# Report on

# Canary Rockfish and Arrowtooth Flounder STAR Panel

**July 30-August 3, 2007**

**NOAA Western Regional Center**

**7600 Sand Point Way, NE**

**Seattle**

**Washington**

**P.L. Cordue**

**Fisheries consultant**

**New Zealand**

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

**10 August 2007**

## Executive summary

The canary rockfish and arrowtooth flounder STAR Panel meet at the NOAA Western Regional Center, Seattle, from July 30-August 3, 2007. The Panel consisted of four 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 assessments accepted by the STAR Panel are technically adequate to provide management advice and, as such, represent the "best available science" on which to base management advice.

Draft assessments were distributed electronically to meeting participants two weeks in advance of the meeting. Although a revised document for arrowtooth flounder was distributed at the meeting, it only contained minor revisions. The arrowtooth flounder assessment was well documented. The documentation for canary rockfish was excellent and revealed some unfortunate details of the Triennial trawl survey time series that had escaped my notice until this meeting.

The canary rockfish assessment was conducted using SS2 and assumed a single west-coast stock, although the author noted that the choice of the Canadian border as the northern boundary was essentially for management convenience. The assessment data included: commercial and recreational catch history beginning in 1916; Triennial trawl survey abundance indices, age and length data, 1980-2004; NWFSC trawl survey abundance indices, age and length data, 2003-2006; and age and length data from the fisheries, 1968-2004.

Two issues dominated discussions on the canary rockfish assessment: steepness and the timing of the Triennial trawl surveys.

The value of steepness in the original base model was chosen to minimize the likelihood component of the Triennial survey biomass indices – which, coincidentally, was a (low) value which was similar to those used in previous assessments. However, the total likelihood suggested much higher values of steepness – which the STAT considered implausible for canary rockfish. The Panel suggested that the selective use of individual likelihood components to estimate steepness was untenable. The only defensible method available was to use the mean or median of the Dorn prior for canary rockfish steepness. This was done, but only after a new prior was constructed which excluded darkblotched rockfish (because the 2007 darkblotched assessment had a substantially different value of steepness than that from the 2005 assessment which was used in the original prior).

The estimates of depletion for canary rockfish are very sensitive to the value of steepness. The new Dorn prior was used to provide base and bracketing values for steepness. The low run is consistent with the 2005 base model with an estimate of depletion equal to 12%; the new base model estimate of depletion is 32%; and the estimate in the high run is

56%. The low and high runs were each assigned probabilities of 25%. This is reasonable given how little is know about steepness.

Surprisingly, the timing of the Triennial trawl surveys has not been constant from 1980 to 2004. The surveys fall into two blocks: mid July-mid September timing for 1980-1992; and June-mid August timing for 1995-2004. Within the second block there is a trend towards earlier start dates and finish dates with the 2004 survey being the earliest. For the canary assessment, in the absence of corrections for day-of-year effects, the Triennial time series was split into the two blocks and separate *q*s estimated for each block. This had no implications for the canary assessment as the Triennial abundance indices had little influence on assessment results. The assessment is inherently uncertain being driven by length and age data, and model assumptions.

A single stock was assessed for arrowtooth flounder off the west coast using SS2. As with canary rockfish, the choice of the Canadian border as the northern boundary was pragmatic rather than scientific. The assessment data included: a catch history beginning in 1940 (although the model started in 1916); Triennial trawl survey abundance indices and length data, 1980-2004; NWFSC slope/shelf survey indices, age and length data, 2003-2006; NWFSC slope survey indices 1998-2002; AFSC slope survey indices 1997, 1999-2001; and age and length data from the fillet fishery, 1986-2006.

The original base model estimated that the current stock size was well in excess of virgin levels (171% depletion). During the meeting it was discovered that this result was primarily driven by the start year in which recruitment deviations were estimated. When all deviations from 1916 were estimated there was an extended period of below average recruitment and a recent period of extremely strong recruitment. When deviations were estimated for a period where data on the cohorts was present in the model, then recent recruitment was estimated to be considerably lower, and estimated depletion was typically in the range 70%-90%.

