# Center for Independent Experts (CIE) Independent Peer Review Report, Stock Assessment Review (STAR) Panel 4. Cowcod and Gopher- Black and Yellow Rockfish

Santa Cruz, California, 22nd – 26th July 2019

**Dr. Sven Kupschus**

**Email: sven.kupschus@cefas.co.uk**

**August 2019**

# Executive summary

Both the cowcod and GYBR stock assessments as prepared for the meeting by the STAT represent a holistic look at the available data sources in a general assessment framework. As such, they are highly informative in describing the relative trends in stock and fisheries development.

The STAR and STAT panel worked on a number of data preparation and evaluation issues during the meeting resulting in both more parsimonious, and at least theoretically less biased models. However, these changes did not materially alter the perceptions of the stocks, both assessments indicating a strong depletion of the population in the 1980s followed by a recovery in more recent times. Both assessments are generally still operating as production models historically with a greater (gopher – black and yellow rockfish (GYBR)) or lesser (cowcod) degree of statistical catch-at-age information included in the recovery of the stocks. The rates of recovery are strongly influenced by the assumption of constant M and k which are poorly defined in the models. Compensatory population growth is therefore facilitated only through a stock recruitment relationship. The age and length composition residuals indicate systematic deviations rather than random variability and model estimates of uncertainty are appropriate only in the latter case. There is little obvious coherence in the index information used and differences in selectivity can only explain a relatively small proportion of those differences so that most indices are considered poorly informative.

From a management perspective, there should be relatively little concern regarding the assessment that the stocks have been recovering and that they are probably both above target levels. Trying to estimate reasonable levels of future exploitation based on the assessment of productivity and the estimate of current stock status is more dependent on the assumptions in the assessment. The sensitivity analysis is only partially informative on the uncertainties. This is especially true for cowcod where there has been no fishery and the longevity of the species and the long-time taken to recovery suggest that a precautionary approach for setting targets when reopening the fishery along with data collections that can systematically address the productivity question would be sensible.

I consider the post-STAR-review assessments the best possible scientific evaluation possible given the available information. Although more could be done to explore the information content of some of the available data sources in my judgement this information would not materially alter the conclusions and would only increase the confidence in the assessments. I would class the quality of the assessment for cowcod as category 2d [1)] and GYBR as category 2d [1)] or 2e [1)] (which ever takes precedence).

[1)] Taken from: TERMS OF REFERENCE FOR THE GROUNDFISH AND COASTAL PELAGIC SPECIES STOCK ASSESSMENT REVIEW PROCESS FOR 2019-2020

**Category d:** Full age-structured assessment, but results are substantially more uncertain than assessments used in the calculation of the P* buffer. The SSC will provide a rationale for each stock placed in this category. Reasons could include that assessment results are very sensitive to model and data assumptions, or that the assessment has not been updated for many years.

**Category e:** Assessments of a complex of species cannot be designated as a category 1 assessment unless there is good evidence that the component species have very similar life-history characteristics and similar rates of biological productivity.

# 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.

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

My role in this review was as an independent CIE reviewer. I read the associated background information provided for the reviews and collated notes as to questions or considerations which I considered relevant for the STAT to answer regarding the appropriateness of the data and the model used to evaluate the stocks. I also read several additional research reports either cited in the assessment reports or discovered independently in the literature to provide a feel for the data quality and the ecology of the species under consideration.

At the STAR panel meeting, I actively interacted with the STAT seeking to evaluate their choices and decisions as well as attempting to prioritise the importance of the different topics in reaching the conclusions proposed by the STAT. I found the STAT team for these assessments knowledgeable and helpful in finding solutions as well as helping me better understand the basis of the assessment. I recommend their efforts in making this an effective review.

I acted as rapporteur for the GYBR assessment presentations and provided the notes to the chair for the purposes of the Summary Report and I reviewed said report for content.

# Summary of Findings numbered ToR for each assessment

The following evaluation of the assessments by ToR do not reflect all of the assumptions and processes involved in producing the assessments reviewed. Most of these were immaterial to the review, and I have focused on those that were specifically discussed and those that led me to my conclusions regarding the assessment. The documents I was asked to review were based on the assessment presented by the STAT as draft. During the meeting small but significant changes to the model were proposed to the initial model (in this review the pre-STAR-base model refers to the model as initially presented) and cumulatively adopted in a stepwise manner to form the final model to be presented to the SSC (referred to in my review as the post-STAR-base model). Because evaluation requests were performed on an evolving model, it was necessary to distinguish between the references as the results of the specific analysis performed may not necessarily be applicable to the pre-STAR-base model and the associated documentation.

## *Cowcod Rockfish assessment:*

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

   I read and reviewed the documents provided ahead of the STAR meeting. I examined figures and tables to explore the consistency of the data with the model assumptions and the fisheries and ecological context.

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

   Generally, I found the reasoning behind choices made for input data and analytical methods appropriate based on principles. However, much of the data beyond the catch data was incomplete / inconsistent either across time or spatial components of the assessment series, so to make the data usable, multiple assumptions were required which were either implemented implicitly in the model or externally as part of the data preparation. The choice of method chosen is in the end somewhat subjective based on expectation, experience and model sensitivity. The panel discussed the benefits and deficiencies of many of the decisions made on a theoretical basis but I found it difficult to formally evaluate these practically by comparing model results mainly because the model seemed to be largely driven by the catch data and the data itself was disparate due to a lack of integration in the data collection program.

