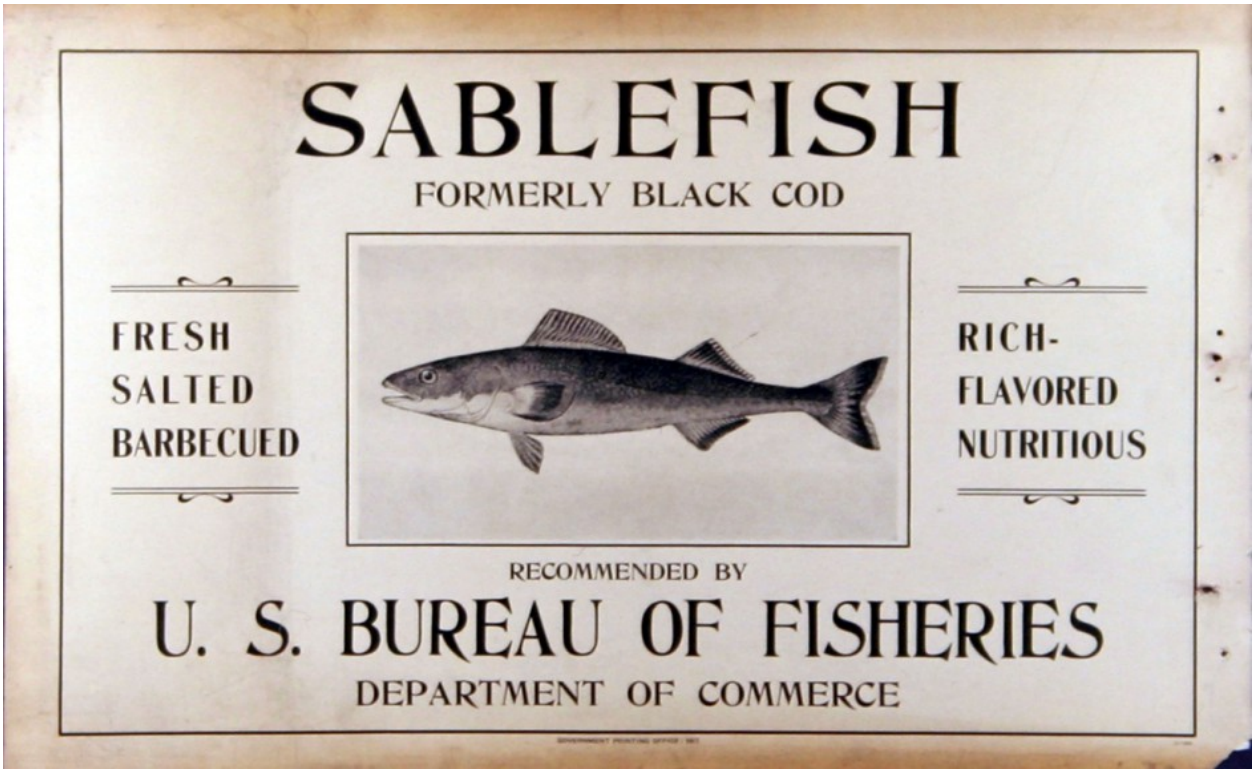
Center for Independent Experts (CIE) Independent Peer Review Report of the Sablefish STAR Panel Review 2019

By

Robin Cook



Prepared for
Center for Independent Experts
Independent System for Peer Review

Contents

Executive Summary

i. An assessment of sablefish in Oregon and California, and Washington was reviewed at the Northwest Fisheries Science Center, Seattle, WA, during a formal, public meeting of fishery stock assessment experts from July 8-12, 2019. Two CIE reviewers were included in the review panel.

ii. The assessment represented the best science available given the existing data, which is mostly limited to recent years, and the decision to use stock synthesis as the principal modelling tool. Stock status is close to limit reference points and the uncertainty associated with this is large.