Considerable time was spent trying to obtain sensible estimates for the selectivity of the fillet fleet. The original base model had a strong prior on the peak selection at 70 cm to prevent the estimate from reaching an upper bound. During the meeting, the problem was partially solved by estimating natural mortality while fixing peak selection at a “sensible” value (60 cm). Higher values of natural mortality were estimated using this approach, which brought them closer to values used for Gulf of Alaska arrowtooth flounder.

The timing of the Triennial survey was of no consequence for the arrowtooth flounder assessment. Like the canary rockfish assessment it is driven by length and age data and model assumptions. The assessment is inherently uncertain but the data are such that there appear to be no current conservation concerns. The Panel specified that model uncertainty be bracketed by a combination of catch history and natural mortality; but there was insufficient time for the STAT to present these runs to the meeting.

For both species it is doubtful that the spatial distribution of the stock stops at the Canadian border. Future assessments should assess plausible biological stocks (using US and Canadian data if necessary).

## Background

The canary rockfish and arrowtooth flounder STAR Panel meet at the NOAA Western Regional Center, Seattle, from July 30-August 3, 2007. This was the last of five 2007 STAR Panels in the biennial meeting schedule (although a "mop-up" meeting is scheduled for October).

Two assessments were scheduled for presentation at the meeting; each species was assessed as a single west-coast stock. The canary rockfish STAT consisted of a single scientist, Dr. Ian Stewart; and the arrowtooth flounder STAT consisted of two scientists, Dr. Issac Kaplan and Dr. Thomas Helser.

The STAR Panel had five members. My four colleagues were Dr Steve Ralston, the SSC representative and Panel Chair, Dr Dvora Hart, Dr. Jim Ianelli, and Dr Paul Medley, my fellow CIE reviewer. This report should be read in conjunction with the STAR Panel reports and Dr Medley's CIE report.

## Review Activities

### Pre-meeting

Meeting documents and materials were received in electronic form well in advance of the meeting (*see* Appendix 2). 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.

### Meeting

The meeting was convened at 12.30 pm on Monday, July 30, 2007 and closed Friday afternoon, August 3, 2007.

We began with brief round-table introductions and a short summary of the Terms of Reference for the meeting by the Chair. As I had at the July STAR Panel meeting, I suggested that the requirement to take detailed notes of the discussion during the meeting should be assigned to non-Panel members, so that Panel members could participate fully in discussions. Dr Hart noted that in the east-coast process the rapporteur roles were assigned outside of the review panel. There was general agreement that this was desirable and the GMT and GAP representatives agreed to rapporteur (as had happened at the July STAR Panel meeting). The task of drafting the formal requests to STATs (and their responses, subsequent main points of discussion, and our conclusions) were assigned to Panel members. I covered canary rockfish and Dr Ianelli dealt with arrowtooth flounder.

I will only give a brief summary of the meeting activities. For both species, details of the requests to the STATs and their responses are contained in the STAR Panel reports.

The assessment of canary rockfish was presented on Monday afternoon by Dr. Stewart. He made an excellent presentation but it proceeded slowly as the Panel had many questions. The task of drafting the first set of requests was complicated by the warning that we had to leave the campus before 6 pm (as the gates would be locked). A resident Panel member suggested that we actually had until 7 pm. This allowed us the time needed to carefully draft the first set of requests for the STAT and the meeting adjourned for the day at some time after 6 pm.

On Tuesday morning the arrowtooth STAT were scheduled to make their first presentations. Dr Helser gave a presentation on the GLMM methods used to construct abundance indices from trawl survey data. He explained why the GLMM biomass estimates were sometimes (often?) much smaller than the usual "area-swept" design based estimates (median vs mean assumptions, and the occurrence of infrequent but very high catch rates). Dr. Kaplan then presented the arrowtooth assessment. As with canary rockfish, the issues that arose during the presentation lead to a set of requests for the arrowtooth STAT (drafted before lunch).

From Tuesday afternoon until Friday morning the meeting proceeded with alternate sessions with the two STATs as the Panel reviewed the responses to previous requests and submitted additional requests.

