Report on

Chilipepper and Bocaccio rockfish STAR Panel

June 25-29, 2007

NOAA Fisheries Service

Southwest Fisheries Science Center

Fisheries Ecology Division

Santa Cruz

California

P.L. Cordue

Fisheries consultant

New Zealand

For **University of Miami Independent System for Peer Review**

13 July 2007

Executive summary

The bocaccio and chilipepper rockfish STAR Panel met at the Southwest Fisheries Science Center in Santa Cruz, California, from June 25-29, 2007. The Panel consisted of three reviewers (two CIE) and an SSC representative as the chair. This is the report of one reviewer and it should be read in conjunction with the other review report and the STAR Panel reports.

The documentation for the two assessments was in various stages of completion when they were distributed to meeting participants two weeks in advance of the meeting. In each case the STAT supplied altered documentation and/or additional documentation to the Panel at the start of the meeting.

The draft bocaccio rockfish assessment was, very surprisingly, an update of the 2005 update of the 2003 full assessment. Meeting participants agreed that the Panel should "pre-review" the assessment as an update and provide advice accordingly to the SSC (whose prerogative it is to review assessment updates). It was also agreed that some exploratory analyses should be done to provide guidance for the next full stock assessment. The Panel recommended that the assessment update be revised to include recent adjustments to CalCOM catch data and concluded that it appeared to meet the requirements for an update. The exploratory analyses suggested by the Panel were diligently performed by the STAT. However, they will be of little benefit to the next assessment author should they follow the recommendation of the Panel for a full review of all data inputs and assumptions before conducting a full assessment using modern stock assessment tools.

The draft chilipepper rockfish assessment assumed a single stock for chilipepper rockfish off California. A full catch history was reconstructed back to 1892 for four fisheries (trawl, hook and line, set net, and recreational). Five abundance indices were used: triennial bottom trawl survey, the NWFSC shelf/slope combination bottom trawl survey, CPUE from commercial trawl, CPUE from the northern California CPFV observer database, and the coast-wide SWFSC juvenile rockfish abundance survey. Age data were used from three fisheries: trawl, hook and line, and set net; and from the NWFSC combination survey. Length data were used from four fisheries: trawl, hook and line, set net, and recreational; and two surveys: triennial bottom trawl survey and the NWFSC combination survey.

The draft assessment document distributed prior to the STAR Panel meeting described a preliminary assessment model that included conditional age-at-length data rather than age frequencies. Problems with tuning this model resulted in the STAT bringing to the meeting a revised assessment model that had length frequencies and age frequencies (with no conditional age-at-length). The Panel agreed that the use of conditional age-at-length data would not be pursued during the meeting (but recommend its use in future assessments).

The Panel and STAT worked together to produce a technically acceptable assessment. After exploring several aspects of the data and performing numerous model runs, the STAT and Panel agreed to an assessment which used the original data sources but with some variations on the actual indices used and their relative weightings. To account for the dependence between some length and age data, length frequencies were down-weighted (lambda = 0.1) where associated age data were also present. The GLMM versions of the triennial and NWFSC combination trawl surveys were used (rather than area swept estimates). Also, the CPFV data were reanalyzed to produce a more consistent time series (data were used from only the "core' depths).

There were some tensions amongst the different data sources, as evidenced by likelihood profiles (of individual likelihood components) on natural mortality and biomass. The Panel recommended and the STAT agreed that the CPFV time series should be used as the "primary tuning" index; that is, efforts were made to properly fit this time series in preference to other input data. The base model adequately fits the downward trend in the CPFV series, but a combined age and length based selectivity was required for it to do so. This is of some concern because there is no direct evidence of such a selection mechanism.

The level of steepness in the stock recruit relationship was determined to provide a suitable dimension of uncertainty, providing good contrast in the estimated stock depletion.

The chilipepper rockfish assessment is at the lower end of acceptability for a modern stock assessment. This is not the fault of the STAT but a criticism of the process. Given the nature of the process there is inadequate time to explore the full range of uncertainty within given models (e.g., using MCMC) or across models (with a complete suite of sensitivity runs).

Background

The bocaccio and chilipepper rockfish STAR Panel met at the Southwest Fisheries Science Center in Santa Cruz, California, from June 25-29, 2007. This was the third of five 2007 STAR Panels in the biennial meeting schedule.

Two assessments were scheduled for presentation at the meeting; each species was assessed as a single west-coast stock. Each STAT consisted of a single scientist: Dr Alec McCall (bocaccio) and Dr John Field (chilipepper)

The STAR Panel had four members. My three colleagues were Dr David Sampson, the SSC representative and Panel Chair, Dr Kevin Piner, and Dr Norm Hall, my fellow CIE reviewer. This report should be read in conjunction with the STAR Panel reports and Dr Hall's CIE report.

Review Activities

Pre-meeting

Meeting documents and materials were received in electronic form in advance of the meeting (*see* Appendix 3). I familiarized myself with the background material and current assessments prior to the meeting. Paper copies of the assessment documents were also made available at the meeting, which was helpful.

