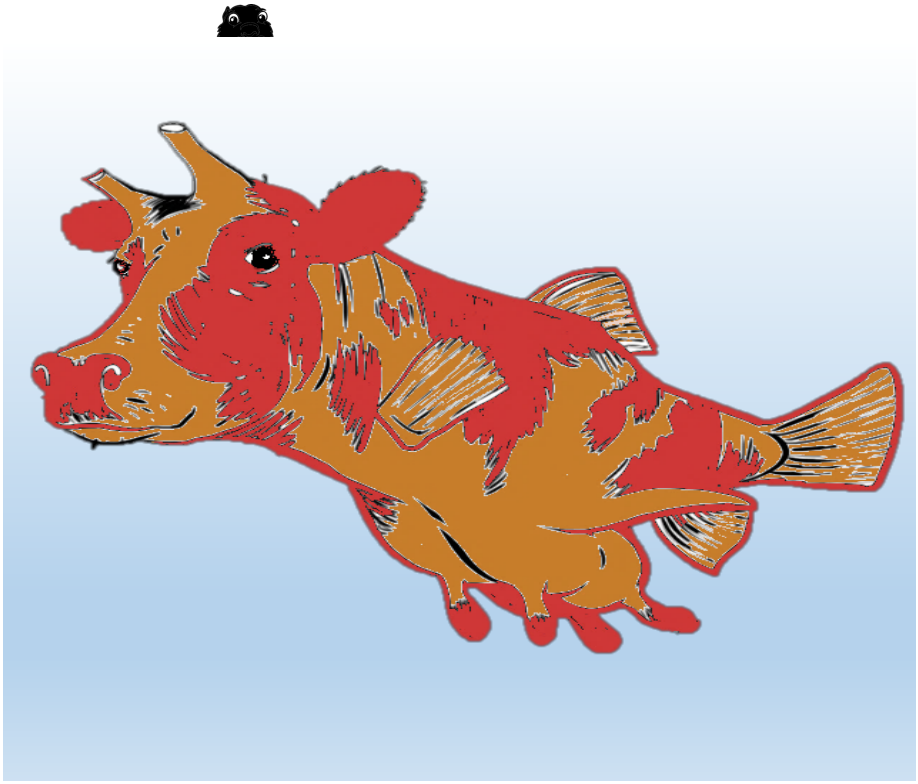
Center for Independent Experts (CIE) Independent Peer Review Report of the Cowcod and Gopher/Black and Yellow Rockfish STAR Panel Review 2019

By

Robin Cook



Prepared for
Center for Independent Experts
Independent System for Peer Review

Contents

Executive Summary

i. Assessments of cowcod and gopher/black and yellow rockfish (GBY) were reviewed at the Southwest Fisheries Science Center, Santa Cruz, CA, during a formal, public meeting of fishery stock assessment experts from 22-26 July 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, which is mostly limited to recent years, and the decision to use stock synthesis as the principal modelling tool. Cowcod appears to be increasing above the management target for depletion while GBY is declining towards the management target. The uncertainty associated with these status evaluations is large.

iii. In the final base model for cowcod agreed at the review meeting, changes were made to the hook and line survey index and a more informative prior was applied to the ROV survey. 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.

iv. In the final base model for GBY agreed at the review meeting, changes were made to the PISCO survey index and the discard fleet was amalgamated with the landings fleet. The number of selectivity curves to be estimated was also reduced. 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.

v. Francis weighting was adopted for both assessments but the sensitivity of the final base model assessments to alternative weighting methods needs to be investigated as it may have an important effect on the results.

vi. The catch data in the assessment are treated as exact and fixed in the model. While this may be a necessary assumption in SS3, it is clearly unrealistic and will affect the perceived stock trend. Alternative catch streams should be investigated as part of sensitivity analyses as well as models that do not assume the catch is error free.

vii. Natural mortality is assumed constant by size and over time, which means the assessments are not able to capture true population dynamics when the stock was lightly fished. This, combined with assumptions of constant growth, fecundity and deterministic recruitment (for cowcod) undermines the veracity of long term stock trends.

viii. 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.

ix. 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. 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. Full MCMC runs can be very time consuming 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.

x. 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 assessments of cowcod and combined gopher/black and yellow rockfish (GBY) stocks.

The stocks were assessed as single units. However, gopher rockfish and black and yellow rockfish are considered different species, albeit closely related and were assessed as combined species forming a unit stock. This is a first full assessment of these species using stock synthesis as the primary modeling framework.