The base model for canary rockfish was essentially agreed by Thursday afternoon (though the responses to some requests were examined on Friday morning). The base model for arrowtooth was not decided until Friday afternoon because of difficulties getting a satisfactory fit to the fillet fishery length frequencies. They could be fitted easily enough but only with a completely unrealistic selectivity curve (with full selection at 80 cm – larger than almost all arrowtooth grow). The arrowtooth STAT was left with a specification for the bracketing runs – these were never presented to the Panel.

The meeting concluded a bit before 4 pm so that the Chair could be sure to catch his flight. First drafts of the reports were not completed before the meeting concluded but the main points of each section were discussed and agreed by the full Panel.

## Post-meeting

I completed my section for the canary rockfish report on requests and responses and forwarded it to the Chair and Panel members. Dr. Ianelli did the same for arrowtooth flounder. The STAR Panel reports were then left with the Chair for the purposes of completing drafts for the Panel to review and revise by email. At the time of writing this report, drafts of the STAR Panel reports have not been circulated. However, I do not anticipate any major problems with the completion of the reports.

During the week of the meeting, Dr. Ralston asked if I would be able to contribute some ideas for "off-year" workshops (given there is an SSC meeting scheduled to plan the 2008 workshops). Shortly after the meeting I drafted a document with some suggested topics for workshops based on my experience over the five STAR Panels (*see* Appendix 1).

## Review findings

The findings for each assessment are discussed below. For each 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.

### Canary rockfish

#### *Assessment summary*

The canary rockfish assessment was conducted using SS2 and assumed a single west-coast stock, although it was noted in the documentation that the choice of the Canadian border as the northern boundary was essentially for management convenience. The assessment data included: commercial and recreational catch history beginning in 1916; Triennial trawl survey abundance indices, age and length data, 1980-2004; NWFSC trawl survey abundance indices, age and length data, 2003-2006; and age and length data from the fisheries, 1968-2004.

Two issues dominated discussions on the canary rockfish assessment: steepness and the timing of the Triennial trawl surveys.

The value of steepness in the original base model was that which minimized the likelihood component of the Triennial survey biomass indices – which, coincidentally, was a (low) value which was similar to those used in previous assessments. However, the total likelihood suggested much higher values of steepness – which the STAT considered implausible for canary rockfish. The Panel suggested that the selective use of individual likelihood components to estimate steepness was not tenable (nor, in my opinion, is the use of total likelihood with these data). The only defensible method available was to use the mean or median of the Dorn prior for canary rockfish steepness (priors on steepness were developed prior to the 2007 assessments using the results of the 2005 assessments).

The STAT was uncomfortable with this suggestion as he was aware that darkblotched rockfish had been included in the prior, with a high value of steepness, and that a much lower value of steepness had been used for darkblotched in the 2007 assessment. He suggested that a new prior be developed for canary which excluded darkblotched, or contained the updated darkblotched results.

A request was made to Dr. Dorn who very promptly developed a new prior on steepness for canary rockfish. This excluded darkblotched as the required input (of a particular steepness profile) was not yet available from the 2007 assessment. The original prior had a mean of 0.6; the new prior had a mean of 0.52. The prior was used to supply a base model value (mean of the middle 50%) and two bracketing values (mean of the lower and upper 25% tails). The estimates of depletion are very sensitive to the assumed value of steepness: low = 12%, base = 32%, high = 56%.

The issue of the timing of the Triennial trawl surveys is relevant for many species but often may be of little consequence, as it was for canary rockfish. The surveys fall into two blocks: mid July-mid September timing for 1980-1992; and June-mid August timing for 1995-2004. Within the second block there is a trend towards earlier start dates and finish dates with the 2004 survey being the earliest. The issue is one of availability. A trend in the timing of the surveys may, for some species, introduce a trend in availability, and hence biomass trends will be confounded. Simple GLM models were applied to the canary and arrowtooth data during the meeting (by the Panel Chair). These were not presented to the full meeting but did find significant monthly effects. The arrowtooth STAT also performed some GLMM analyses with seasonal effects included – but these, also, were not presented to the full meeting (due to lack of time).

For the canary assessment, in the absence of corrections for monthly or day-of-year effects, the Triennial abundance indices were split into two blocks (1980-1992, 1995-2004) and separate *q*s were estimated for each block. This had no implications for the canary assessment as the Triennial abundance indices had little influence on assessment results (both before and after the split).