I also inquired of the Chair as to why an update of the bocaccio assessment was being supplied to the meeting – rather than a full assessment. The Chair admitted to sharing my surprise and initiated some inquiries. Prior to the meeting, panelists were informed by the Chair that we would discuss how to proceed with regard to bocaccio as the first agenda item of the meeting.

Meeting

The meeting was convened at 12.30 pm on Monday, June 25, 2007 and closed Friday afternoon, June 29, 2007. I will only give a brief summary of the meetings activities, concentrating on bocaccio rockfish. For both species, details of the requests to the STATs and their responses are contained in the STAR Panel reports. I was not required to rapporteur at this meeting – with three Panelists (not the Chair) and only two assessments. Dr Piner took detailed notes for bocaccio rockfish, and Dr Hall covered chilipepper rockfish.

The first item on the revised agenda on Monday afternoon was to decide how to proceed with regard to the bocaccio assessment given that a simple update had been produced instead of a full assessment. Dr MacCall explained why he had only done an update (the

reasons are not relevant to the review and are not stated here). Meeting participants discussed how to proceed.

I expressed concern that management decisions would be based on an update of the previous full assessment in 2003. I was on the STAR Panel that reviewed bocaccio in 2003 (as was Dr Sampson) and that Panel had major concerns about the assessment. I stated that I thought that the 2003 assessment was at the lower end of acceptability in 2003, and with nothing but simple updates since then, it would necessarily fail to meet today's standards. The Chair pointed out that we didn't have a full assessment to review in any case – the STAT confirmed that they would not be providing the extra runs and documentation that would be needed for a full assessment – so there was no point reviewing the update with regard to the terms of reference for a full assessment.

It was eventually agreed that the Panel could "pre-review" the updated assessment and provide some advice to the SSC (who are charged with reviewing assessment updates). Further, the Panel and STAT agreed that some exploratory runs could be done with the objective of aiding the next assessment author.

The meeting then moved on to the chilipepper assessment presented by Dr Field. In the assessment document submitted to the Panel, prior to the meeting, conditional age-at-length data had been used in the model runs (which had not been fully tuned). Just prior to the meeting, further documents were distributed to the Panel by email with the results of tuned runs, and some additional runs not using age-at-length data. The tuned and un-tuned model results were quite different, and so much so that Dr Field indicated that he was no longer comfortable using conditional age-at-length data and suggested to the Panel that the more traditional approach (at least on the west coast) of using length and age frequencies was preferable.

The meeting closed for the day on Monday evening after the Panel had agreed some tasks for the chilipepper STAT to perform prior to Tuesday afternoon. On Tuesday morning the bocaccio STAT presented the assessment update. The Panel considered the update relative to the terms of reference for an update and suggested that the catch history be "refreshed" (using the latest CalCOM data) as was done with other recent assessment updates. Some tasks for the bocaccio STAT were agreed before we broke for lunch.

From Tuesday afternoon until the end of the meeting the Panel concentrated on aiding the chilipepper STAT to produce an acceptable base model and to identify a suitable dimension of uncertainty. There were some interruptions to consider the results of the bocaccio requests. Some sessions were also spent on drafting the STAR Panel reports (the bocaccio report was almost completed by the end of the meeting, but the chilipepper report – being much longer and more complex - was only in a very rough draft at the end of the meeting).

A suitable base model was identified for chilipepper on Thursday afternoon and on Friday a suitable dimension of uncertainty was determined. The meeting closed at 3.15 pm on Friday.

Post-meeting

The STAR Panel reports were left largely with the Chair for the purposes of completing drafts for the full Panel to review and revise by email. I provided the Chair with an Appendix for the chilipepper report which considers in some detail the appropriate use of age and length data (*see* Appendix 1). At the time of writing this report, full drafts of the STAR Panel reports have been circulated by the Chair, and they are in the final stages of revision by email.

Just before the STAR Panel meeting concluded, there was an incidental conversation through which I was alerted to the fact that the SSC had found a problem with a method used by the Newport STAR Panel. In their review of sablefish they noted a problem with the method used to obtain a prior on the trawl survey proportionality constant (q). At the Santa Cruz meeting, after a brief discussion, I took the problem to be with the calculation of the "low q" and "high q" (being respectively the mean of the lower and upper quartiles in the prior). I discovered, from subsequent email exchanges, that this was not the problem, but the real issue was the use of a lognormal distribution when forming a prior from a 99%-range and a "best guess". After some analysis, I reached the same conclusion as the SSC that the use of the lognormal was not generally appropriate. I went somewhat further by proposing an alternative distribution which I believe is generally appropriate (*see* Appendix 2).

I drafted my second set of "suggestions to STATs" for distribution by email (*see* Appendix 2). My suggestions included recommendations concerning the use of age and length data, and a discussion of the lognormal problem identified by the SSC, with my proposed alternative distribution.

