

**Report on
2007 STAR Panel meetings
West Coast, USA**

**P.L. Cordue
Fisheries consultant
New Zealand**

For University of Miami Independent System for Peer Review

7 December 2007

Executive summary

This is a report on the STAR Panel meetings held on the West Coast during 2007 from my perspective as the CIE reviewer who attended all of the meetings. This is an “overview” report which considers the following: the approaches and findings of STAR Panels that may or may not be consistent with each other or previous Panels; the general reporting of uncertainty in the stock assessments; and the general strengths and weaknesses of current approaches along the West Coast. The last aspect is considered very much within the context of the STAR Panel process rather than being restricted to the “assessment approaches”.

Six STAR Panel meetings were held on the west coast during 2007, five regular STAR Panels, during May-August, and then a “mop up” meeting, in October, held as part of a groundfish sub-committee meeting of the SSC. Each meeting was scheduled for four and half days (beginning Monday lunchtime, ending Friday evening) and involved three, four, or five reviewers including two CIE reviewers and an SSC member as the Chair (regular meetings). I attended all meetings including the “mop up” (as the only CIE reviewer).

Most meetings began with the Chair reminding the Panel that the meeting was a “review” and not a “workshop”. STAR Panel members were then assigned as rapporteurs for each assessment being reviewed. Their important role was that of drafting the formal requests to the STAT for exploratory analyses or model runs. Each STAT would then present their draft assessment and during the week the Panel and STATs would collaboratively work towards technically acceptable assessments.

As the “continuity” CIE reviewer on the 2007 STAR Panels, where possible, I encouraged each Panel to adopt similar approaches – to achieve some level of “consistency” across assessments. However, different issues arose at each meeting; different STATs and Panelists tended to emphasize different aspects of assessments or be concerned about different technical issues. Also, my experience over the meetings was cumulative rather than “simultaneous”. I took lessons from earlier Panels and applied them to the later ones – the converse was not possible.

There was, of course, much greater consistency across assessments within each meeting than between meetings. Assessments considered together, by the same panel, naturally had the same technical issues considered. Across meetings, this was much more difficult to do – because of a change in personnel and assessments.

A comparison of the values of natural mortality, steepness, and recruitment variability between the 2007 assessments and the most recent earlier corresponding assessments showed a great deal of consistency for natural mortality and recruitment variability. There were some marked changes in steepness, but this was generally due to use of new data.

The reporting of uncertainty in 2007 followed the usual practice of concentrating on model uncertainty through the presentation of a base model together with “low” and “high” runs. Catch history uncertainty was a common component in many two-dimensional representations of uncertainty (used to define the “low” and “high” runs); often used in combination with natural mortality. Steepness and relative data weights were also used to define “low” and “high” runs.

In all of the 2007 assessments “statistical uncertainty” (within a given run) was dealt with using confidence intervals calculated using the inverted Hessian and making asymptotic multivariate normal assumptions. This is a poor substitute for using Bayesian posterior distributions (results can often be very different – for the intervals as well as point estimates).

The “current (assessment) approaches” on the West Coast, as seen in the 2007 assessments are singular: the use of SS2 on all/most available data. The assessments consist of point estimates obtained by minimizing the total negative log-likelihood including prior probabilities. The strengths and weaknesses of this approach cannot really be isolated from the STAR Panel process. The process is such that full Bayesian assessments cannot be conducted.

There is no doubt that the assessments which leave a STAR Panel meeting are technically better than those that enter the meeting. However, the quality of the assessments, on an absolute scale, is a moot question. If judged by New Zealand standards, I doubt that any of the 2007 assessments would have been accepted, because a full Bayesian assessment is generally required.

The 2007 process was not without its strengths:

- The inclusion of CIE reviewers provides the Panels with a good mix of local and international expertise.
- The workshop atmosphere lends itself to open discussion which can lead to the development of new ideas and useful generic recommendations.
- The discussions and exploration of assessments do lead to improved assessments.