iii. There are a number of fishery independent surveys available. The west coast shelf survey (WCGBT) makes an important contribution to the estimation of recent population trends and recruitment. However, the older surveys are not well fitted by the assessment. Priority should be given to maintaining the WCGBT survey as it appears to offer the most promising data source for future assessments. It tracks stock abundance well and provides information on recruitment.

iv. In the final base model agreed at the review meeting, all the length data were removed apart from the conditional age at length data for the WCGBT survey. This arose due to variations in growth that could not be accommodated in the assessment model and lead to conflicting signals in the length and age data. Leaving out data is not very satisfactory and the problem needs to be investigated by developing an appropriate way to handle the length data either through a more realistic growth model or converting length to age data outside the assessment.

v. While overall stock trends are robust, the scale of stock biomass is not well determined. It means that biomass reference points are highly uncertain though ratio reference points such a percent depletion will be more robust.

vi. The choice of weighting method in SS3 had an important effect on the parameter estimates. Francis weighting was adopted because it resulted in a base run with an improved retrospective pattern, but the sensitivity of the assessments to alternative weighting methods is indicative of uncertainty.

vii. The catch data in the assessment are treated as exact and fixed in the model. While this is probably a necessary assumption, it is clearly unrealistic. However, it appears that the fishery catch is a minor contributor to the stock dynamics and represents a lesser fraction of the total mortality.

viii. Natural mortality appears to be estimable in the models and accounts for a large share of the mortality acting on the population. As it is assumed constant by size and over time, the assessments are not able to capture true population 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.

ix. 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.

x. 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 would produce more realistic estimates of posterior distributions and should be a routine output of the analyses.

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

Background

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

The stock was assessed as a single unit in US waters comprising California, Oregon and Washington. However, Sablefish are found in the Bering sea off Alaska and westward to Japan. This is a further full assessment of the species which was last considered in 2011 and 2015.

The technical review of pre-STAR assessments took place during a formal, public, meeting of fishery stock assessment experts from 8th-12th July in Seattle, WA. Two CIE reviewers 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 25th June. An updated pre-STAR assessment report was received on the 4th July. 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 STAR panel report was prepared and agreed by correspondence after the meeting.

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 an assessment of sablefish in Washington, California and Oregon. In addition, material relating to 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.

The assessments used data quantifying total catches by fleet, indices of abundance, length compositions and age composition data. In addition, a sea-level index was used as an index of recruitment.

Catch data

These data are an important input to the assessment as they provide information on fishing mortality and normally help scale the assessment to the absolute fishable biomass. The assessment attempts to characterize removals dating back to the beginning of the fishery which is judged to be 1890. Catch data have improved over the period of the assessment but rely on reconstructions for the earliest years.

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, discard rates and discard survival rates. These uncertainties mean that the catch data are subject to error and possibly some bias. The assessments 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. For sablefish catch records are more reliable than for many minor species so the reconstructed catches are considered suitable for the assumption of low measurement error. As the catch represents a relatively small fraction of the estimated biomass (~5% in recent years), errors in the catch data should not be a major problem.

Indices of abundance

A number of fishery independent surveys were available for the assessments. These include the NWFSC shelf (WCGBT) and slope surveys, the AFSC slope survey and the triennial survey. An environmental index measuring sea level was used as a predictor of recruitment.

The slope surveys and the triennial survey cover a limited range of years ending in the early 2000s. The WCGBT began in the early 2000s and continues to the present. As a result, there is little overlap between the older surveys and the WCGBT meaning that only the WCGBT informs the critical recent period of the assessment. The earlier surveys do not appear to be fitted well by the assessment model and hence contribute little information on biomass trends but are important in contributing age data that inform recruitment deviations and growth.

Indices were standardized using the VAST package assuming gamma errors. Although the indices provide swept area estimates of biomass, catchability is estimated within the model effectively assuming the indices provide information on relative abundance only. Without an informative prior on catchability for at least one survey, this contributes to indeterminacy in the biomass scale of the assessment.

The assessment model was able to fit the WCGBT well and the estimates of the precision of the survey were not inflated during the model fit suggesting this survey is the most consistent with other data included in the assessment.