Review findings

The findings for each assessment are discussed below under a variety of sub-headings. The "standard' headings are only appropriate for chilipepper rockfish (for which a full assessment was presented). For this stock, I summarize the draft assessment and the changes that were made to reach the final assessment. I then summarize the main uncertainties and finally summarize the merits and deficiencies of the accepted assessment.

For bocaccio rockfish, only an assessment update was presented and this was only reviewed relative to the terms of reference for an update. The merits, deficiencies, and uncertainties of the assessment would simply be a re-statement of those found by the STAR Panel in 2003 and are not presented here. For bocaccio, I briefly summarize the assessment and the few exploratory analyses that were conducted.

Bocaccio rockfish

Assessment summary

A single stock was assumed for bocaccio rockfish occurring in waters off the state of California (fish residing in U.S. waters south of Cape Mendocino).

A catch history was input back to 1951 with an assumed annual "historical catch" of 2000 t (i.e., the initial biomass is assumed to be less than virgin biomass and equal to the equilibrium biomass under a constant *F* which delivers an annual catch of 2000 t given constant recruitment). Five fisheries were modeled: three state-wide commercial fisheries and separate southern California and central/northern California recreational fisheries. Length data from each fishery and the triennial trawl survey were used with six abundance time series: trawl logbook CPUE, three recreational Cpues, triennial Survey abundance, and CalCOFI larval index of spawning output.

The assessment update consisted of updates of two model runs accepted by the 2003 STAR Panel and an additional run proposed by the STAT and accepted by the SSC in 2003. There was an extra run which used a fixed value of steepness which was the mean of a prior derived from meta-analysis.

The Panel considered the TOR for an update and concluded that the assessment meet the criteria. We were careful in our wording in the STAR Panel report to make it clear that we had not "reviewed" the update as such – since this was the role of the SSC. We did suggest that the recent catch history should be "refreshed" using the latest CalCOM data. This was done by the STAT and the revised results were, as expected, changed very little.

Exploratory analyses

The Panel suggested that a reference run be constructed against which to investigate various sensitivities. The new run with a fixed steepness was considered suitable except that it included contradictory indices: recreational CPUE and the triennial survey abundance indices. It was proposed by the STAT that much of the contradiction would be removed if the recent points in the CPUE time series were removed. There was also good reason to do this as the recent points were somewhat contrived having been derived under assumptions about the impact of bag limits (i.e., not actually based on observed catch rates, but catch rates that might have occurred if there hadn't been reduced bag limits in place). The proposed changes were made to the fixed steepness run to give a new reference run.

Sensitivity runs relative to the reference run included: an extended catch history with no annual historical catch; lower and higher annual historical catches; a change in selectivity for the recreational fisheries in recent years; and a logistic selection pattern for the triennial survey. Model results were confirmed to be sensitive to the assumed level of annual historical catch. The requested run with an extended catch history couldn't be

done properly because of array size limitations in the executable version of SS1 that was available.

Other exploratory analyses included: a comparison of total rockfish catches with the assumed historical catch of bocaccio – it appeared that the assumed level of 2000 t was perhaps 25% too high; an investigation of species changes over time in species composition in the northern and southern recreational fisheries – there appeared to be an inshore shift in the north and an off-shore shift in the south (this result was consistent with the selection patterns estimated by the model when recent selection was allowed to depart from earlier selection).

There had been some concern that estimates of good recent recruitment may have been driven by changes in fishery selectivity patterns. On the basis of the exploratory analyses and model runs the Panel concluded that there probably had been good recent recruitment – or at least, that similar results were obtained whether constant selectivities were assumed or not.

Chilipepper rockfish

Assessment summary

A single stock was assumed for chilipepper rockfish off California. A catch history extending back to 1892 was input for four fisheries (trawl, hook and line, set net, and recreational). Five abundance indices were used: triennial bottom trawl survey, the NWFSC shelf/slope combination bottom trawl survey, CPUE from commercial trawl, CPUE from the northern California CPFV observer database, and the coast-wide SWFSC juvenile rockfish abundance survey. Age data were used from three fisheries: trawl, hook and line, and set net; and from the NWFSC combination survey. Length data were used from four fisheries: trawl, hook and line, set net, and recreational; and two surveys: triennial bottom trawl survey and the NWFSC combination survey.

The draft assessment document distributed prior to the STAR Panel meeting described a preliminary assessment model that included conditional age-at-length data rather than age frequencies. Problems with tuning this model resulted in the STAT bringing to the meeting a revised assessment model that had length frequencies and age frequencies (with no conditional age-at-length).

The STAR Panel accepted that the conditional age-at-length approach should not be pursued during the meeting. Initially we were unsure whether the likelihood in SS2 was appropriate for age-at-length data obtained by sub-sampling a length frequency which was also used in the model. Later, it became clear that the likelihood was appropriate, but by then it was too late to pursue a conditional age-at-length model and it was also unclear how the associated age-at-length and length frequencies should be jointly tuned. I developed a recommendation for this after the STAR Panel meeting (see Appendix 2).