The technical review of pre-STAR assessments took place during a formal, public, meeting of fishery stock assessment experts from 22nd-26th July in Santa Cruz, CA. 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 8th 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 cowcod in southern California and combined gopher/yellow and black (GBY) rockfish. 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 use data quantifying total catches by fleet, indices of abundance, length compositions and age composition data.

Catch data

For both stocks, catch data are an important input to the assessment as they provide information on fishing mortality and can help scale the assessment to the absolute fishable biomass. Unlike many analytical approaches where catch data are treated as observations to which the model is fitted, the model employed in these assessments uses the catch data to drive the model and is therefore particularly influential on the results.

The catch data attempt to characterize removals dating back to the beginning of the fishery which is judged to be 1916. Catch data have improved over the period of the assessment but rely on

reconstructions for the earliest years. This is due to the way landings were recorded in earlier years where these minor species were often not explicitly identified or were mis-classified.

A considerable amount of effort has gone into the reconstruction of the catch time series and these have been reviewed and accepted as the best available. However, ultimately these reconstructions are reliant on pragmatic assumptions about the development of the fishery, species associations, discard rates and discard survival rates. These uncertainties mean that the catch data are subject to error and possibly some bias. In the case of cowcod, for example, the early catch data are estimated to have CVs ranging between 40-80%. 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. The catch is instrumental in determining the stock trajectory from the unfished state to the period when data inform the assessment. If the inter-annual variation in catch and its scale are incorrectly estimated, this will give a misleading impression of stock development during the period of the assessment.

Indices of abundance

Cowcod

A number of surveys were included in the assessment. These include:

- NWFSC West Coast Groundfish Bottom Trawl Survey (WCGBT)
- Sanitation District Surveys
- NWFSC Hook-and-Line Survey
- CalCOFI Ichthyoplankton Survey
- SWFSC Submersible Survey
- SWFSC ROV Survey

The WCGBT survey only catches low numbers of cowcod and is less likely to provide a good abundance index due to the habitat preference of the species for untrawlable grounds, though it is otherwise an effective and important survey for many species. The sanitation surveys are not specifically designed for the purpose of stock assessment but provide a longer time series. The STAT had preprocessed the survey data to extract an index for cowcod using a binomial GLM model.

The NWFSC hook and line survey uses a standard protocol for sampling and will overcome the limitation of the trawl survey in being able to access rocky habitats. The survey area has changed to include the cowcod conservation area (CCA) and the initial base model developed two indices of abundance; one for outside the area and another inside the CCA. While this overcomes the problem of the change in survey area it creates indices that only partially represent the whole stock biomass. During the review meeting the two indices were combined into one on the assumption that the proportion of positive samples inside and outside the CCA were the same. This created an inconsistency with the associated length frequency samples and resulted in the exclusion of samples for the period when the survey only sampled outside the CCA. While this was necessary, it indicates a need to review the index to find the best way of using the available data.

The CalCOFI survey is a long running ichthyoplankton time series that is used in the assessment as an index of spawning biomass. In some years the survey was absent while in others there were no catches of cowcod and this limited the use of the binomial model being applied to individual years to derive a standardised annual index. The STAT overcame this issue by creating “super years” where data from

adjacent years were combined. This was an innovative way to make best use of data but leads to some inconsistency in matching the modelled spawning biomass to the observations. However, as the longest survey time series that appeared to be fitted fairly well by the model, it is an important source of information.

Two underwater surveys of cowcod abundance are available, one in 2002 and another in 2012. These surveys provide and absolute estimate of biomass in the area surveyed. The 2002 survey is included in the assessment as a relative index to account for the area coverage but with a prior on q assuming 25% of the biomass lies outside the surveyed area. Similarly, the 2012 survey is included with an informative prior on q based on the proportion of the stock surveyed. As these two observations are single points, an informative prior is necessary (or q must be fixed) if they are to influence the biomass estimates when q is estimated. These two measurements of stock size appear to be influential in estimating M and hence the strength of the recent stock recovery.

GBY

The pre-STAR assessment included 7 indices but this was reduced by excluding the CCFRP and MRFSS southern indices. These indices represent a small part of the stock and tend to show differing trends from the equivalent northern index. In addition, the PISCO dive survey was split into two following discussion at the review meeting. This was because the length frequencies appear to measure young of the year in the smallest length bins, but this signal is not tracked through the length frequencies in subsequent years suggesting the fish move away. The split was made on the basis of fish up to 6cm and while reasonable, does need further investigation.