   The STAT performed extensive sensitivity analysis trying to determine the contribution of specific data sources using the 'leave out' approach. However, due to the large number of parameters involved and the correlation between many of them, it was not possible to discern the effects of specific information sets beyond some rescaling on the absolute biomass level which was inconsequential on the relative scale on which the management metrics operated. Examination of the stock dynamics in the separate data sources is needed before combining in an integrated model in order to make more sense of how the model is evaluating the data.

   One interpretation of this is that the choices were inconsequential to management advice and that the base model therefore represented a sound basis for management advice. For me, however, this was an indication that the model was having trouble finding a sensible solution to the questions asked, and instead it converged to a default solution very similar to the previous XDB-SRA model despite the substantial increase in information provided. This is not to imply that the new model is not suitable for management in general, but that the ambition of being able to provide more information in terms of stock productivity and selection pattern is not independent of the untestable assumptions made in the construction of the SCA-model. The following evaluation is based on the development process from the pre- to the post-STAR-base model and remaining uncertainties.

Catch data reconstruction:

Catch trends are the major driving force in this assessment, and so must be considered in detail though there appear to be few if any alternatives to corroborate or test the final estimates used in this assessment.

A major issue with the historic catch data has been that the market categories for most rockfish species have changed over time in different ports and that species due to previous management have been recorded as mixed categories. In past assessments, historic catch reconstruction was carried out by individual assessors independently resulting in concerns that the summed catches across stocks differed from the total rockfish catch. A formal reconstruction (Ralston, 2010) was carried out for all rockfish stocks based on species ratios in sampled catches that were expanded by market category and location to the reported landings. However, the reconstruction did not cover the cowcod management area so that a different reconstruction 1916-1968 had to be formulated using similar methodologies but more extensive datasets.

Cowcod are a high value market species and have generally been recorded in market categories where their contribution has been relatively high (large ratio of cowcod to other species) so that concerns over high variability in the raised catches due to the ratio estimator are minor compared to 'rare' species. An analogous methodology was applied to the intervening period 1969-1983 using an updated dataset.

Since 1984, cowcod market categories have been more consistently recorded but cowcod have generally been reported in categories where the species represents a smaller portion of the catches than previously, and sampling of different categories has been more intermittent resulting in a greater uncertainty due to a smaller ratio estimator (variability) and the need to 'borrow' ratio estimates from other categories (bias). Landings of cowcod since 2001 have been prohibited. While uncertainty remains in the landings estimates, the scale of the estimates of the data appear representative of the scale of catches and variability is likely on the interannual scale as opposed to biases across multiple years.

A sensitivity analysis of the base model allowing variability on catches, accounting for the different aggregation periods would be of interest if there was more size and age data available in this period. The sensitivity analysis prepared on catch scaling is uninformative as it merely results in a rescaling of the assessment. However, for the purposes of this assessment, I consider the commercial landings data appropriate.

Recreational catches:

Similar to commercial catches, early estimates for the recreational fleet are based on compositional data from the 1970s to 1980s. The report acknowledges that this assumes that the fishing practices are the same historically as during the latter period. Not formally acknowledged is that it also assumes that the relative abundance of cowcod to other rockfish is also the same, whereas the assessment assumes that the steep decline in the abundance of cowcod has already started during the latter period.

Given the high desirability of large fish, the assumptions that discarding in the recreational fishery is negligible as evidenced in the 1980s is sensible.

Recreational catches 1984-2000 were estimated from the MRFSS data with some assumptions necessary for converting numbers caught in some strata to the weights and interpolating across some years with missing data. Overall, the impact of these assumptions on the relative trend in catches is difficult to quantify, but because of the relatively small scale of catches during the periods where gaps in data existed, they are unlikely to have major impacts on the assessment trajectories. Catches since 2001 for the fleet are release mortalities and are negligible on the scale of the assessment, and so are appropriately treated here.

Age and length information from the fishery is generally sparse but critically only available for certain periods of the timeseries, and it is unclear to what degree such information is representative of the selectivities of fisheries in other periods. In addition, the contrast in stock abundance is poorly covered, making it difficult to attribute changes in length and age distribution with changes in stock size. Generally, the model found it hard to match length and age information and much of the information is inconsequential to the model outcome making it difficult to assess the suitability to estimate selectivities.

Bias correction to the same reader (through the whole time series) has been carried out, which means aging is at least consistent. However, ages have not been verified and there are few data where lengths are very informative on age. As the assessment is not really tracking cohorts, this is less important in this model; however, it is still critical to the external estimate of k and the prior on M which are strongly influential in determining the recovery rate of the population.

<u>Fisheries independent indices:</u>

The data selected for use as indices in the pre-STAR-base model is generally suitable, but there were questions regarding the methodologies use to develop them into indices. Firstly, separating stock components for the NWFSC H&L survey (inside and outside the Cowcod Conservation Area (CCA)) when these partial-indices are potentially divergent in trend but penalized against a single population could potentially exclude the information from one or both partial-indices even if when added together they might be highly informative. For the Sanitation district trawl survey index and the trawl and CalCOFI-ichthyoplankton surveys, data was binned into “super-years” which will have a tendency to obscure cohort tracking as well as potentially be sensitive to the choice of years within a block.