The Panel suggested that the "standard approach" (of 2007) be used for age and length frequencies where the age data were derived from sub-sampling a length frequency for age. Length frequencies associated with age frequencies were down-weighted in the likelihood by setting their lambda equal to 0.1. Also, the Panel suggested that the length frequencies of aged fish be compared with those of un-aged fish. This revealed a number of years in which the aged fish appeared to have been selected in a non-random fashion (typically being larger fish). In subsequent model runs suspect age frequencies were eliminated from the input data (the choice was subjective and was made by the STAT). Late in the meeting I realized that we had made a mistake by doing the comparison of length frequencies on the raw data – the comparison should have been made on properly scaled data (to allow for potential spatial and temporal variation in length). However, the approach should not have biased the results (a worse case scenario is perhaps only that some data that should have been used were not).

As the STAT and Panel explored alternative model configurations during the meeting it became apparent that there were some tensions between data sets within the model (as evidenced by likelihood profiles on M and later, and more informatively, on R_0). In particular, the fits to biomass indices were compromised "by the model" in order to better fit length frequency data. This often happens in these models as the contribution to the total log-likelihood from abundance indices can be over-whelmed by the relatively vast amount of length (and/or age) frequency data. As I put it to the meeting, we would perhaps need to choose an abundance time series on which to "hang our hat".

Chilipepper rockfish are known to be semi-pelagic, so there were concerns that the two available bottom trawl surveys may not provide reliable biomass indices. Of all the available indices, the CPFV time series, based on observed angler catch rates at defined fishing sites, seemed the most likely to provide a reliable abundance index. However, model fits to this time series were poor, with the predicted trend being flat while the observed indices showed a decline. The predicted trend also seemed inconsistent with the large 1984 cohort estimated by the model. After some experimentation it was realized that the length based selectivity for the CPFV time series was such that the 1984 cohort never entered the CPFV-selected biomass to any large extent. When an age-based selectivity was used, the predicted trend in the CPFV time series showed the expected increase when the 1984 cohort entered the selected biomass. However, the trend in the observed time series was not fitted well unless a very high emphasis was placed on the time series (i.e., a high lambda).

A base model was eventually found, but to achieve a reasonable fit to the CPFV time series, without an increased emphasis factor, a combined age and length based selectivity had to be used. Unfortunately there was no obvious rational for such a selection pattern. The base model also included time-varying growth, where the five blocks were derived from low-frequency changes in the Pacific Decadal Oscillation index (in the original assessment an arbitrary three-year blocking pattern had been used).

After exploring the accepted base model along several dimensions of uncertainty, including M and the level of the historical catch history, it was agreed that the major axis

of uncertainty should be the steepness parameter (*h*), which provided good contrast in the estimated level of stock depletion. Low and high values of *h* for a decision table were calculated as the means of the lower and upper quartiles of a normal prior probability distribution (which had been derived from a meta-analysis of rockfish steepness parameters).

Primary sources of uncertainty

Major uncertainties:

- It is not clear that the CPFV time series is a reliable index of abundance (the assessment was built on the assumption that it is).
- It is not clear that an appropriate selectivity has been estimated for the CPFV time series.
- The historical catch history is very uncertain.
- Natural mortality and steepness remain uncertain.
- Stock structure.

Strengths and weaknesses of current approach

The final assessment is at the lower end of what I consider to be acceptable in a modern stock assessment. There is a base model and two sensitivity runs, none of which were taken forward to MCMC runs to obtain posterior distributions and an appropriate measure of within-run uncertainty (approximate confidence intervals will no doubt be obtained for each run using boot-strapping or likelihood profiles but these are poor substitutes). This is a fault with the process rather than a criticism of the STAT.

Merits:

- SS2 was used and as such brings the advantages of a standard and well tested package.
- Discrimination and discernment were applied to the choice of data sets (as opposed to simply putting in everything and hoping for a sensible outcome).
- The use of environmental data to sensibly block time-varying growth was a useful and interesting innovation.

Deficiencies:

- All final runs used a composite length-age selection curve for the main tuning index (the CPFV survey), but currently there is no obvious rationale for such complex selection.
- The full uncertainty associated with the historical catch history has not been explored.

- The approach applied in this assessment of down-weighting length frequencies when associated age frequencies are included is ad hoc. Alternative approaches are suggested in Appendix 2.
- The tuning process used to equalize the "input" and "effective" sample sizes for length and age frequencies treated the age and length frequencies as independent even though length/age data for some fish were included in both length and age frequencies.
- "Good" length data may have been excluded from the model because the data filtering to detect biased age samples was based on unscaled length frequencies. (The fault of the Panel, not the STAT.)
- No MCMC runs were done.

Conclusions and Recommendations

The chilipepper rockfish assessment was technically improved by the STAR Panel process. The Panel's contribution with regard to bocaccio is more difficult to judge. I believe that the outcome was as good as could have been expected given the unprecedented circumstances.

I support the recommendations given in the STAR Panel reports many of which are repeated below.