A considerable amount of work has been invested in establishing a relationship between year class strength and environmental co-variates. Various metrics have been investigated but recent research suggests that sea-level alone is an adequate correlate of recruitment acting as a proxy measure for environmental conditions. The sea level index is crucial to the estimation of recruitment deviations for the mid-20th century period when no length or age data are available. Environmental indices as predictors of year class strength are notoriously unreliable and while this index should not be rejected, it should be kept under review at every assessment to ensure its predictive properties are useful.

As implemented in the assessment, the first dynamic factor (DF1) derived from an analysis of a range of tide gauges along the coast was used as the index for sea-level. However, analysis of the sea level data suggests that the best model accounts for ~35% of the variance in recruitment and included DF3 both as a linear and squared term implying the relationship with recruitment is non-linear. The choice to use only DF1 which accounts for <25% of the recruitment variability therefore needs more careful thought as it appears to have been adopted primarily for simplicity.

In the initial base model, the DF1 index was treated as proportional to recruitment and no additional variance was estimated within SS3 as is commonly done for conventional survey indices. During the meeting this additional variance was estimated and led to improved fits to the WCGBT survey. This was due to high additional variance being added to the survey CVs effectively down-weighting it to give little influence. This somewhat contradicts the rationale for including it.

Length compositions

Available length composition data for surveys and the fishery began around the 1980s. Annual sample sizes at fleet level are generally low compared to the WCGBT which provides most of the data in recent years. The latter is likely to be the most important fishery independent data source for the assessment. Visual inspection of the length frequency data does show some evidence of year class strength signal and might facilitate estimation of recruitment deviations.

Given that most data are for the post 1980 period, it means that there is almost no 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.

During the meeting all length data were removed from the base case assessment because the model was unable to adequately model growth in a way that fitted the observations. In effect the length data appeared to give conflicting signals in estimates of M and R0 compared to the age data. Since age data are more information rich, the length data were removed to avoid this conflict.

Age compositions

Age data is available from about 1985 onwards for the fishery and the surveys. Most data are available from the WCGBT survey. Age reading studies suggest that age determination is difficult especially at the older ages. It seems likely that age determination for the older fish is biased low and could be in error by more than 10 years. Aging error is taken into account in the assessment model but this does not allow for bias. Model runs assuming no aging error on young fish gave poorer fits to the data indicating it may be an important source of uncertainty.

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. It also is a major source of information on recruitment strength. The sablefish assessment appears to have sufficient data of this type to facilitate the estimation of recruitment deviations, at least since the 1980s. It does seem possible, however, that problems with age determination for the oldest fish may add to uncertainty in recruitment estimates and may have implications for how the plus group is chosen. In this assessment the plus group age was increased compared to the last assessment and is likely to affect the scale of the assessment through estimates of the selectivity parameters.

In the final base model agreed with the STAT, age data for the WCGBT survey were included as conditional age at length as opposed to the initial base model that used unconditional age compositions.

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

Model framework

The assessment makes 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 assessment.

Maximum likelihood forms the basis for parameter estimation but can be modified through the use of penalty functions referred 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 usually characterized by point estimates with approximate asymptotic

variances rather than their full posterior distributions. MCMC sampling of posterior distributions in other SS3 assessments suggests that the posteriors may be asymmetric and the use of asymptotic variances may not therefore adequately characterize parameter uncertainty. The STAT indicated that full MCMC runs were very time consuming (48hrs) and could not realistically be undertaken at the meeting. Nevertheless, these runs should be performed for the final base model and reported in the final assessment document.

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 specified model 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 follow invariant growth rates. This inevitably raises the question as to whether this somewhat approximate growth assumption is sufficiently robust in the light of possible changes in growth by cohort, month and year. This issue may be of most importance for hindcasting the population back to 1890, since there is little data pre-1980 to estimate growth and the projected trajectory is predicated on constancy of growth over a period of nearly 90 years.

A conventional von Bertalanffy curve is applied with each sex following a separate growth curve with females growing somewhat larger than males.