Other issues that arose for the canary assessment were: in which year to start estimating recruitment deviations; how to best estimate recruitment variability (sigmaR, tune or not tune); and how many selectivity blocks to estimate for the major fisheries. Of these, only the estimation of sigmaR was of any consequence (with higher sigmaR giving lower estimated depletion). The canary length and age data do not reveal any consistent cohorts that can followed “by eye”. Perhaps because of this lack of a consistent cohort signal recruitment variability is estimated to be very low (0.3 under iterative tuning). In reality, recruitment variability may be much higher. In the final assessment sigmaR was set equal to 0.5 (being not too different from values used for other rockfish).

The canary assessment is inherently uncertain being driven by length and age data, and model assumptions. Steepness was chosen as the primary dimension of uncertainty, and estimated depletion ranges from 12% to 56% (base = 32%). This is a very reasonable representation of the true uncertainty. Little is know for sure.

### *Primary sources of uncertainty*

The major sources of uncertainty in the canary rockfish assessment are assumed stock structure and steepness. The assessment is inherently uncertain being driven by length and age data, and model assumptions.

### *Strengths and weaknesses of current approach*

The current assessment is adequate in terms of capturing the nature of the assessment uncertainty. A better assessment could be produced using full Bayesian methods but that is not currently possible with the current process and technology.

Merits:

- SS2 was used and as such brings the advantages of a standard and well tested package.
- Steepness was correctly identified as a dominant dimension of uncertainty and a "risk neutral" approach was taken to determining the value of steepness used in the base model.
- The use of conditional age-at-length data is technically superior to the common practice of using dependent length and age frequencies (i.e., where the length data have been sub-sampled for age).
- The procedure used to specify initial multinomial effective sample sizes for age and length data was a useful attempt at standardization.

Deficiencies:

- The uncertainty associated with the historical catch history was not fully explored.
- An ad hoc approach was used for weighting of dependent fisheries age and length data (i.e., lambda = 0.5 for associated age and length data).
- The procedure used to specify initial multinomial effective sample sizes for age and length data has an unsound theoretical basis.
- No consideration was given to the problem of violating the assumption of constant proportions of age-at-length in conditional age-at-length data (E.g., when fish are growing during the sampling period.)
- No MCMC runs were done.

## Arrowtooth flounder

### *Assessment summary*

A single stock was assessed for arrowtooth flounder off the west coast using SS2. As with canary rockfish, the choice of the Canadian border as the northern boundary was pragmatic rather than scientific. The assessment data included: a catch history beginning in 1940 (although the model started in 1916); Triennial trawl survey abundance indices and length data, 1980-2004; NWFSC slope/shelf survey indices, age and length data, 2003-2006; NWFSC slope survey indices 1998-2002; AFSC slope survey indices 1997, 1999-2001; and age and length data from the fillet fishery, 1986-2006.

The original base model estimated that the current stock size was well in excess of virgin levels (171% depletion). During the meeting it was discovered that this result was primarily driven by the start year in which recruitment deviations were estimated. When all deviations from 1916 were estimated there was an extended period of below average recruitment and a recent period of extremely strong recruitment. When deviations were estimated for a period where data on the cohorts was present in the model, then recent recruitment was estimated to be considerably lower, and estimated depletion was typically in the range 70%-90%. The latter approach was adopted in the final base model.

Considerable time was spent trying to obtain sensible estimates for the selectivity of the fillet fleet. The original base model had a strong prior on the peak selection at 70 cm as when it was unconstrained it hit an upper bound. The model was able to fit the fillet fleet length frequencies adequately, but only with full selection at very high values (larger than most arrowtooth). Selectivity had been modeled as a function of just length. There was a large imbalance in the sex ratio in the fillet fishery landings so it was suggested that sex-specific selectivity would be appropriate. However, this alone was not adequate to estimate a sensible full selection size.

The problem was eventually "solved" by fixing peak selection at a "sensible" value (60 cm) and estimating natural mortality. This approach gave higher estimated values of natural mortality than in the original base model, which brought them closer to values used for Gulf of Alaska arrowtooth flounder.