For most indices, the final base model fit shows systematic lack of fit though some of the general trends are captured. The PISCO 0-group index does appear to inform the model reasonably well, suggesting that further work to refine this index may be productive.

Length compositions

Cowcod

Available length composition data are mainly from around 2003 onwards and are derived from the WCGBT and the hook and line survey. A few samples from the recreational fishery are available from the 1970s. Sample sizes are generally low with most coming from the hook and line survey.

GBY

Length compositions are available from the commercial and recreational fleets dating from the 1970s, with additional samples from the surveys becoming available in the 1980s and 1990s. The longer time series means that there is the potential to estimate recruitment deviations within the model.

Age compositions

Cowcod

Age data are available from the two sources for which length composition data are available. In addition, a few age samples are available from research projects in the 1970s for the recreational fleet and the 1980s for the commercial fleet. In total, there is very limited age composition data and this will limit the ability of the model to estimate recruitment deviations. The final base model did not attempt to estimate recruitment deviations and hence the ability of the model to fit the age data is weak.

Aging error is taken into account in the assessment model but this does not allow for bias. The species is estimated to have a maximum age of at least 55 years, and it is likely that older fish are subject to both bias and error in age determination.

GBY

Age composition data are patchy and cover some early years of the recreational fishery with additional data available more recently from the CCFRP. A range of miscellaneous age samples was aggregated into a dummy fleet as a way of including age data in the model.

Aging error was estimated using a standard tool which indicates that the error increases with age.

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

Model framework

Both 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 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. Full MCMC runs can be very time consuming 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 length at age. 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 the unfished state since there is little data pre-1980 to estimate growth and the projected trajectory is predicated on constancy of growth over a very long period.

A conventional von Bertalanffy curve was applied with each sex following the same growth curve. For the GBY assessment, both species were assumed to have the same growth. This assumption illustrates the more general difficulty that arises with SS3 where the growth model is not sufficiently granular to be consistent with information in the age composition data, a problem that is exacerbated when recruitment deviations are not estimated. Consequently, there is frequently stress between the information on growth and natural mortality in the two data components that was evident in both assessments at this review.

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. Selectivity is critical to the estimated scale of the stock biomass.

Cowcod

Selectivity curves were estimated for the commercial and recreational fishery and 5 surveys. For the commercial fishery, few relevant length compositions are available so the selectivity was fixed to reflect the maturity ogive. Clearly this is somewhat arbitrary, though necessary, but may be important in the determination of the biomass scale. It will also be of particular importance in the calculation of OFLs.

GBY

The pre-STAR model split the commercial fishery into two fleets, one representing landings and the other discards. Modelling discards as a separate fleet is often adopted in assessment models to overcome problems in modeling the discard process. However, in doing so, fleet selectivity is confounded with post-capture discard selection. After discussion at the review meeting the STAT combined the landings and discard data into a single fleet, and therefore only needed to estimate a single selectivity for the commercial fishery.

During the review meeting, inspection of the estimated selectivities for a number of survey fleets appeared indistinguishable from the commercial fishery and the STAT agreed to mirror these selectivities to one pattern while still retaining a separate selectivity for the commercial fishery. The justification for the latter was to allow this fleet to change in future update assessments. Mirroring at least some fleets helps in reducing the number of model parameters and is a preferable approach.

Natural Mortality, M

Cowcod

Natural mortality is estimated in the assessment 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 and was based on a maximum age of 55. The model appears to be able to estimate M but various model runs investigated during the meeting gave values that varied substantially. Model runs that included one survey at a time suggested the ROV and submersible survey contributed most to the final estimate of M, which is higher than the value used in the last assessment of the stock.

As there is no significant fishery catch at present, much of the recent stock dynamics will be driven by natural factors external to the fishery. In particular, the perceived rate of the recovery of the stock will be heavily dependent on M and the value of R0, neither of which is well determined in the assessment.

GBY

In the pre-STAR assessment, M was estimated but subject to a prior based on the Hamel approach. This assumed a maximum age of 28. The estimate of 0.212 was close to the prior of 0.193. During the review meeting, the estimates of k and M appeared to be highly correlated and the STAT agreed to fix the value at the prior value of 0.193. Given the relative scarcity of age data and typically fairly poor fits to the index data, it is hard to believe the assessment can estimate M well and setting a fixed value appears sensible.