As described above, length data were not included in the model due to conflicts with signals in the age data. Inspection of the size at age in the raw data strongly indicated a cohort effect that undermines the constancy of growth assumption. As there was no obvious way to account for the cohort effect the panel felt that omitting the length data was a preferred option.

Model parsimony

The number of parameters is large when considering the available data. Some thought should be given as to whether such highly parameterized models are really justified by the information in the data. Building a model from the simplest possible might provide insights into just how much complexity is supported by the data and may simplify the number of assumptions that need to be made. There was some discussion about the estimation of M by sex since when estimated separately, there was very little difference in the values estimated. Furthermore, the various model runs investigated during the meeting showed that there was no consistency between the relative value of M between sexes suggesting that the data do not contain sufficient information to estimate sex specific values.

As noted above there were also problems with the growth model in SS3. A way around this would be to use observed weights at age to avoid estimating growth parameters within the model.

During the meeting the model was also simplified by collapsing the hook and line and pot fleets into a single fleet reducing the number of selectivity parameters to be estimated. In addition, the sex specific selectivities were constrained to the same shape using an offset rather than trying to estimate a full set of parameters for each curve. These constraints appeared to be critical to their estimation given the assumed dome shaped selection for all fleets.

Selectivity

An important element of the SS3 approach is the need to model selectivity. The selectivity curves filter the length or age composition of the underlying population to explain the observed fleet specific compositions. For sablefish dome shaped selectivity for age was assumed in the model for most fleet components but, initially asymptotic selection assumed (and estimated) for the WCGBT survey. It is desirable to fix at least one section function to be asymptotic as was done for the WCGBT survey. However, during the meeting the base model was revised to allow the WCGBT survey to have dome shaped selectivity. This improved the model fit but in order to estimate the fixed gear selectivity, constraints on the selectivity parameters for this fleet were required as discussed above.

Selectivity is critical to the estimated scale of the stock biomass and the lack of information in the data to estimate all the selection parameters means that the biomass scale is heavily influenced by modeling choices.

In addition to the form of the selection curve, temporal changes in selectivity were accommodated through the use of time blocking. This adds to realism by accounting for changes in fleet behavior, especially in relation to management regulations, but comes at the expense of increased model complexity. Given the lack of information in the data, these additional parameters are likely to be highly uncertain.

Natural Mortality, M

Natural mortality is estimated in the assessments but is 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. The model appears to be able to estimate M but various model runs investigated during the meeting gave values ranging between 0.05 and 0.09, which suggests it is heavily influenced by the prior (with a mode of 0.05).

Sex specific values of M were estimated in the final base run with a slightly higher value for females. However, the 95% CIs for the male and female M values overlap substantially and there is no evidence from the assessment that M varies by sex.

As the estimated fishing mortality rate is low and is of a similar magnitude to M, much of the stock dynamics will be driven by natural factors external to the fishery. Whatever the true level of M, it is likely to vary over time, and since M cannot be reliably included in the model dynamically (as there are not data to support it) the interpretation of hindcast stock trends is extremely difficult.

Weighting multinomial data

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. Sensitivity to the choice of weighting was investigated using Francis, Dirichelet-multinomial (D-M) and harmonic mean weighting. D-M weighting was initially adopted in the base model, but when Francis weighting was used, a better fit to the WCGBT was evident. Harmonic mean weighting gave similar results to the Francis method. The Francis weighting also resulted in an improved retrospective pattern and also estimated recruitment values that appeared more realistic, especially for the period in the 1960s when some very large year classes appear to have occurred. The Francis run supported these year classes being large but did not give extreme values seen with D-M

weighting. As a result of the improved model diagnostics, the Francis method was adopted for the new base model.

Recruitment model

The assessment uses the Beverton-Holt stock-recruitment function parameterized in terms of steepness, h, and recruitment at unexploited biomass, R0. For this stock, steepness was fixed at 0.7 to reflect the likely productivity of sablefish based on meta-analyses. 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 and there is little evidence of a change in mean recruitment over the range of spawning biomass observed.