There was much discussion about how best to calculate the catch history for the "by-catch fleet" – in essence being the discard history. In the original base model the by-catch fleet catch history began in 1956 and was equated to 13% of the coast-wide landed English, petrale, and Dover sole. Alternative catch scenarios were constructed for the by-catch fleet to give low and high alternatives. This was a reasonable attempt to investigate sensitivity of assessment results to assumed catch history, but it was not nearly exhaustive and rigorous enough. The discards from the very large catches of rockfish in the 1940s had not been considered, nor had other specific references in the literature to flatfish landings in the 1940s. The use of a constant proportion of landings of sole is of course dubious and other approaches need to be explored.

The timing of the Triennial survey was of no consequence for the arrowtooth flounder assessment. Like the canary rockfish assessment it is driven by length and age data and model assumptions. The assessment is inherently uncertain but the data are such that there appear to be no current conservation concerns. The Panel specified that model uncertainty be bracketed by a combination of catch history and natural mortality; but there was insufficient time for the STAT to present these runs to the meeting.

### *Primary sources of uncertainty*

The major sources of uncertainty in the arrowtooth flounder assessment are assumed stock structure and catch history. The assessment is inherently uncertain being driven by length and age data, and model assumptions.

*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 will be 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.
- The use of conditional age-at-length data is technically superior to the common practice of using dependent length and age frequencies (i.e., where the length data have been sub-sampled for age).

Deficiencies:

- The uncertainty associated with the historical catch history was not fully explored.
- Constant proportions of age-at-length were assumed for the NWFSC slope/shelf survey despite substantial growth in the smaller fish during the sampling period.
- The procedure used to specify initial multinomial effective sample sizes for age and length data has an unsound theoretical basis.
- It was not clear that the runs had properly converged.
- No MCMC runs were done.

## Conclusions and Recommendations

The canary rockfish and arrowtooth flounder assessments were technically improved, and substantially altered by the STAR Panel process.

The canary rockfish assessment brought to the meeting used a fixed value of steepness that was similar to values used in previous assessments. However, there was no reasonable basis for using such a level of steepness. A better approach to choosing steepness was adopted in the final assessment. Steepness was chosen as the axis of uncertainty as the assessment results are very sensitive to the choice made. Although improved, the assessment is unsatisfactory in two respects. First, it is doubtful that the spatial distribution of the “biological stock” is actually truncated at the Canadian border. Second, age and length data rather than biomass indices are driving the assessment results. This suggests that the results are very dependent on model assumptions.

The arrowtooth flounder assessment represents a good first modern assessment for this species. It appeared that there were no conservation issues with regard to the species and so it was held to a lesser standard than the canary rockfish assessment. The final assessment was judged acceptable for management purposes but it is not ideal. As with

canary rockfish, the spatial distribution of the stock almost certainly extends beyond the Canadian border. Also, the catch history is very uncertain, and the assessment results are largely independent of the biomass indices. There is also the minor but troubling technical issue of whether the runs actually converged.

I support the recommendations given in the STAR Panel reports many of which are repeated below. I have made similar recommendations in my other 2007 reports. Below, they are generalized to "groundfish" rather than being specific to "rockfish".

## Generic (all groundfish)

- Establish a *meta* database of all data relevant to groundfish 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 groundfish stock assessment, so that assessment authors can obtain the *raw* data if required.
- Establish a database for historical groundfish catch histories, "best" guesses and estimates of uncertainty (and processes for updating and revising the database). There must be a coordinated and comprehensive approach to developing this database (it must *not* be a compilation of individually constructed catch histories.)
- 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.
- For "hook and line species" which are not suitably sampled by trawl surveys, 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 "groundfish 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.)
- Develop standard and appropriate methods for modeling age and length data, including choice of distribution, initial variance assumptions, and tuning methods (current methods can and should be improved).
- Routinely produce and present supporting documentation for any derived indices or inputs which are included in a stock assessment model (e.g., GLMM derived trawl survey abundance indices; informed priors on steepness).

## Canary rockfish

This should be aimed towards a fully Bayesian assessment when the process and technology allow. More importantly, a biological stock needs to be assessed. It seems very unlikely that the stock stops at the Canadian border (given the high catch rates near the border).