Weighting multinomial data

Age and length 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 of the pre-STAR models to the choice of weighting was investigated using Francis, and harmonic mean weighting with the Francis approach giving an apparently better interpretation of the data. In both cases, Francis weighting was retained in the final base models but needs to be tested when the final base model is subjected to full sensitivity analysis. For GBY, attempts to fit the model using Dirichlet-Multinomial weighting were unsuccessful, illustrating the sensitivity to choice of weighting method.

Recruitment model

The assessments use the Beverton-Holt stock-recruitment function parameterized in terms of steepness, h, and recruitment at unexploited biomass, R0. For these stocks, steepness was fixed at 0.72 to reflect the likely productivity based on meta-analyses.

Cowcod

In cowcod, recruitment deviations were not estimated in the final base model. During the review meeting, a run of the pre-STAR base model where recent recruitment deviations were estimated gave unrealistic estimates. Consequently, there is almost no information in the data to inform the stock recruitment relationship, and steepness in particular.

GBY

In the pre-STAR GBY assessment, recruitment deviations were estimated for the full time series and included autocorrelation. These appeared to be too low for the early period and the Panel suggested restricting the estimation period to 1979 onwards and omitting autocorelation. The revised base model gave recruitment estimates that were far more consistent with the Beverton-Holt model although the 1991 year class remains a very large outlier and appears unrealistic. The plotted stock recruitment relationship suggests there is little information to estimate steepness.

Uncertainty

Systematic sensitivity analyses which consider the principal sources of uncertainty were presented for the initial base model assessments. The analyses consider the influence of a range of modelling assumptions on the principal stock metrics. However, the final base models that were agreed during the meeting still need a full set of sensitivity runs to be completed to demonstrate the robustness of the final 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.

As noted earlier, the catch data in both assessments are assumed exact and are used to drive the model. This may be a necessary assumption for model convergence though it is clearly unrealistic. The problem will be less severe in the estimation of current stock status since this is largely determined by recent data which is of better quality. For cowcod, as there is no recent catch, the issue is not relevant for

current stock size estimates. However, both assessments seek to reconstruct the population biomass and harvest rate back to 1916 and these trajectories may be highly misleading since they are determined by a theoretical value of B0 fixed for 1916 which is then projected forward on the basis of unreliable catch data and productivity parameters estimated for a period nearly 100 years later.

Jitter analyses suggested that the final base models agreed during the meeting converged on the lowest negative log-likelihood.

Stock status

Cowcod

In the absence of fishing there is good reason to believe the harvest rate is well within management limits. The pre-STAR assessment and the numerous investigative runs made during the review meeting all supported the view that the stock is above the management target. The stock apparently shows continuous recovery since 1989, but this perception is heavily dependent on the projected stock given assumed steepness and estimated growth and natural mortality. The ROV and submersible surveys that provide absolute estimates of biomass also support the improvement in the stock.

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.

GBY

Most model configurations show the current depletion to be less depleted than the management target but that the stock is declining. SS3 provides asymptotic estimates of the variance of depletion and the lower bound of the interval estimate lies very close to the management target. Great care is needed in the interpretation of the variance estimates because they assume symmetry in the interval and are derived from a model in which there is clearly misspecification (e.g., fits to the surveys). Furthermore, profiles over R0 indicate that for some values the stock may be below target.

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

General

The reliance on SS3 as the only tool for assessment and sensitivity analysis hides some of the qualities of the data and range of uncertainty. Almost by definition, SS3 will not fit the data if recruitment variability is not included in the model and growth is fixed. These issues need to be explored externally to the SS3 assessment.

A great deal of modelling effort has gone into the derivation of survey indices, but it would be helpful to see additional analyses that explored the properties of the indices themselves to demonstrate they contain useful population signal. It is not difficult to fit a population model to indices alone and this might reveal whether or not a consistent population signal is recoverable.

The use of fixed catches is another assumption that needs to be explored. Fitting a surplus production model with year effects to account for recruitment variability is feasible with these data and would allow for errors in catches to be accommodated. The assessments undertaken suggest there is contrast in the catch and biomass trajectories that may allow such model to be fitted. Such analyses may not be definitive in their own right but may help to identify how SS3 should be configured and the extent to which catch errors need to be considered.

Cowcod

The fecundity of cowcod is not well known and there is uncertainty as to the number of broods individual females produce. In this new assessment, fecundity estimates were revised and led to a major change in the scale of the estimated spawning output. A sensitivity run that allowed for multiple spawning was performed on the pre-STAR assessment which suggested that scale was affected but the depletion estimates were largely unchanged. Clearly, this is an issue which requires further investigation.