Generic (all rockfish)

- Establish a *meta* database of all data relevant to rockfish stock assessment. The database should include enough detail about the nature and quality of the data that a stock assessment author can make a well informed decision on whether it could be useful for their stock assessment.
- Establish *accessible* online databases for all data relevant to rockfish stock assessment, so that assessment authors can obtain the *raw* data if required.
- Establish a database for historical rockfish catch histories, "best" guesses and estimates of uncertainty (and processes for updating and revising the database).
- Develop a concise set of documents that provide details of common data sources and methods used for analyzing the data to derive assessment model inputs.
- Develop fishery independent time series using fixed sites and volunteer fishers properly supervised using standard protocols
- Publish a full descriptive analysis of the recreational fisheries and fleets for CPUE interpretation (not limited to "rockfish trips" – interactions with other target species are important)
- Develop standard and validated methods for producing recreational CPUE indices which deal with the peculiarities of the recreational data and regulation changes. (The method of Stephens and MacCall for filtering recreational fishing trips is promising but remains largely unvalidated.)

Bocaccio rockfish

- The next assessment of bocaccio rockfish should be a full assessment and should use SS2 or some comparable modeling package.
- All the bocaccio rockfish data and assessment assumptions need a critical review and potential revision before being included in the next assessment.

Chilipepper rockfish

- Fully capture the uncertainty in historical catch. At least three alternative catch histories should be constructed: a “best guess”, an upper bound and a lower bound. Alternative assumptions in the timing of small and large catches could also be explored.
- Explore use of conditional age-at-length data rather than coupled age and length frequency data.
- Explore time-varying growth as influenced by environmental changes.
- Explore alternative selectivities for the CPFV time series.
- Explore possible spatial structuring of the data and model.
- Read chilipepper rockfish otoliths from the triennial and combination bottom trawl surveys to provide better data on the early stages of growth and possible time variations in growth
- Age validation for chilipepper rockfish should be pursued.

Appendix 1: Modeling of age and length data

The appropriate use of age and length samples in stock assessments is important in obtaining robust stock assessment results. In a likelihood setting, the key is the application of appropriate likelihoods given the nature of the data – which is dependent upon how it was collected.

Age frequencies and length frequencies for a given fishery or abundance survey may be obtained independently or in combination. The usual likelihood used for both is a multinomial with an "effective sample size" which is smaller than the actual number of fish measured or aged (for length frequencies, the effective sample size is often similar in magnitude to the number of samples taken rather than the number of fish measured).

When a length frequency is sub-sampled for age, it is not immediately clear how the dependence between the length frequency and the age data should be represented. Two approaches have been taken in rockfish assessments. The most common method is to use both the length and age frequency in the assessment but to down-weight the joint contribution of the data to the total likelihood by adjusting emphasis factors on the individual components (e.g., lambda = 0.1 for length samples where a sub-sampled age frequency is also present; or lambda = 0.5 for both the age and length frequency). An alternative, which is theoretically better, when both age and length are used, is to use the age data as conditional age-at-length.

The latter method requires the input of the proportions at age for given length (class). The same approach is used when there are independent age and length samples, but the age sample was obtained from non-random length samples (e.g., to obtain a growth curve). The age frequency is biased, but the conditional age-at-length data are not.

The distinction between the two situations is the issue of independence between the length frequency and the age-length data. When there is sub-sampling of a length frequency for age, the length data and the age sub-sample are clearly not independent. It follows, in comparison to independent samples, that there must be an additional likelihood component which "links" the two data sets. It is very instructive to derive the likelihood and see why this component is important but also why it does not contribute to the total likelihood when fish are selected at random for the age sub-sampling.

Suppose that n_l fish are sampled at random for length from a population (in a statistical sense). Further, suppose that n_a fish are then sub-sampled at random for age.

Assume that there are m length classes and let L_i denote the number of fish in the ith length class for the length sample. Let X_{ij} denote the number of fish in the ith length class and jth age class in the sub-sample for age. Adopting the notation of lowercase letters for observations of the random variables and bold notation to represent vectors or matrices, it follows from conditional probability theory that,

$$P(\boldsymbol{L} = \boldsymbol{l}, \boldsymbol{X} = \boldsymbol{x}) = P(\boldsymbol{L} = \boldsymbol{l})\, P(\boldsymbol{X} = \boldsymbol{x} \mid \boldsymbol{L} = \boldsymbol{l})$$

The likelihood for $\boldsymbol{L}$ is a multinomial:

$$P(\boldsymbol{L} = \boldsymbol{l}) = \text{Mult}(\boldsymbol{l} \mid n_l, \boldsymbol{p})$$

where $\boldsymbol{p}$ is the vector of proportions at length in the population.

The conditional likelihood is derived by applying a further conditional construction:

$$P(\boldsymbol{X} = \boldsymbol{x} \mid \boldsymbol{L} = \boldsymbol{l}) = P(\boldsymbol{U} = \boldsymbol{u} \mid \boldsymbol{L} = \boldsymbol{l})\, P(\boldsymbol{X} = \boldsymbol{x} \mid \boldsymbol{U} = \boldsymbol{u}, \boldsymbol{L} = \boldsymbol{l})$$

where U_i is the number of fish in the ith length class in the age sub-sample.