- The next assessment should include a full stock structure review and assess a plausible biological stock.
- It seems likely that Canadian data will need to be incorporated in the next assessment (perhaps not a joint assessment with Canada, but at least the catch history for the portion of the stock found to be in Canada should be included).
- Carefully consider the most appropriate use of the available data:
    - Not all data need be used.
    - Not all age data need be included as conditional age-at-length (omit length frequency data when the corresponding age data are adequate)
    - All trawl survey abundance indices and associated length and age data should be carefully evaluated before inclusion. If catches are highly variable it may be because the trawl survey is not sampling representatively from the population (e.g., due to the bulk of the fish being on non-trawlable ground).
- 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.

## Arrowtooth flounder

As with canary rockfish this should be aimed towards a fully Bayesian assessment when the process and technology allow and a biological stock should be assessed (the stock does not stop at the Canadian border).

- The next assessment should include a full stock structure review and assess a plausible biological stock.
- It seems likely that Canadian data will need to be incorporated in the next assessment (perhaps not a joint assessment with Canada, but at least the catch history for the portion of the stock found to be in Canada should be included).
- 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.

## Appendix 1: Suggestions for “off-year” workshops

During the STAR Panel meeting, Steve Ralston asked if I could contribute some ideas for the “off-year” workshops. The following document was emailed to Martin Dorn (Chair of SSC groundfish sub-committee) and Steve Ralston on 5 August 2007.

**Suggestions for generic topics for “off-year” workshops**

P.L. Cordue
5 August 2007

Given there is a scheduled meeting to discuss workshop topics for 2008, Steve Ralston asked if I would make some notes on issues which I thought could be considered for the workshops. There have certainly been a number of important generic issues which have arisen during the 2007 STAR Panel meetings – which do need to be addressed.

I am not convinced that previous workshops have been as effective as they might have been in addressing generic issues. I know that some good work was presented at the 2006 workshops, but I get the impression that there was less than a fully coordinated approach taken to solving identified problems.

I see two potential extremes in the process that could be used for the 2008 workshops.

One extreme goes like this: there is an announcement to potential participants that there will be a workshop on such and such a theme; a date and venue are specified and people let the organizer know if they want to present something; everyone then gets together for the day, there are presentations, a general discussion, and some conclusions and recommendations are made and written up in a report.

At the other end of the spectrum: specific topics are identified for a workshop (with an identified theme); projects are defined, in each case, with a detailed specification of the problem that needs to investigated/solved; researchers with the requisite skills are identified and contracted to work on the projects; the researchers present their results at the workshop; there is a general discussion, and some conclusions and recommendations are made and written up in a report.

I believe that the latter approach is preferable to the former. I suspect that the 2006 approach was perhaps closer to the former than the latter.

Below I list some issues, under general headings, which I think could be usefully addressed by some funded projects – the results of which could be discussed at workshops. Alternatively, perhaps a workshop is needed to discuss research priorities and make recommendations on projects to be funded. I am not familiar with your research planning procedures so it is difficult for me to judge. I am well aware, that several issues

have been identified many times and the same recommendations have been made by STAR Panels, year after year.

*Data accessibility and catch histories*

It is somewhat inefficient for assessment authors to rely on the composition of STAR Panels to inform them of relevant data sources for their assessment. By the time the STAR Panel has convened it is often too late to obtain relevant data, let alone to include it in the assessment.

- Establish a *meta* database of all data relevant to groundfish 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 groundfish stock assessment, so that assessment authors can obtain the *raw* data if required.
- Establish a database for historical groundfish catch histories, "best" guesses and estimates of uncertainty (and processes for updating and revising the database). There must be a coordinated and comprehensive approach to developing this database (it must *not* be a compilation of individually constructed catch histories.)

*Abundance indices*

With many fisheries under severe regulation it is difficult or impossible to monitor abundance using fishery data. Fishery independent abundance indices are needed. A number of trawl survey indices are developing but there are also a number of important species which are poorly surveyed by trawl. Other methods are needed for these species.

- Consider all species and stocks which need to be monitored.
- Identify which species are adequately monitored by current time series and which are not.
- Identify suitable methods for species which are not adequately monitored.
- Develop a prioritized schedule for conducting the required surveys (development of new time series or continuation of existing time series).