GBY

The stock is assessed as a single species and by implication well mixed. However, there is clearly ecological separation between the species by depth at least which would mean differences in vulnerability to fishing. This issue was understood by the STAT but with current knowledge and data their approach is the best that can be achieved. However, further research in this area is desirable with a view to separating the species in the assessment.

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

Given the available data and the adoption of SS3 as the primary modelling tool, these assessments represent the best science available. The reliance on selecting a “best model” based on statistical fit criteria, while clearly sensible, is vulnerable to type 2 errors, especially where there is lack of fit and model misspecification as occurs in these assessments. Exploration of the base model sensitivity helps a great deal in understanding uncertainty but does not fully address the problem of model uncertainty. It would have been helpful to see other models that make different assumptions about population dynamics and statistical error distributions to understand better the envelope of uncertainty.

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

These stock assessments have modest amounts of data yet there is a desire to fit the most elaborate models to the data. If complexity is the preferred direction, then there needs to be greater investment in basic data collection. Age data are probably the most valuable so that a time series of age compositions can be established from which individual cohorts can be tracked. These data need to be collected both from the fishery and from surveys.

It is also of importance to continue to extend the survey time series as these are crucial to calibrating the model. In particular, the CalCOFI, hook and line survey, CCFRP and PISCO surveys all appear to make

important contributions to the assessments. In the case of cowcod, the ROV and submersible surveys appear important and should be repeated at some stage if possible.

Modelling approach

I felt that while model fit was improved marginally during the review meeting, there was no reason to believe the resultant model had better predicative power and this matters when forecasting catches and biomass. While estimated parameters may allow the model to fit the observed data, many are correlated (M and k for example) and error in the estimates will be amplified in the forecast period. 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 aid model stability and predicative power.

As mentioned above, it would have been helpful to see other models that make different assumptions about population dynamics and statistical error distributions to understand better the envelope of uncertainty.

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 for cowcod:

- Re-estimation of the hook and line survey index
- Appropriate creation of super-years for the CalCOFI index
- The prior chosen for catchability of the ROV index
- Investigating whether recruitment deviations for recent years was worthwhile

And for GBY:

- Amalgamation of the landings and discard fleets
- Fixing the value of M
- Splitting the PISCO index to form a recruitment index
- Reducing the years for which recruitment deviations are estimated
- Removal of autocorrelation for recruitment deviations

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 cowcod and GBY represent 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 status of the cowcod stock appears to be above the management target for depletion with GBY above but declining close to the target.

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 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.**

The data available for these assessments is somewhat limited and **I would recommend that more age composition data are collected from the fishery and that the most promising surveys are fully supported. These include the CALCOFI, the hook and line, CCFRP and PISCO surveys. If possible, a further ROV survey should be conducted.**

It would helpful to see other models that make different assumptions about population dynamics and statistical error distributions to understand better the envelope of uncertainty. **I recommend that analyses be undertaken to fit population models to appropriate surveys to understand their utility and to perform assessments with models that make alternative assumptions about population dynamics and error distributions of the observations.**

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, for this reviewer (the continuity reviewer), the timing of the STAR 3 panel meant that only a single week was available to read the assessment documents for STAR panel 4 and complete the CIE report for panel 3. The CIE report deadline for panel 3 fell during the panel 4 meeting. This meant that pre-meeting preparation was restricted. Some thought should be given to scheduling to avoid overlap.

In common with STAR panel 3, the base models at this meeting changed significantly during the meeting and it was not really feasible to look fully at all the sensitivity runs and model diagnostics of the final model. The involvement of the panel in model revision with the STATs during the meeting leads to some blurring of the lines between active participation in modelling and external review. Since the STATs ultimately are responsible for the assessment, it may be better to limit the meeting to review only with

no model revision, but for the STATs to consider the reviewers' comments and then prepare a final assessment responding to those comments.

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

Dick, E.J. and He, X. 2019. Status of Cowcod (Sebastes levis) in 2019. Pacific Fishery Management Council, Portland, OR. Available from http://www.pcouncil.org/groundfish/stock-assessments/

Monk, M and He, X. 2019. The Combined Status of Gopher (Sebastes carnatus) and Black-and-Yellow Rockfishes (Sebastes chrysomelas) in U.S. Waters Off California in 2019. Pacific Fishery Management Council, Portland, OR. Available from http://www.pcouncil.org/groundfish/stock-assessments/

Background

Mohn, R.K. Report to CIE of STAR Panel May 9-13, 2005 Long Beach, CA Cowcod, gopher rockfish, scorpionfish, vermillion rockfish.