But, there are some serious problems:

- Modifications to many of the assessments were so substantial that Panels obtained part “ownership” of the assessments. Therefore, the final assessments were not fully and *independently* reviewed.
- The process does not allow the use of the full Bayesian method, which is accepted (by many) as the best approach currently available.

Introduction

This is an “overview” report covering the 2007 STAR Panels and addressing the following general questions as specified in the TOR for this report (Appendix 1):

- Comment on the approaches and findings of STAR Panels that may or may not be consistent with each other or previous Panels.
- Comment on the general reporting of uncertainty in the stock assessments.
- Comment on the general strengths and weaknesses of current approaches along the West Coast.

I begin with a summary of the 2007 STAR Panel meetings, setting out the basic format and illustrating the general nature and operation of the meetings. I then cover each question in turn, and finish with a set of recommendations with regard to the process. I have answered the last question fairly broadly. The “current approach” on the West Coast is assessment with SS2 within the STAR Panel process – and it is the process that shapes the assessments.

The 2007 STAR Panel meetings

Six STAR Panel meetings were held on the west coast during 2007, five regular STAR Panels, during May-August, and then a “mop up” meeting, in October, held as part of a groundfish sub-committee meeting of the SSC. Each meeting was scheduled for four and half days (beginning Monday lunchtime, ending Friday evening) and involved three, four, or five reviewers including two CIE reviewers and an SSC member as the Chair (regular meetings). I attended all meetings including the “mop up” (as the only CIE reviewer).

Most meetings began with the Chair reminding the Panel that the meeting was a “review” and not a “workshop”. STAR Panel members were then assigned as rapporteurs for each assessment being reviewed. Their important role was that of drafting the formal requests to the STAT for exploratory analyses or model runs (rather than keeping detailed minutes – although the STAR Panel terms of reference do require that minutes are kept). Each STAT would then present their draft assessment and during the week the Panel and STATs would collaboratively work towards technically acceptable assessments (a base model, together with a “low” and “high” run – ideally with the base model having a 50% probability, and the “bracketing” runs each with 25%) – pretty much like a workshop.

The process began in May with a meeting in Newport, Oregon, covering sablefish and long-nosed skate. This meeting was characterized by the development, during the meeting, of prior distributions for trawl survey qs , which were very influential in both assessments – because fixed values of q were used to define the model runs. Both assessments were substantially modified by the STAR Panel process (especially long-nosed skate) and both final assessments were considered, by the Panel, an adequate basis on which to provide management advice.

Two weeks later in Portland, Oregon, three stock assessments were considered by a Panel: blue rockfish, black rockfish (north), and black rockfish (south). This meeting was unique in having two of the assessments rejected (and being the only meeting at which any assessments were rejected). Blue rockfish and black rockfish (south) were referred to the October “mop up” meeting, albeit for quite different reasons.

In the case of blue rockfish, the STAT had prepared the available data for use with SS2 but had failed to get the model working. They had then hurriedly constructed an ASPIC based assessment. The Panel rejected the assessment because it did not come close to meeting the standards required of modern stock assessments (*not* because it was done using ASPIC – although this certainly made it harder for it to be accepted).

For black rockfish (south), many runs were explored during the week, but none satisfied the STAT’s belief (based on tag-experiment estimates of fishing mortality) of low fishing mortality off Oregon. Therefore, the STAT had no model to offer, and the Panel was forced to send the assessment to the “mop up” meeting.

Black rockfish (north) was accepted by the Panel after some modification. This STAT also had strong beliefs with regard to fishing mortality (again based on tag based estimates – not available to the model). Their initial model gave results consistent with their belief. However, when an error was corrected in the model specification, the results were no longer consistent with their belief. The STAT made progressive increases to the assumed value of natural mortality during the meeting. By the end of the meeting they appeared willing to accept the results of their base model (which the Panel accepted). However, after the STAR Panel, the STAT yet again increased natural mortality (just for females) in the base model and the SSC accepted the larger value.