*Triennial time series*

The Triennial trawl survey has had a shift in timing. The surveys fall into two blocks: mid July-mid September timing for 1980-1992; and June-mid August timing for 1995-2004. Within the second block there is a trend towards earlier start dates and finish dates with the 2004 survey being the earliest. The 2004 survey is also notable for many species showing very large increases from 2001. Further, for some species the Triennial survey is unlikely to adequately sample the population. These species need to be identified. It is

unacceptable to throw everything into the stock assessment model and hope that something sensible will emerge. Discernment is needed.

Conduct a comprehensive multi-species study of the Triennial trawl survey results:

- check for years with unusual "catchability" (i.e., do "too many species" show a marked increase or decrease in abundance in some years – look for indicator species which are less likely to have been affected by fishing)
- identify species for which the survey cannot be expected to provide abundance indices (those with higher densities on non-trawlable ground; those that are "too" semi-demersal; those which have highly variable catch rates)
- check for day-of-year effects for species for which abundance indices are defensible (e.g., perform a GLM on the Triennial survey data; GLM on NWFSC survey data; examine seasonal CPUE in fisheries data)
- if necessary incorporate day-of-year effects into the GLMM analysis used to produce abundance indices
- consider approaches to using the abundance indices from the Triennial survey in stock assessment (e.g., seasonally corrected or splitting the time series into two blocks).

*Development of informed priors*

Ideally, an informed prior should be developed for the proportionality constant ($q$ or "catchability") associated with each abundance time series used in a stock assessment model. This is often done for fishery independent surveys but can, in theory, also be done for CPUE indices which retain some measure of units. Even if a prior is not used in the estimation model, it is a necessary to have it before the estimated value of $q$ can be used as a legitimate diagnostic. Many times I have heard people say "that value of $q$ is just not plausible". They clearly have in their mind an "informed prior", but it may be very uninformed in that they do not have a clear understanding of all of the factors that affect a particular $q$. The correct equations need to be used in the development of informed priors for survey $q$s. Ancillary data needs to be made available to help bound some components. Expert opinion will also be needed. Groups of related species are best done together (as they will share ancillary data sources and experts will have opinions on the relative values of their components).

I suggest that trawl surveys for groundfish be tackled first:

- identify defensible trawl survey abundance time series for a range of species (and stocks)
- identify the appropriate equations for trawl survey $q$s for each stock (e.g., proportion of non-trawlable ground will matter for some species and not others; as will their relative densities on trawlable and non-trawlable ground)
- identify, collate, and analyze relevant sources of ancillary data on the parameters within the equations
- identify *small* groups of experts to develop ranges and "best guesses" for each parameter (and hence to priors for each trawl survey $q$)

*Recreational CPUE indices*

For some important recreational species, there may be little choice but to use CPUE indices despite the imposition of regulations. However, it is crucial to have the full context within which to interpret and analyze CPUE indices. For many species, the same type of data is available and the same regulations have been implemented. Therefore, it would be efficient to do a comprehensive study over the whole recreational sector.

- Conduct and publish a full descriptive analysis of the recreational fisheries and fleets for CPUE interpretation (not limited to "groundfish 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.)
- Specifically consider the use of random variables as explanatory variables. These have been used as proxies for habitat, but they introduce the "errors within variables" problem, and potentially may remove valid biomass signals from the response variables.
- Specifically consider the use of combined models (binomial model combined with a positive catch rate model) and whether they are robust to non-biomass factors that could drive the occurrence of zeroes.

*Stock assessment modeling issues*

**Use of age and length data**

The whole issue of how best to use age and length data in a stock assessment has not been resolved. The over-riding consideration for addressing these issues is whether the approach leads to a "better" stock assessment or not. Often, assessment authors appear to strive for greater reality through greater complexity and the inclusion of each and every data source that could conceivably be relevant. More data and more complexity do not necessarily mean a "better" assessment. There is much work that could be done looking at the following questions:

- What are the appropriate statistical distributions to use when modeling length and age data? (Properties of the data must be examined analytically and/or through bootstrapping.)
- If multinomial distributions are appropriate, how should effective sample sizes be determined (the existing equations of Stewart and Miller are not based on the observation error inherent in the data – rather on modeling choices and assumptions made in the 2005 stock assessments – again, analytical and/or bootstrap methods are needed).