<u>NWFSC H&L survey:</u>

This survey is the most important contributor of length and age information as well as a potential indicator of stock recovery through CPUE trends for this assessment. However, the survey in the pre-STAR-base was treated as two timeseries in the assessment (inside and outside the CCA) because of a desire to keep the consistency in the time series intact while taking account of the more recent data 2014 onwards from within the CCA. A more reasonable approach that avoids inappropriate conflict between partial indices would be to model a combined index weighted by the proportions of the area inside and outside the CCA, particularly since the atrial indices were modelled logistically anyway to account for differences between stations, depths, and hook deployments, etc. The STAT conducted some analysis suggesting the partial trends were roughly orthogonal and developed a couple of potential solutions for dealing with the differences in the associated size and age distribution. These potential replacement indices were tested in exploratory assessments based on the pre-STAR-base model.

<u>Sanitation District index and CalCOFI ichthyoplankton index:</u>

All of these indices, although monitoring recruits / juvenile abundances, inform the model backward on spawning output rather than forward on juveniles. Available data are sparse and multiple observations are rare suggesting a binomial approach is preferable under current conditions (future changes in biomass may necessitate revision of the preselection of areas/sites to avoid ignoring distributional expansion and / or saturation of positive tows. The Bayesian binomial modelling approach used struggled to compute mean probabilities since some years did not contain any cowcod, so samples were grouped over 5-year periods. This assessment poorly tracks cohorts so the “misinformation” provided on recruitment by grouping years is unlikely to have consequences. The biomass fluctuations in the stock should be strongly autocorrelated in time because of the large age range in the stock Grouping *per se* is arguably not an issue in the current assessment. But the model's ability to track cohorts is not helped by the treatment of the index.

Absolute biomass indices:

Treatment of the absolute abundance estimates was appropriate given that there is only a single value in each of the time series. In addition, the ROV index was restricted to the CCA, which previously had used an uninformative prior to estimate the proportion of the population outside the CCA based on the recreational fleet. During the panel, a more informative prior was developed based on the uncertainty of the proportion inside and outside the CCA in the SUB-survey. The new treatment is inconsistent with the assumption made for the H&L survey concerning the length distributions inside and outside the CCA. Overall, it made relatively little difference to the parameter estimates because there is little information in the model that contradicts the estimates. The new treatment seems to be the best possible compromise in informing the model.

My personal feeling is that these absolute indices of abundance should be considerably more variable than estimated based on the sample variability alone. Given the extremely small proportion of the population that is actually observed, the uncertainty estimate used would only be representative of the population if site selection was random: however, it ignores the lack of mixing in the population, the relationship between relative abundance and habitat, and the uncertainty around the quality of habitat maps. Additional data points using different survey designs would provide more confidence in the appropriateness of using these data as absolute estimates of abundance. Alternatively, a more detailed analysis of the H&L survey using the observed ages (all fish caught are aged) could provide a better assessment of the uncertainty in the rate of stock recovery than is currently made in the assessment, due to the uncertainty around growth in conjunction with the use of conditional age data.

Natural Mortality:

Natural mortality has been estimated based on the maximum observed age and this estimate is used with an uninformative prior in the model which moves the estimate. The value was not very different from a rough estimate of the Z curves and this was given as an argument for using the former. However, Z curves are estimates of total mortality and not just the natural part. Since the original Z curves were calculated based on the period where exploitation was still occurring, it is not surprising that the model is estimating the M to be lower, so the prior must ultimately have a significant impact on the model estimates of M. The model struggles with the apparent conflict between the compositional data and the starting value for M.

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

The reasoning behind moving from the much simpler biomass production model with a single index to a more complex integrated SCA model was to obtain better information on fisheries selectivities, better estimates of stock productivity and to use all available data effectively. Though laudable, these objectives have a number of consequences in data preparation. A substantial number of assumptions were necessary due to the general sparsity and more importantly the unbalanced availability of length and age samples. Some of these assumptions have the potential to influence results. For example, the composition of the commercial fleets (in terms of gears used) is not uniform over time. If as one might assume that hook and line fleets have a different selectivity from the net-fleet, then selectivities will at best reflect the average of the two fleet compositional samples which are unlikely to be representative of the combined catches. Moreover, unless future fleet composition represents the same proportions, the combined selectivities have little use in forecast and management decision making, which is one ambition of the integrated model.

As it turns out, estimating selectivity for the commercial fleet was not possible within the model anyway so the STAT arbitrarily used the maturity ogive. Undoubtedly, the lack of informative length data in conjunction with the complex dependencies of M, growth and q were major factors in this. More fundamentally with the changes in fisheries over time, particularly the expansion of the net fishery, it is clear that a single selectivity curve for the commercial fleet will introduce unavoidable process error. The parameter uncertainty in selectivity and its relationship with M were crudely tested in the sensitivity analysis by fixing the logistic location parameter at different lengths while also fixing M at different levels. This is an appropriate assessment of the unknown uncertainties given the model, but is contrary to the objective of providing better information on selectivities because it fails to consider the process uncertainty.

More generally, there was insufficient data to actually test the suitability of many of the assumptions quantitatively, and somewhat surprisingly, credibility of different options seemed to be based predominantly on the perception of stock status (SSB / reproductive potential) and the comparison to the previous assessment results.

Overall, the SCA-model appeared to treat the data as two assessments integrated by a number of parameters that were loosely defined. This was largely an artefact of the imbalanced data collection, but also because the early part of the model dealt with the depletion of the stock, while the latter focused on the recovery. At least theoretically, we would not expect populations to have the same productivity parameters given the difference in the dynamics.