The conditional likelihood for $\boldsymbol{U}$ is another multinomial:

$$P(\boldsymbol{U} = \boldsymbol{u} \mid \boldsymbol{L} = \boldsymbol{l}) = \text{Mult}(\boldsymbol{u} \mid n_a, \boldsymbol{s})$$

where $s_i = l_i / n_l$ is the proportion of fish in the ith length bin in the length sample.

The final component in the joint likelihood is the conditional age-at-length likelihood:

$$P(\mathbf{X} = \mathbf{x} \mid \mathbf{U} = \mathbf{u}, \mathbf{L} = \mathbf{l}) = \prod_{i=1}^{m} \text{Mult}(\mathbf{x_i} \mid u_i, \mathbf{p_i})$$

where $\boldsymbol{p_i}$ is the vector of proportions at age in the population for the ith length class.

Hence, the joint likelihood of the length sample sub-sampled for age is the product of the likelihood for the length frequency, the conditional age-at-length, and the "linking" component being the sub-sample for length associated with the age sampling.

If the sub-sample of length is truly at random then the linking component consists entirely of "constants" (in terms of population parameters) and so does not need to be included for estimation purposes. Alternatively, if the sampling is biased, but the bias depends only on the characteristics of the length sample, then the linking component can still be ignored (even across a time series, despite the fact that the "constant" varies).

However, if sub-sampling for length is not random and depends upon population parameters then the linking component is potentially important. To adhere to a strict likelihood approach, it would be necessary to include the population parameters driving the bias in an appropriate parameterization to account for the biased selection process. When a time series of length and age data are used it is important to check for potential bias in the length sub-sampling and to consider if it could be driven by population parameters. If that could be occurring in some years, then the associated age data should perhaps be removed or the annual biases should be estimated using a joint likelihood which includes an appropriately parameterized bias function in the probability vector of the linking likelihood component.

Of course, one does not necessarily need to adhere to a strict likelihood approach. It can be argued that any bias in the sub-sampling for age can be ignored when the age data are used as conditional age-at-length. The argument being that the linking component may potentially provide information about population parameters, if the bias truly is driven by them, but by ignoring the component, potential information is forgone, but existing information in the other data is not compromised.

An important point emphasized by the full joint likelihood is the linkage of the length and age data in terms of their sample sizes. This is perhaps obvious in hindsight, but when "tuning" of age and length data is done during a stock assessment (i.e., an iterative adjustment of effective sample sizes to ensure that input variance assumptions are consistent with residual variances) it is crucial to maintain the consistency of the age and length sample sizes. That is, they must *not* be tuned independently. The relative contributions of each year's age and length data to the total log-likelihood of a full age and length time series will be proportionally maintained if effective sample sizes are scaled by the same multiplier both between and within years.

Appendix 2: Suggestions distributed to STATs (2)

The following document was distributed to STATs and other relevant people by Stacey Miller, on my behalf, by email on 11 July 2007.

Suggestions for STATs (2)

Patrick Cordue
9 July 2007

As the person with the "honor" of serving on all of the STAR Panels this year I thought it would be useful to provide some suggestions to STATs which are yet to present their assessments to STAR Panels. This is my second set of suggestions (see A-F in "Suggestions (1)"). As before, the following are merely suggestions – many special cases arise for individual assessments and STATs may have good reason for taking somewhat different approaches.

G. Use of age and length data

The simultaneous use of age and length frequencies where the age data is obtained from sub-sampling the length data is, of course, technically incorrect. The age and length frequencies are not independent. The ad hoc approach of down-weighting the length and/or age data to allow for the dependence is an option often used – but it is far from ideal.

I offer three alternative approaches:

- use the age data as conditional age-at-length;
- do not use the length data when "good" age data are available;
- do not use the age data when the length data are far superior.

Each of the above approaches is technically correct in avoiding the "double" use of data. Do not strive to use every piece of data. The approach of throwing everything in and hoping for the best is not at all appropriate in stock assessment. Discernment and judgment are the key – and the judicious use of sensitivity runs.

The use of conditional age-at-length data is quite robust to biases which can occur when length frequencies are sub-sampled for age. Of course, the use of conditional age-at-length data presupposes that no significant growth occurs (at the selected lengths) during the period over which the data were collected. If such growth does occur then the length frequency data (and conditional age-at-length data) are of little use unless they can be split into appropriate seasonal components (and fitted accordingly in SS2). The age data

used as an age frequency will still be useful (assuming that there are no otolith edge effects which might compromise the age readings).

When length data have been sub-sampled for age it is useful to compare the *properly scaled* length frequencies of the aged and non-aged fish to check for any obvious differences. If, in a given year, the length frequency of the aged fish is significantly different from that of the non-aged fish then the age data can only be used as conditional age-at-length. It must not be used as a normal age frequency or to provide mean length-at-age inputs.