For earlier data, the estimated deviations are largely driven by the sea level index. While there are analyses to support the relationship between the index and recruitment, without a clear and specific causal mechanism such relationships need to be treated with considerable caution. In the final base model, the low weight given to the sea level index meant that recruitment deviations showed little variation.

Recruitment deviations are modelled as random effects assuming a lognormal distribution. Inspection of the estimated recruitment values suggests that the pattern is one of typically low year classes with periodic large ones that sustain the stock. One might question whether the lognormal distribution adequately captures this pattern of recruitment. One possible line of investigation would be to include the environmental covariate (sea level) as part of the stock-recruitment function to remove the environmental signal before treating the residuals as lognormal. For example, steepness could be modeled as a function of seal level, though more thought into the biological processes might help inform how this might be modelled.

Uncertainty

Systematic sensitivity analysis which considers the principal sources of uncertainty was presented for the initial base model assessment. The analyses consider the influence of a range of modelling assumptions on the principal stock metrics. However, the final base model was agreed late on the last day of the meeting and it was not possible to look in detail at a full set of sensitivity runs.

The results of the initial sensitivity runs provide a clear indication of where the main issues occur. They show that the greatest uncertainty is in the estimate of the scale of the biomass. The scale is affected by the selectivity assumptions and since there is insufficient information in the data to estimate all these parameters, there is a high degree of uncertainty. Reference points based on absolute biomass will therefore be highly uncertain though ratio values such as Bratio_2019 are more robust. Estimates of the latter from a range of different model configurations resulted in depletion estimates that bracketed the limit reference point and illustrate the uncertainty in stock status.

Recruitment deviations are determined largely by the age data and, for the historical period, the sea level index. Provided the age determinations are reasonably accurate, recruitment deviation should be well estimated for recent years. Earlier recruitment values are more speculative as they depend on an environmental index.

As noted earlier, the catch data in all assessments are assumed exact or estimated with high precision. This is likely to be a necessary assumption for model convergence though it is clearly unrealistic. As the

catch represents a small part of the total mortality, the assessment is less affected by catch estimation errors.

Retrospective runs showed that the final base model agreed at the meeting had a tendency to underestimate the biomass. This tendency was much reduced compared to earlier base models considered at the meeting and suggests that at least some model mis-specification has been mitigated in the new base model.

Jitter analyses suggest that the base model described in the pre-STAR assessment report converged on the lowest negative log-likelihood. A similar analysis for the final base model was not available and will need to be carried out.

Stock status

Precise stock status is subject to high uncertainty since the estimate of depletion is very close to (and possibly below) the overfished reference point. While it is possible the stock is not experiencing over-fishing, the 95% CI associated with the SPR ratio includes the possibility of overfishing. The final base run suggests that for the last decade at least, the exploitation rate has varied around the overfishing threshold without little long term trend.

Estimated stock trends suggest that the stock has increased recently following a long period of decline from the mid-1970s to 2011. The increase appears to be due to a number of stronger year classes. The trends appear robust to a range of sensitivity tests, at least for the more recent decades covered by the WCGBT survey and where age data are available. Earlier trends are more problematic to interpret given the dependence of the estimated SSB on recruitment informed only by the sea level index.

The uncertainty in the scale of the biomass means that the calculation of OFLs will be heavily conditioned on the assumptions made in the base model.

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

The two main areas of concern are the exclusion of the length data and the choice of data weighting method. Excluding data is not desirable and highlights a modelling (growth) problem rather than a data problem. It may be possible to circumvent the growth modelling issue by using observed weights at age. These could be used directly or smoothed using a statistical model, external to the assessment model, to account for cohort effects.

The choice of Francis or D-M weighting made a difference to the model fit to the WCGBT survey which appears to be the most important data source. It was resolved during the meeting by adopting Francis weighting which gave more weight to the survey. This improved the retrospective pattern and seemed to give more realistic recruitment estimates. The weighting issue is an enduring problem with SS3 and arises partly due to the assumption of multinomial data for the age and length compositions. It would be worth reviewing whether this assumption really is justified and whether other ways of entering the age composition data (such as raised age compositions where length data are converted to age) might offer a less sensitive formulation. It could also avoid the common problem of conflict between the age and length data, often associated with the difficulty of trying to model growth. Digital data (i.e., age) are simpler to model that analogue data (length or growth).