An exploratory model presented at the panel, for example, indicated that blocking growth in the two periods significantly improved the model fit to the available composition data. While acknowledging that the selectivity for the composition data may not have been the same, as it came from different fleets as felt by the majority of the STAR, I considered the model should have been able to adjust for this using the selectivity parameters if it was just a differences in selectivity and not a change in growth. I feel more could have been done to explore this using different selectivity settings.

Similarly, the Z-curves produced during the high exploitation period versus the closed period did not indicate a big difference in total instantaneous mortality rate (F+M). And in fact, both rates were quite close to the value of M estimated in the assessments. Sensitivity runs on estimates of M were limited to constant increases and decreases in M, so are not representative of the uncertainty around whether M has changed. The old production model does not deal with M separately but an integral of the population response, and so can deal with the interactions more adequately. The integrated model has strong correlations between some of the productivity estimators, so that some had to be fixed, which would make it difficult for the model to reconcile all the data sources if M or k had changed.

As a reflection of relative biomass trends in a production model, the assessment appears to be performing reasonably, the changes that have been suggested to the pre-STAR-base model largely deal with developing more robust methodologies / implementations, but do not materially change the trends nor the perception of the stock status relative to virgin biomass. It does leave some minor inconsistencies necessitated by the data availability, such as assuming the same selectivity for the population inside and outside the CCA for the ROV survey, which is dependent on the actual distributional differences which although not likely to be influential could be inappropriate.

I do not consider viewing the integrated SCA model in absolute terms with the ambition of defining selectivities and stock productivity warranted. In particular, the rate of recovery of the stock since the closure is more uncertain than the uncertainties in the assessment and the provided sensitivity analysis suggest, because: Conceptually, the model is removing around 50% of virgin biomass in a matter of 8 years. This is thought to have been a consequence of the rapid expansion of the net fishery with a much higher q, therefore not requiring a major expansion of the fleet in terms of numbers of boats. However, cowcod tend to be closely tied to a territory which would mean that the fishery would have to cover wide and diverse areas in order to access such a large portion of the population without mixing, unless the territories are strongly aggregated in a small number of areas. I find the rate of change in the exploitation rates not entirely credible.

The Z curves corresponding to periods of high and low exploitation, though making some quite general assumptions, tend to suggest that the difference in exploitation is rather smaller between the periods than expressed in the assessment and in fact closer to M, which could be interpreted as larger population in absolute terms than assessed at least in productivity models (but at least the ROV survey provides some evidence of the absolute scale of the population). Similarly, the historic recreational trophy size of fish had been decreasing slowly. Although one can argue that trophy size only looks at the very largest of fish and therefore is hyper stable on the decreasing slope it should still decrease much more rapidly in a mixed population near 10% of virgin biomass with the main fleet selectivity being asymptotic. At reduced exploitation, the trophy length should increase slowly in contrast to the observed trends in the paper.

Examination of the post-STAR-base model evaluating the individual index information (leave one in) indicated that it was the relationship between the two absolute estimates of abundance (ROV and SUB indices) that defined the model estimates of M and hence the recovery rate of the stock. The H&L survey tended to estimate higher rates of stock recovery, but in the final model was down weighted likely due to conflict between age and length comps associated with the poor fit of a single growth curve over time.

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

   My comments with regards to options for improving the pre-STAR-base model given the data are described in relation to the pre-STAR-base model in TOR 3, many of which have been incorporated into the post-STAR-base model. Remaining deficiencies or uncertainties predominantly depend on the available data but do not permit quantitative evaluation, so alternatives are largely a matter of opinion. I have tried to highlight where the combined opinion of the panel differed from my own perspective. By and large, more data are required to critically evaluate the differences in perspectives and, when defining research needs, the specific types of data should be defined at a level of specificity that enable such comparisons in future.

   The CalCOFI ichthyoplankton index seems to be one of the more influential indices in the assessment despite observing only a very small portion of the larval population. In fact, the data is so sparse that it was modelled as a Bayesian binomial model grouped over “super years” covering differential periods with the weight assigned to a mid-period year. Future work should consider:

   - Whether the years to group has some influence on the model outcome. For example, high catches in two adjacent years could provide a distinct peak if considered in one “super year” while the estimates would provide much less contrast when considered in adjacent groups, similarly the terminal group will change over time in update assessments and could change the basis of the assessment input.
   - Using a model to determine the variability external to the model has some benefits in providing precision estimates but circumvents the integrated assessments of probabilities and in this case necessitates the grouping into super years. In addition, the index modelled index has a different error distribution than assumed in the SS3 model which could potentially lead to misinterpretation of the likelihoods.
   - It is acknowledged that there is considerable variability in the spawning success due to environmental conditions and in many ways the “super year” approach can ameliorate this concern if it causes interannual variability evenly throughout the time period. The

more episodic the environmental conditions are, the more problematic this effect will be. Future models should consider using environmental covariates to improve the linkage to SSB particularly if there are long-term environmental trends.