- How should non-independent age and length data be jointly tuned? (E.g., when an age sub-sample of a length frequency is included as conditional age-at-length data, together with the length frequency.)
- Is it always best to estimate growth within the model? If so, how much conditional age-at-length data is desirable?
- How much violation of the assumption of constant proportions of age-at-length is allowable in conditional age-at-length data, before seasonal growth should be modeled? (E.g., when fish are growing during the sampling period.)

**Estimation of $R_0$, recruitment deviations, sigmaR, natural mortality, and steepness**
It is not clear how best to determine which year to start estimating recruitment deviations. Nor is it clear how best to estimate sigmaR (should sigmaR be tuned or not?). Estimation of steepness is also a thorny issue, as is the imposition of a stock recruitment relationship. Natural mortality is of course another problem.

There are least three general ways to configure a "forward projection statistical stock assessment model". An integrated model with a fully specified catch history and internally consistent relationship between $R_0$, recruitments, stock recruitment relationship, and $B_0$ can be configured with or without a penalty forcing recruitment deviations to follow the stock recruitment curve (in the latter case, recruitments are simply estimated to best fit the data and the stock-recruitment relationship is an output of the results). A third alternative is to start the model in a non-equilibrium state when data first become informative (and hence a full catch history is not needed).

It would be useful if some guidance was available on when different configurations were preferable – in terms of the conditions under which each method delivers the most "reliable" estimators. Some help will be available in the literature but there are no definitive studies. Retrospective analysis and bootstrapping methods are *not* adequate to investigate these questions. Nothing short of a full simulation study with a "complex" operating model and alternative (simpler) estimation models will do. A number of generic stock assessments will need to be simulated over a multi-dimensional operating model space (e.g., different true values of $R_0$, steepness, natural mortality, sigmaR, etc) to investigate the relative performance of the alternative estimators (in terms of accuracy – not just bias) and their robustness to violation of estimation model assumptions.

## Appendix 2: Bibliography of supplied material

### I. Meeting Materials

A. Terms of Reference for Stock Assessments and STAR Panels
B. Draft Agenda
C. Meeting Location Information
D. Driving Directions & Map
E. List of Panel Participants

### II. Previous Assessments and STAR Panel Reports

A. Status of the U.S. canary rockfish resource in 2005. Richard D. Methot and Ian J. Stewart.
B. STAR Panel Report for Canary Rockfish. August 15-19, 2005
C. Follow-Up STAR Panel Report for Canary Rockfish. September 26-30, 2005
D. Status of the Coastal Arrowtooth Flounder Resource in 1993. Martha H. Rickey. October, 1993.

### III. 2006 "Off-Year" Workshop Reports

A. A Summary Report from the NWFSC Bottom Trawl Survey Workshop held October 31 – November 2, 2006 in Seattle, Washington. NOAA Fisheries, NWFSC, FRAM Division.
B. A Summary Report from the WC Groundfish Data/Modeling Workshop held August 8-10, 2006 in Seattle, Washington. NOAA Fisheries, NWFSC, FRAM Division.
C. 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
D. 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.

### IV. SS2 Documentation

A. SS2 Zip File – includes User's Manual, example files, and powerpoint presentations
B. 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.

## V. Additional Background Materials

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

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

## Appendix 3: Statement of work

**Statement of Work**

**July 30 – August 3, 2007 Canary Rockfish and Arrowtooth Flounder Stock Assessment Review (STAR) Panel**

### 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 canary rockfish and arrowtooth flounder 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 canary rockfish and arrowtooth flounder 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 draft of the canary rockfish and arrowtooth flounder stock assessment;
- Most recent previous stock assessment and STAR panel report for canary rockfish (this is the first assessment of arrowtooth flounder since 1993);
- An electronic copy of the data, the parameters, and the model used for the assessment (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 assessment and background materials.
2) Actively participate in the STAR Panel to be held in Seattle, Washington, July 30 - August 3, 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 August 17, 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 August 31, 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.