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

The principal limitation in these assessments is the available survey data. While there are a number of fishery independent surveys, most do not, unfortunately, appear to assist in estimating stock trends though the WCGBT is an exception and is a critically important source of information. This survey began in 2003 and therefore only informs the recent period of the assessment. During the meeting, a sensitivity run was done where the start year was 1970. This run resulted in an estimate of B0 that was very similar to values obtained from the full time series assessment and strongly suggests the parameters used in the SPR calculation are largely determined by the data from this survey and that the early data are of minor importance in assessing current stock status. 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 on scale as in these assessments, the model is not well anchored and a wide variety of possible interpretations of absolute biomass are possible.

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.

Data

Priority should be given to the WCGBT survey which is the most important source of data to determine stock trends and estimate recruitment deviations. Given the problems of biomass scale, there may be value in trying to develop and informative prior on survey catchability to condition the model.

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 be 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.

For much of the period considered in the assessment natural mortality is a dominant source of mortality but is assumed constant through time and by size. There is of course little alternative to making this assumption since there is insufficient information up to the start of the WCGBT survey to inform the model about M. It makes the interpretation of the early stock trajectory problematic. In view of this, there is value in trying to make more effective use of the WCGBT survey data on its own to avoid older less reliable data contaminating critical parameter estimates. There does not seem to be any need to hindcast the model back to 1890 since the WCGBT data provide an abundance estimate from 2003 and it would be quite feasible to run the assessment forward from the late 1990s when catches are best known. Not only are the early catch data less reliable, but the assumptions of invariance in the hindcast are very unlikely to be correct.

Estimating temporal trends in M would be difficult but a more realistic treatment of mortality by size should be possible using, for example, the Lorenzen meta-analysis (Lorenzen 1996). M could then be estimated within the assessment model as a scaling constant on a chosen size dependent function.

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 helpful and extremely 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:

- The exclusion of length data due to their incompatibility with the constant growth assumption
- The use of dome shaped selectivity for all fleets and surveys
- The use of Francis weighting in preference DM weighting
- Relevant weighting given to the sea-level index.

Towards the end of the meeting there were discussions on the states of nature for decision tables. It was proposed by the STAT to use the range of uncertainty in estimated M as the axis of uncertainty. This may have merit given the range of M values estimated in the various runs considered during the meeting. However, there was little time at the meeting to consider this topic and further consideration of the appropriate states of nature is needed. In particular, the uncertainty in the biomass scale should be given careful thought.

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

Recommendations for future assessments are discussed in the next section.

Conclusions and Recommendations

The assessment of sablefish represents the best science available given the existing data and the SS3 software available to the STATs. The analyses were thorough and considerable work has gone into making good use of data from a variety of sources. The precise status of the stock is unclear given the proximity of current biomass and productivity to reference points, but it is likely to be borderline overfished and possibly not experiencing overfishing.

Given the value of the WCGBT survey, I would **recommend that priority be given to continuing this survey. It is of particular importance to maintain the surveys to calibrate estimates of biomass.**

Natural mortality is the largest component of total mortality in the stock and will drive much of the stock dynamics. I was not entirely convinced that modelling M as a constant value was the best approach. 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 might avoid the need to model M by gender and would capture some of its annual variation.

More focus should be placed on analysis of the WCGBT survey to better understand the population dynamics of the stock. Restricting an analysis to the period from the mid-1990s is feasible if an

informative prior on q was developed and might avoid contamination by strong assumptions and poor data applied during the early period of the assessment.