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

   The pre-STAR-base model as presented contained some technical inconsistencies with regards to its treatment of fisheries independent indices, most of which were revised to my satisfaction during the meeting. However, the changes did not materially alter the stock status estimates or the management metrics. Given the data, I feel the assessment provides the best scientific information available. The SCA-model treats stock productivity explicitly, but it is clear from the model that the parameter estimates, and their interactions are poorly defined and treated like a general production function. Consequently, it is not possible to credibly interpret these parameters so the assessment should be treated in management as a category 2 assessment. The axis of uncertainty use in the sensitivity analysis are, in my opinion, more informative of the uncertainty than the statistical evaluation of the assessment likelihoods since the unresolved process errors seem more prominent than in other models I have reviewed.

   This report so far has focused on the assessment shortcomings largely due to data availability which should not be misinterpreted as an indication of suitability for providing advice, more as a means of expressing the process uncertainty remaining in the evaluation of the stock dynamics.

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.

   Models can always do better with more data because one can statistically afford to estimate more parameters. During discussions at the meeting, I could not get a clear steer on the objectives for this assessment beyond the current need to establish the stock status. While I feel the assessment reasonably accomplishes this task, it seems that future objectives are crucial to providing advice going forward.

   The panel did provide a wish-list of data and analyses that would address certain weaknesses in the assessment, and this was based on the analytical uncertainties in the model, not necessarily the uncertainties that are likely to impact the quality of management.

   While I can see that the research needs as described can improve the model, they are neither prioritised in terms of what is important for management nor in relation to the cost-benefit for improving the model. Furthermore, the elements of the list are provided very generally (collect age information) but will actually require multiple data sources. From the list it is therefore not possible to optimise the sampling efforts across different objectives resulting in unnecessary duplication of effort, and fails to provide the coherence across studies and objectives. Insufficient time was dedicated to the tasks in the review.

   For example, if a better growth function is required, more age data is needed and this could be provided most efficiently through additional otolith collections in many existing data collection programs. Benefits (and likely costs) will vary. For example, collecting many more ages from a fleet that predominantly samples young fish may not be particularly helpful. Increasing survey sampling may be more representative of the entire length distribution and is likely to be less effected by autocorrelation in sampling due to its generally smaller sample sizes. More importantly, it may permit using the age data directly in translating the SSB index into the cohorts which would provide a lot more information on future stock trajectories than the current model. However, it would likely do little to improve our understanding of the commercial selectivity needed to interpret the OFL from the model.

Without more consideration of the actual objectives, I find it hard to make more specific recommendations on data needs beyond suggesting that it is essential the managers more explicitly express what their management needs are for this stock. With the current list as provided in the STAR-report, I see rapid progress in the assessment unlikely since so much hinges on being able to tie the data collections together consistently in the assessment framework. Much of the uncertainty in the current assessment is caused by the need for assumptions on how the data link and only a more integrated approach to future data collection will avoid having to make the same or new untestable assumptions.

In my opinion, most relevant to the management is to test the assumptions of constant stock productivity over the full assessment period and the sources of evidence for changes in growth and mortality. This is particularly relevant given that a much simpler production model seems to be adequately replicating the results and uncertainties of the SCA-model. If production has been variable over time, it seems unhelpful to retain the long time series of catches and a shorter cohort-focused model will be more informative on future stock trajectories. If things have been constant a much simpler production model seems to be adequately replicating the results and uncertainties of the SCA-model so a focus on age data may not be effective.

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

   Relatively little consideration given of research needs and often not at level of detail sufficient to implement in a way that is sufficient to address the objectives or to evaluate the integrated impact across different research activities to ensure efficient data collection across multiple objectives.

## *Gopher – Black and Yellow Rockfish assessment:*

Gopher – Yellow and Black rockfish assessment:

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

   I read and reviewed the documents provided ahead of the STAR meeting. I examined figures and tables to explore the consistency of the data with the model assumptions and the fisheries and ecological context.

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

   Gopher and Black and Yellow Rockfish are considered a species complex due to the strikingly distinct colouration, so it is important to consider whether a stock-complex such as this that can be reasonably assessed in a single assessment is valuable, and if so, what additional considerations are needed in the derivation of advice. The similarity in growth, ecology and the similarity in distribution despite some depth and latitudinal differences suggests at least that the different components of the complex are likely to respond similarly to exploitation or other pressures. Despite evidence of geneflow restrictions, the two species are the only rock fish species that are currently not uniquely identifiable genetically. The recent divergence of the species may explain the similarity in life history parameters. It is desirable to work towards separating the assessments, particularly the catch data, in the future. Given the currently available data, and the similarity in the species, its more beneficial for management to work towards a stock complex assessment with more certain catch data than trying to develop two species specific assessments with uncertain catch data limited only by the sum of the catches.

This is likely to require additional precautionary measures, since the live fish fishery does tend to target the more strikingly coloured BY species which tends to represent the younger / smaller individuals in the fishery.

Commercial Catches:

Like other rockfish, the complex has over time been landed in different market categories, so the assessment made use of the Ralston (2010) reconstruction for data up to 1968. This is the best in-depth analysis on the topic, though it leaves considerable uncertainty in the estimates due to the small proportion that the complex makes up in the different market categories resulting in a small ratio estimator. 1969-1980 followed an analogous methodology but used the original data sources since the data has not been formally included in a data base. Data since 1980 is available from two different data bases (CALCOM (state) and PACFIN (federal)). For this stock complex, the data should be identical with comparable extraction methods, but the two data sets provided different data particularly with regards to the number of lengths. PACFIN was used as the final version because it provided a greater range of lengths. It seems essential that the differences are explored and explained as the length data (i.e., estimation of growth) is particularly poor in the assessment and this could be contributing to that uncertainty. Also, CALCOM data forms the basis of 1969-1980 data, and so is potentially inconsistent with the subsequent period, relevant only if it is a data versus a methodological issue.