Cordue, P. Report On The Southern California Rockfish Star Panel 9-13 May, 2005 Long Beach, California.

Jaio, Y. Center for Independent Expert (CIE) review report on the 2013 STAR Cowcod and Pacific Sanddab Stock Assessments.

Roel, B. A. Center for Independent Experts (CIE), Independent Peer Review Report on the Stock Assessment Review (STAR) Panel for Pacific Sanddab and Cowcod.

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 4

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.

Benchmark stock assessments will be conducted and reviewed for the gopher rockfish and yellow rockfish complex, and cowcod. These stocks were identified within the top five rankings for assessment consideration 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.

Gopher rockfish was assessed for the first time in 2005 and estimated stock depletion under the base model was 97 percent of its unfished biomass at the start of 2005 (Key, et al. 2006). Although the distribution of gopher rockfish extends south into the Southern California Bight (SCB), the assessment was restricted to the stock north of Point Conception. There were no fishery- independent indices of stock biomass for gopher rockfish available at that time and the assessment was based on landings and length composition data from commercial and recreational fisheries (primarily hook and line gear) and an index of relative abundance (CPUE) from the commercial passenger fishing vessels (CPFV) Sportfish Survey database. These data sources were used to estimate population trends from 1965 to 2004. New genetic evidence suggests that gopher rockfish and black-and-yellow rockfish are the same species, and so the assessment will likely be conducted as a complex.

Cowcod in the Southern California Bight was last assessed in 2013 (Dick and MacCall 2013), which estimated stock depletion to be 33.9 percent of unfished spawning biomass at the start of 2013. The 2013 assessment suggested that cowcod in the SCB constitute a smaller, but more productive stock than was estimated from previous assessments. Median unfished and 2013 spawning biomasses were estimated to be 1,549 mt and 524 mt, respectively. The 2013 assessment used the XDB-SRA modeling platform to estimate stock status, scale, and productivity. Dick et al. (2013) fit five fishery-independent data sources: four time series of relative abundance (California Cooperative Oceanic Fisheries

Investigations (CalCOFI) larval abundance survey, Sanitation District trawl surveys, Northwest Fisheries Science Center (NWFSC) 42 2018 Groundfish Stock Assessment and Fishery Evaluation (SAFE) trawl survey, and NWFSC hook-and-line survey), and the 2002 Yoklavich et al. (2007) visual survey estimate of absolute abundance. Cowcod is one of two remaining rebuilding rockfish stocks on the West Coast and is predicted to rebuild by the start of 2019.

Assessments for these stocks will provide the basis for the management of the groundfish fisheries off the West Coast of the U.S. including providing scientific basis for setting OFLs and ABCs as mandated by the Magnuson-Stevens Act. The technical review will take place during a formal, public, multiple-day 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 4 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 4 review meeting in scheduled in Santa Cruz, CA during the dates of July 22-26, 2019 as specified herein, and conduct an independent peer review in accordance with the ToRs.
3) No later than August 9, 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 Santa Cruz, CA.

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 22-26, 2019	Each reviewer participates and conducts an independent peer review during the panel review meeting

August 9, 2019	Contractor receives draft reports
August 23, 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 4

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 4

NMFS Southwest Fisheries Science Center (SWFSC)

Santa Cruz Laboratory

110 McAllister Way

Santa Cruz, California 95060

July 22-26, 2019

TBD

Annex 3: Panel membership and participation

Panel Members

Owen Hamel, National Marine Fisheries Service Northwest Fisheries Science Center (Chair)
Chantel Wetzel, National Marine Fisheries Service Northwest Fisheries Science Center
Sven Kupschus, Center for Independent Experts
Robin Cook, Center for Independent Experts

Stock Assessment Team (STAT) Members

Melissa monk, National Marine Fisheries Service Southwest Fisheries Science Center
E. J. Dick, National Marine Fisheries Service Southwest Fisheries Science Center
Xi He, National Marine Fisheries Service Southwest Fisheries Science Center

STAR Panel Advisors

Melissa Mandrup, Groundfish Management Team representative
Gerry Richter, B&G Seafoods, Groundfish Advisory Subpanel representative
Todd Phillips, Pacific Fishery Management Council representative