In view of the high frequency of small year classes and periodic occurrence of large year classes **there may be merit in reviewing the lognormal distribution used to characterize recruitment deviations. Since there is apparently a relationship between an environmental index and recruitment, considerations be should be given to include this in the stock-recruitment function rather than a recruitment index.**

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 be 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 unless an MCMC simulation is performed. 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 MCMC runs are performed on the final base run to provide full parameter posterior distributions and these should be used to characterize states of nature in projections.**

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. However, the pre-STAR assessment report was updated very shortly before the panel meeting and the assessment had undergone further change by the first day of the panel discussions. This meant that much of earlier document review was redundant and that the level of scrutiny given to the final base model was less than ideal. Understandably, there is a compromise to be struck between the completion of assessment documents and time available for review.

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.

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%203%20-%20Sablefish/

Pre-STAR assessment reports

Haltuch, M.A., Johnson, K.F., Tolimieri, N., Castillo-Jordán, C.A., and Kapur, M.R. 2019. Status of the sablefish stock in U.S. waters in 2019. Pacific Fisheries Management Council, 7700 Ambassador Place NE, Suite 200, Portland, OR. 392 p.

Background

Kupschuss, S. Independent Peer Review Report on the STAR Panel for Sablefish and Dover Sole, held 25–29 July 2011, in Newport, Oregon.

Stokes, K. Report on the Stock Assessment Review (STAR) Panel for Dover Sole and Sablefish.

In addition to the materials listed above further documents were made available during the review. These can be found in the ftp site listed above.

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 3

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 four 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 external review of stock assessments;
5) increase understanding and acceptance of stock assessments and peer reviews by all members of the Council family;
6) identify research needed to improve assessments, reviews, and fishery management in the future; and
7) use assessment and review resources effectively and efficiently.

A benchmark stock assessment will be conducted and reviewed for Sablefish. The sablefish stock was identified as the top ranked candidate 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.

Sablefish is one of the most important groundfish stocks on the West Coast and the most commercially valuable groundfish stock on a per pound basis. Sablefish is a major target species in commercial trawl and non-trawl fisheries and is readily caught with trawls, longlines, and sablefish pots/traps on the shelf and slope is an important component of the west coast groundfish fishery. The last full assessment of sablefish was in 2011 with an update completed in 2015. The update assessment indicated spawning biomass to be 34.5 percent of its unfished level in 2015. Following the review of the 2011 update assessment, the SSC recommended the next assessment of this stock be a full assessment.

An assessment for the sablefish stock will provide the basis for the management of the groundfish fisheries off the West Coast of the U.S. including providing scientific basis for setting OFLs and ABCs as mandated by the Magnuson-Stevens Act. The technical review will take place during a formal, public, multiple-day 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. The CIE reviewer's duties shall not exceed a maximum of 14 days to complete all work tasks of the peer review described herein.

Tasks for Reviewers

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

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 3 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 3 review meeting in scheduled in Seattle, WA during the dates of July 8-12, 2019 as specified herein, and conduct an independent peer review in accordance with the ToRs.
3) No later than July 26, 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 Seattle, WA.

Period of Performance
The period of performance shall be from the time of award through September 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
July 8-12, 2019	Each reviewer participates and conducts an independent peer review during the panel review meeting
July 26, 2019	Contractor receives draft reports
August 9, 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.

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 3

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 3

NMFS Northwest Fisheries Science Center

2725 Montlake Blvd, NE

Seattle, WA 98112

July 8-12, 2019

TBD

Annex 3: Panel membership and participation

Panel Members

John Field, National Marine Fisheries Service Southwest Fisheries Science Center (Chair)
Jim Ianelli, National Marine Fisheries Service Alaska Fisheries Science Center
Yong Chen, Center for Independent Experts
Robin Cook, Center for Independent Experts

Stock Assessment Team (STAT) Members

Melissa Haltuch, National Marine Fisheries Service Northwest Fisheries Science Center
Kelli Johnson, National Marine Fisheries Service Northwest Fisheries Science Center
Nick Tolimieri, National Marine Fisheries Service Northwest Fisheries Science Center
Maia Kapur, University of Washington
Claudio Castillo-Jordán, University of Washington

STAR Panel Advisors

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