Discarding in the fleet is relatively small proportion and needs to be sampling depth adjusted to account for discard mortality rates varying with depth. It was possible to achieve this back to 2003, and for the previous period the average rate for the commercial catches was applied. Because of a lack of length composition data in the previous period, this was added as a separate fleet in order to be able to monitor the commercial fleet with a single discard selectivity. Using the preferable retention ogive was not possible due to the inconsistency in the way the landings were reported.

Unfortunately, as implemented in the pre-STAR-base model, this results in loss of connectivity in the information and they could theoretically at least be representative of different levels of fishing effort. It was also thought that this effect may be reflected in the banding pattern in the residuals from the fleet trying to bend around a substantially sharper curve at the bottom of the logistic caused by the sharper retention ogive. Therefore, the STAT was asked to combine the discard and catches in a single fleet and also combining the length compositions weighted by the ratio of discards to retained in numbers of individuals. Length compositions prior to 2000 were not affected by discarding, while 2000-2003 length samples were only representative of the retained portion, which is why it was necessary to remove the length composition for those years. This also removed the need for the selectivity time block used in the pre-STAR-base model.

The post-STAR-base is a more parsimonious solution to the length comp issue and is more representative of the process of discarding. However, it still did not improve the banding in the length composition residuals probably because the catch was selected logistically, but discards were only partially counted towards mortality. Other inconsistencies remain in the treatment of these data, but they are not thought to be serious since they have little or no overlap with the period during which recruitment deviates are being estimated.

Recreational Catches:

In previous assessments, the Southern recreational catch had been estimated as unrealistically large for the period 1966-1980 based on the use of a ratio estimator from the north which had seen a disproportionate expansion in that period when compared to the south. When a separate catch estimate was available in 1980, this produced a significant discontinuity in the catch stream. A smoother transition was developed using a ramp that although not very realistic, is unlikely to have the highly influential effects in the assessment caused by the extrapolation from the northern data since it is likely that southern catches have been small compared to catches from the north since the 1960s. I consider this the best that can be done at this time.

Associated length distributions of catches and discards of the recreational fleet were compiled from various monitoring and research observer programs while ensuring that these were consistent with the relevant selectivities and retention ogives.

Indices:

A number of fisheries dependent indices were used in the assessment derived from observer programs (either onshore or on the boat) focused on the recreational charter boat fleet. Two observer programs were split into southern and northern components because they could not easily be combined into a single population. However, they were providing conflicting trends with the southern indices despite being only representative of a small portion of the stock. The southern indices were removed entirely from the assessment in the post-STAR-base model. Because the data did not follow a standard sampling design, it was necessary to model the catch rate as a function of the observed trips and the drops made. A significant amount of prefiltering was conducted in an attempt to standardise to the fishing activity around the habitats where GYBR occurred.

Fishing locations were overlaid with reef information and drops, with starting locations a certain distance from the reefs were excluded as not being the right habitat despite the fact that a small number of individuals were captured. Similar drops that finished the same distance off the reef as the others started from the reef could not be dropped because ending locations were not available. I think there is a risk that this approach potentially can create biases or misinterpretations unless a drop ceases more quickly if initiated on the reef before dropping off compared to the other way around. Either way, it is necessary to assume that the number of drops spending little or no time on the reef is roughly similar between years and distributed spatially.

Data on the catch weight of GYBR were not available in all years and these years were dropped from the indices. From this, it is difficult to ascertain if there were no catches, which would be indicative of low biomass, or if there were catches, but the necessary data was not available to turn these into weights for the index. The indices as used in the assessment implied the latter was the case.

The PISCO index exhibited a very peculiar length distribution with a substantial gap in observations between the very young (0 and 1 group) and fish greater than 25cm in the original pre-STAR-base model had been modelled as a single index with a logistic regression. This resulted in a strong residual pattern in length. It was suggested the STAT removing all fish less than 15cm (the juveniles) from the index and to use the individual less than 8cm which were identified as being almost exclusively 0-group in a separate recruiting index to help the model fit rec-devs. An age selectivity was chosen since the growth model was not able to identify the fish as 0-group based on their length alone. Year where no positive hauls were observed could not be modelled in the logit model and the STAT had to resort to replicating the lowest value and its associated CV to indicate to the model that abundance must have been low.

The splitting of the index is not necessarily ideal as it essentially splits a dataset and provides extra weight to the data. In this case, the data is split and representative of non-overlapping ages, so I feel the post-STAR-base model is justified in this case. More importantly, it now provides a better index of recruitment that previously was drowned in the signal from the adult biomass signal within the combined as well as blurred by the poor resolution of the growth model.

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

The model is considerably more data rich than the cowcod assessment in terms of sheer numbers of data points. However, it is only marginally better in terms of its temporal data balance and the data types that stretch across the period of highest exploitation. The difficulty in sensibly tying the data sources together because they largely represented ad-hoc collections as opposed to monitoring data. This is also the reason that most of the age information ends up in the conditional dummy fleet where it is least effective on informing on cohort strength. In addition to his temporal bias it seems to also have a species bias in the age data (young = BYR, old = GR) and this may well be partly responsible for the conflict between ages and lengths.