A related issue is whether to estimate growth within the model or to do so externally. Technically, estimation within the model appears to be the best option. However, in practice this is often not the case. In many assessments it appears that enormous time and trouble can be spent trying to fit length frequencies in the model so that growth and selectivities can be estimated – and the trends in biomass indices get ignored (by the model). I believe that it is always best to have at least one run where growth is estimated externally and data inputs are restricted to biomass indices and age data alone. This may well just be a sensitivity – but it will provide a useful comparison of the effect of including length data and/or estimating growth internally.

H. Tuning of age and length data

If conditional age-at-length data are used (with the corresponding length data) then the *whole* data set must be tuned together. That is, effective sample sizes for the age and length data within and across years must be scaled by the same multiplier. SS2 is not currently set up to do this. You will instead get two scalars to multiply by – one for the length data and one for the age-at-length data. I suggest that you use the average. The key diagnostic for a properly tuned run is that the standard deviation of the Pearson residuals is near to 1 for both the age and length data.

I. Construction of priors

At the Newport STAR Panel meeting, priors were formed for the catchability coefficients (q) of two key trawl surveys (one for each of sablefish and longnosed skate). The method involved the construction of a range and a "best guess" using the different components of a trawl survey q (the lower bound of the range being the product of the lower bounds on each component; the best guess being the product of best guesses; similarly for the upper bound). The resulting "triplet": [*low, high, best*] was then used to construct a lognormal prior on the q: $Q \sim LN(\mu,\sigma)$ where $P(low < Q < high) = 0.99$ and median(Q) = *best*.

The SSC reviewed the Newport STAR Panel report and noted a problem with this method. Their concern was that the lognormal distribution was unsuitable for use with triplets where *best* was too far from the midpoint of the range (in logspace). This could have the consequence, for example, that no density was less than *low* and hence the lognormal distribution was entirely determined by *best* and *high*.

I investigated this issue and I agree that the lognormal distribution is not a good choice in such cases. As an alternative to the lognormal I suggest the use of a “double lognormal”. This is a natural extension of a double normal:

$$Q \sim DLN(\gamma, \sigma_l, \sigma_r) \quad \text{iff} \quad \ln(Q) \sim DN(\gamma, \sigma_l, \sigma_r)$$

where γ is the mode of the double normal, and σ_l, σ_r are the left and right hand standard deviations. The triplet is used in the same way as with a lognormal except now it is guaranteed that 0.5% will be in each tail. I have a standard set of R functions available for the double normal (ddnorm - density, qdnorm - quantile, pdnorm – cumulative density, rdnorm – random generation) should anyone require them.

Appendix 3: Bibliography of supplied material

I. Current Draft Stock Assessments

A. Status of Chilipepper rockfish off of the West Coast in 2007. John Field
B. Status of bocaccio off California in 2007. Alec D. MacCall

II. Background Materials

A. 2006 Workshop Summary Reports
1. A Summary Report from the NWFSC Bottom Trawl Survey Workshop held October 31 – November 2, 2006 in Seattle, Washington. NOAA Fisheries, NWFSC, FRAM Division.
2. A Summary Report from the WC Groundfish Data/Modeling Workshop held August 8-10, 2006 in Seattle, Washington. NOAA Fisheries, NWFSC, FRAM Division.
3. Report of the Groundfish Harvest Policy Evaluation Workshop, Southwest Fisheries Science Center, La Jolla, California. December 18-20, 2006. A Workshop Sponsored by the Scientific and Statistical Committee of the Pacific Fishery Management Council
4. Pre-Recruit Survey Workshop. September 13-15, 2006. Southwest Fisheries Science Center, Santa Cruz, California. A Summary Report Prepared by Jim Hastie NOAA Fisheries, Northwest Fisheries Science Center and Stephen Ralston, NOAA Fisheries, Southwest Fisheries Science Center.

B. Previous Stock Assessments and STAR Panel Reports
1. Status of chilipepper rockfish stock in 1998. Stephen Ralston, Donald E. Pearson and Julie A. Reynolds. 1998.
1a. STAR Panel report on the chilipepper rockfish (Sebastes goodei) assessment. 1998.
2. Status of bocaccio off California in 2003. Alec D. MacCall.
2a. STAR Panel report on bocaccio. 2003.
3. Status of bocaccio off California in 2005. Alec C. MacCall.
3a. STAR Panel report on bocaccio. 2005.

C. SS2 Model Related
1. SS2 Zip File – includes User's Manual, example files, and powerpoint presentations
2. R Software Zip File – Code developed by Ian Stewart to perform model diagnostics and plotting of SS2 output. This is not an official SS2 add-on and is not part of the NOAA toolbox. File contains User's Guide, example files as well as powerpoint presentations.

D. Terms of Reference (TORs) for the West Coast Groundfish Stock Assessment and Review Process for 2007-2008. The Scientific and Statistical Committee (SSC) of the Pacific Fishery Management Council. 2006.