Despite the comp data, the model is still in essence a biomass dynamic model based on accurate assumptions. The period of highest exploitation in the 1980s extends over multiple years, but I was suspicious of the rapid rate of increase in the exploitation. Z curves produced by the STAT on request for the periods pre- and post-2000 confirm that the earlier period was on average reflective of higher total mortalities and that even the scale was reasonable with respect to the external estimates M. The proximity of the species to the shore and the lack of specialist equipment (i.e., H&L), particularly the live fish fishery, may make the species more susceptible to rapid increases in effort. While it seems possible that the fishery really expanded so quick, it is also plausible that the assessment is struggling with trying to tie together two periods of data with very different information contents with the transition being interpreted as an excessive exploitation rate.

The pre-STAR-model had a number of deficiencies particularly with respect to its complexity. Much has been simplified at the meeting with multiple fleets being matches, for efficiency and stability and discard ogives / fleets eliminated. This undoubtedly will have introduced some process error into the assessment, but judging by the diagnostics, surprisingly little was evident.

Further simplification of the model was undertaken with the recruitment deviates. This now starts considerably later, which means fewer parameters, smaller recruitment bias adjustment, and more realistic recruitment patterns, especially after also removing the autocorrelation on recruitment. Interestingly, with a reduced period of recruitment deviates to play with before the new PISCO index takes over, the model has had to create one very large recruitment year in 1991 which previously had been split over several years in order to arrive in the fishery by the time the length frequencies shift to smaller fish. The year 1991 is close to the smallest observed biomass in the assessment, and so produces a large outlier in the stock recruit relationship, the effects of which should be examined in more detail.

More work should also be done to investigate whether one or multiple strong cohorts appeared in the fishery at that time. This is important to evaluate if the treatment of the rec-devs as implemented now is justified, but it also could identify if the poor specification of the growth function is hampering the assessment process. As an indication, during the sensitivity runs, k was fixed at slightly different levels around the uncertainty which shifted the estimation of the strong cohort from 1991 to 1993. While this is not surprising from a growth perspective it does suggest that age compositional data which should be informative is either not present or ignored.

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

My comments with regards to options for improving the pre-STAR-base model given the data are described in relation to the pre-STAR-base model in TOR 3 many of which have been incorporated into the post-STAR-base model. Remaining deficiencies or uncertainties predominantly depend on the available data but do not permit quantitative evaluation, so alternatives are largely a matter of opinion. I have tried to highlight where the combined opinion of the panel differed from my own perspective. By and large, more data is required to critically evaluate the differences in perspectives.

The major sources of uncertainty in this assessment are described by the STAT as being around recruitment (R0) and (k) and a sensitivity of these values is provided for the post-STAR-base model. Not surprisingly, these are the part of the productivity set. If the model describes productivity adequately, then while the individual parameters may be highly uncertain, the combination of them would be much less uncertain were it not for the fact that this is a species complex. Unfortunately, I was unable to come up with a reasonable way to test for sensitivity to the single species assumption. This has to be the major axis of uncertainty for fisheries management in the longer term. Certainly, the STAT should consider whether a more precautionary biomass reference point is not sensible for a stock complex.

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

The pre-STAR-base model as presented was data poor, not necessarily the quantity of data, but independent way the data was collected and its representation of the stock complex with imbalance in the data for the two species. In contrast, the model was overly ambitious in terms of the parameters being estimated many of which were strongly correlated. During the meeting, STAT and STAR together developed a more parsimonious approach and provided some guidance on the methodological improvements to the model. The resulting post-STAR-base has reduced correlation amongst some of the parameters of interest and has gained in stability and hence confidence in the appropriateness for management, without large changes in the perceived long-term trends in stock dynamics between the two models.

I find the new model to be the best scientific information available given the data, and class it as a category 2d or 2e dependent on which takes precedence in the classification since both are applicable. Essentially, this is still a biomass production model despite the application of an SCA-model with recruitment deviates estimate. The comparatively poor fit to the composition data and remaining residual patterns in length suggest that the model is not able to track cohorts. The imbalance of the data and the stock complex characteristics of this assessment mean that managers must consider additional process uncertainty when developing management measures based on the outcome. The sensitivity analysis provided in my opinion was only able to deal with process uncertainty associated with the correlation between some of the parameter estimates particularly the productivity ones. It does not resolve the uncertainty around differences in the dynamics of the two stocks.

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.**

The issue of the species complex is clearly something that will require further investigation. The similarity of the life histories, ecology and genetics suggest that consideration as a complex is appropriate for advice in the short to medium term. However, there are some indications that longevity is different, implying different levels of productivity which is likely to have bigger long-term management implications, especially if the management aim is to optimise yield.

Beyond this, it seems difficult to make a list of very general comments based on the quality of the assessment. These cannot be prioritised in terms of their effect on improving management without a clearer steer on what the aims of the management are. I agree that the list of research needs of the summary report will aid the development of a more coherent assessment and probably management. It is unlikely that there will be sufficient funding for all the research needs across all of the stocks, and it certainly will not be efficient if it were implemented. So, it seems unlikely that without a more integrated approach to funding research, much progress will be made by the next full assessment.

I bring this up here because this stock, more than most, exemplifies how the lack of an integrated approach to data collection leads to very disparate data that is difficult to combine in an assessment, and requires untestable assumptions which ultimately end up as uncertainties in the sensitivity analysis. As an example, a call for more age and size at age information can have multiple benefits. Age information could lead to better estimates of recruitment / cohorts or better estimates of fishing mortality if they are the older ages in the fishery that aggregate F across many years, and so help build towards a more detailed SCA model. But these data are also informative on productivity through growth, and so could help a biomass production type models if they considered the impact of stock size on growth specifically.

How you would collect the data for these different purposes and their likely impact on management is then dependent on the management objectives. There is little point in collecting more information on ages to estimate cohort strength as long as the aim is to maintain the long time series of landings data, because it is likely to have any impact in the foreseeable future in a production model or the assessment as it currently exists. Just like getting more information on the species, contributions in the complex is of little value since it appears the historic data cannot be split reliably by species. Information on growth helps the biomass production models much more than a SAC model which tries only to use it as a means of estimating recruitment (since growth is fixed). But if the aim is to move to more of an SCA approach implied by the SS3 assessment used here, then it is a much lower priority.

Efficiency is important in data collection as it is the most expensive part of the management process. We can see from the above that data on ages potentially can aid both approaches though would collect them very differently. Being more specific as to how to collect the data is important. Sometimes this can be done as part of fishery dependent collection, sometimes surveys both would be more efficient than setting up a separate collection program, but it needs to be clear if this is helpful or a hinderance as it could be that the selectivity of the fleet is poor or biased for the size or the age of greatest interest to the research outcome.

The aim of this section was not to discuss the specifics of the data needs, as I found this difficult to judge and by no means exhaustive of possible research needs. It is aimed at illustrating that the possible research needs are vast and that progress can only be made through a more structured integrated approach and should not be the last thing that is done in a very short period of time at the end of the review if significant progress is to be made by the next assessment.

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

I found the discussions around the technical elements of the assessment implementation very constructive and informative. The STAT was knowledgeable and helpful in implementing the requests from the STAR, and virtually all the meeting focused around these elements despite the fact that it was clear that the model was unlikely to dramatically change its perception of the historic stock dynamics.

More attention on the evaluation of the stock trends implied by the individual data sources and greater clarity on representativeness, i.e. the imbalance between species, etc., would go some way to understanding why the assessment is relatively insensitive to the treatment of the indices and the compositional data.

At the other end of the process, virtually no time was given to the consideration as to how to derive future catch options or how to formally consider the uncertainty in management. The sensitivity analysis only provided a range of what might be reality in terms of M and k. I leave it to the managers to consider whether more formal discussions about how to interpret the outcome of the assessment along with a formal categorisation of the assessments and an in-depth look at research needs would have been helpful to managers.

The lack of more consideration of the management objective makes it likely that future assessments (really applies to both stocks) will struggle with the same or similar problems in future. Much of what is collected right now is based more on consistency of data collection than on appropriateness/utility, and some of it is certainly not efficient. Deciding on what would work better is not quick. Usually this topic does not get the attention it deserves and having come back to some assessment in the CIE review process, I find it difficult to see what has changed other than the model used. In this review, the topic received even less attention than in other reviews and these stocks need it more than most if a meaningful progress is to be made towards higher 'categories' of assessment quality. If not, then it seems likely that one could produce the same information with significantly less data and only a marginal increase in uncertainty and use the resources in other assessments more effectively.

## Appendix 1: Literature provided for the review

- ACCEPTED PRACTICES GUIDELINES FOR GROUNDFISH STOCK ASSESSMENTS
- An update to developing a prior for the natural mortality rate for fishes for use 1 in the 2017 and 2019 assessment cycles for the U.S. West Coast
- NOAA guidelines on Information Quality
- Stock Synthesis User Manual Version 3.30.13
- TERMS OF REFERENCE FOR THE GROUNDFISH AND COASTAL PELAGIC SPECIES STOCK ASSESSMENT REVIEW PROCESS FOR 2019-2020
- Overview of West Coast Groundfish Fishery-Independent Surveys
- Methot Jr, R. D., & Wetzel, C. R. (2013). Stock synthesis: a biological and statistical framework for fish stock assessment and fishery management. Fisheries Research, 142, 86-99.
- **Status of Cowcod (*Sebastes levis*) in 2019**
- Status and Productivity of Cowcod, *Sebastes levis*, in the Southern California Bight, 2013
- Cowcod Stock Assessment Review (STAR) Panel Report 2013
- Bellquist, L., & Semmens, B. X. (2016). Temporal and spatial dynamics of 'trophy'-sized demersal fishes off the California (USA) coast, 1966 to 2013. Marine Ecology Progress Series, 547, 1-18.
- **The Combined Status of Gopher (*Sebastes carnatus*) and Black-and-Yellow Rockfishes (*Sebastes chrysomelas*) in U.S. Waters Off 3 California in 2019**
- STOCK ASSESSMENT of the GOPHER ROCKFISH (*Sebastes carnatus*) 2005
- GOPHER ROCKFISH STAR Panel Report 2005

# APPENDIX 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 fulfil 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.

<u>Pre-review Background Documents</u>: 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 panellists, 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

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

# 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**

# APPENDIX 3: Panel membership

## STAR 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 for Gopher and Black and Yellow Rockfish

Melissa Monk, National Marine Fisheries Service, Southwest Fisheries Science Center
Xi He, National Marine Fisheries Service, Southwest Fisheries Science Center

## Stock Assessment Team (STAT) Members for Cowcod

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