E. GAO Report: Pacific Groundfish: Continued Efforts Needed to Improve Reliability of Stock Assessments. United States General Accounting Office, Report to Congressional Requesters. June 2004.

F. Coastwide Pre-Recruit Indices from SWFSC and PWCC/NWFSC Midwater Trawl Surveys (2001-2006). Stephen Ralston. April 6, 2007.

III. Meeting Materials

A. Draft Agenda
B. STAR Panel Meeting Location Information
C. Driving Directions to NMFS/SWFSC/FED
D. List of STAR Panel Participants

Appendix 4: Statement of work

Consulting Agreement between the University of Miami and Patrick Cordue

Statement of Work

June 6, 2007

General

The Stock Assessment Review (STAR) meeting is a formal, public, multiple-day meeting of stock assessment experts who serve as a peer-review panel for one or more stock assessments. External, independent review of West Coast groundfish stock assessments is an essential part of the STAR panel process that is designed to make timely use of new fishery and survey data, analyze and understand these data as completely as possible, provide opportunity for public comment, and assure the best available science is used to inform management decisions.

The stock assessments will report the status of the bocaccio and chilipepper rockfish resources off the west coast of the United States using age and/or size-structured stock assessment models. Specifically, the information includes a determination of the condition and status of the fishery resources relative to current definitions for overfished status, summaries of available data included in the models, and impacts of various management scenarios on the status of the stocks. The information is provided to the Pacific Fishery Management Council and NOAA's National Marine Fisheries Service to be used as the basis of their management decisions, which are subsequently approved and disseminated by the Secretary of Commerce through NOAA and NMFS.

The consultant will participate in the Stock Assessment and Review (STAR) Panel of the Pacific Fishery Management Council (PFMC) for the review of the bocaccio and chilipepper rockfish stock assessments. The consultant should have expertise in fish population dynamics with experience in the integrated analysis type of modeling approach, using age-and size-structured models, use of MCMC to develop confidence intervals, and use of Generalized Linear Models in stock assessment models.

The Pacific Fishery Management Council's Scientific and Statistical Committee requests that "all review panelists should be experienced stock assessment scientists, i.e., individuals who have done actual stock assessments using current methods. Panelists

should be knowledgeable about the specific modeling approaches being reviewed, which in most cases will be statistical age- and/or length-structured assessment models" (SSC's Terms of Reference for Stock Assessments and STAR Panel Process for 2007-2008)

Documents to be provided to the consultants prior to the STAR Panel meeting include:

- Current drafts of the bocaccio and chilipepper stock assessments;
- Most recent previous stock assessments and STAR panel reports for bocaccio and chilipepper rockfish;
- An electronic copy of the data, the parameters, and the model used for the assessments (if requested by reviewer);
- The Terms of Reference for the Stock Assessment and STAR Panel Process for 2007-2008;
- Summary reports from the West Coast Groundfish "Off-Year" stock assessment improvement workshops held in 2006;
- Stock Synthesis 2 (SS2) Documentation; and
- Additional supporting documents as available.

Specifics

Consultant's duties should not exceed a maximum total of 14 days: several days prior to the meeting for document review; the 5-day meeting; and several days following the meeting to complete the written report. The report is to be based on the consultant's findings, and no consensus report shall be accepted.

The consultant's tasks consist of the following:

1) Become familiar with the draft stock assessments and background materials.
2) Actively participate in the STAR Panel to be held in Santa Cruz, California from June 25-29, 2007. Participants are strongly encouraged to voice all comments during the STAR Panel so the assessment teams can address the comments during the Panel meeting.
3) Comment on the primary sources of uncertainty in the assessment.
4) Comment on the strengths and weaknesses of current approaches.
5) Recommend alternative model configurations or formulations as appropriate during the STAR panel.
6) Complete a final report after the completion of the STAR Panel meeting.
7) No later than July 13, 2007 submit a written report consisting of the findings, analysis, and conclusions (see Annex I for further details), addressed to the "University of Miami Independent System for Peer Review," and sent to Dr. David Die, via e-mail to ddie@rsmas.miami.edu, and to Mr. Manoj Shivlani, via e-mail to mshivlani@rsmas.miami.edu.

Submission and Acceptance of Reviewer's Report

The CIE shall provide via e-mail the final reports of the consultants in pdf format to Dr. Lisa L. Desfosse for review by NOAA Fisheries and approval by the COTR, Dr. Stephen K. Brown by July 27, 2007. The COTR shall notify the CIE via e-mail regarding acceptance of the report. Following the COTR's approval, the CIE shall provide the COTR with pdf versions of the final report.

ANNEX 1: Contents of Panelist Report

1. The report shall be prefaced with an executive summary of findings and/or recommendations.

2. The main body of the report shall consist of a background, description of review activities, summary of findings (including answers to the questions in this statement of work), and conclusions/recommendations.

3. The report shall also include as separate appendices the bibliography of all materials provided by the Center for Independent Experts and a copy of the statement of work.