The third STAR Panel met in Santa Cruz in June and was scheduled to consider chilipepper rockfish and bocaccio rockfish. However, the bocaccio STAT only offered an update of the 2005 *update* of the 2003 assessment. The Panel decided to do a “pre-review” of the update to aid the SSC who were charged with reviewing updates. For chilipepper, a full assessment was presented and reviewed. The STAT had used conditional age-at-length data in his initial assessment but had found the results to be very unstable. For the meeting he returned to annual age frequencies, with associated length data. Many alternative models were explored during the week as the Panel and STAT worked towards an acceptable assessment – which was achieved.

The STAR Panels then shifted to Seattle for the last two normal meetings and the October “mop up”.

In July, cowcod and darkblotched rockfish were reviewed. Cowcod had originally been scheduled for only an update but errors had been found with the 2005 assessment that the SSC determined were too serious to allow an update to be accepted – and they called for a “full” assessment. However, there was very limited time available to the STAT to perform a full assessment and this was taken into consideration by the Panel during the

review. The most crucial aspect of the assessment was the relative weights given to conflicting abundance indices. The Panel and the STAT included this aspect together with natural mortality in a two-dimensional representation of model uncertainty. However (as I discovered at the mop-up meeting), in the rebuilding analysis only natural mortality was used (which was a minor component of the uncertainty).

The darkblotched assessment was perhaps modified least of all the assessments by the STAR Panel process. There was a dispute between the Panel and the STAT with regard to the estimation of steepness. The Panel contended that the available data contained no useful information with which to estimate steepness (data preference, in terms of improved fit, does not necessarily equate to information – often it is driven by model assumptions). The STAT was sympathetic to this contention but nevertheless insisted on estimating steepness (with an informed prior) and then fixing it at the estimated value. The SSC later supported the STAT's position.

The final regular STAR Panel reviewed canary rockfish and arrowtooth flounder. For canary, steepness was again an issue, with the STAT being unhappy about using an informed prior, from the Dorn meta analysis, which included the 2005 steepness profile for darkblotched rockfish. Their reasoning was that steepness for darkblotched had been revised downwards from 0.95 to 0.6, so that the previous Dorn prior was, in some sense, “biased” high. Dr. Dorn was on hand to produce a revised prior for canary steepness that excluded the old darkblotched profile. The Panel, somewhat reluctantly, accepted the use of the new prior. Also, the Panel Chair was keen to have the SWFSC pre-recruit indices used in the base model. The STAT eventually agreed to this.

The arrowtooth assessment was noteworthy for an innovative approach to estimating recruitment deviations. The model was started some 20 years before the catch data began and *all* recruitment deviations were estimated. This turned out to have a large effect on the assessment results as the early recruitments, for which no observations were available, were estimated to be below average, while recent recruitments were estimated to be above average. The net effect was a very optimistic estimate of depletion (170%). When recruitment deviations were estimated only when supported by data, estimated depletion was approximately 80% (the final base model used this approach).

The “mop up” meeting in October reviewed new assessments of blue rockfish and black rockfish (south). In addition, several rebuilding analyses were considered and there was time put aside for a general discussion on rebuilding analyses and the SSC's role in reviewing them. The Panel now consisted of the SSC groundfish sub-committee with myself as the only CIE reviewer.

The review of the new black rockfish assessment was relatively straight forward. However, data weighting issues were critical with the assessment results being very dependent on the weight given to mean length-at-age data. The blue rockfish review was more problematic as the assessment was not well crafted. SS2 was used this time, although the STAT still professed a preference for ASPIC models (but these were never

presented). An acceptable assessment was reached but, strangely, with the STAR Panel somewhat more confident of the acceptability of the assessment than the STAT.

Comparison of findings and approaches

For the purposes of this report I made a comparison of key parameter values and assumptions used in the 2007 assessments and those used in the preceding assessment. Sablefish, cowcod, darkblotched, and canary were last assessed in 2005, with black rockfish (south) last assessed in 2003, and chilipepper in 1998. For this comparison I have not included black rockfish (north) which was last assessed in 1999 (using a very different model), or bocaccio, which was an update. Long-nosed skate, blue rockfish, and arrowtooth flounder were being assessed, with a model-based approach, for the first time.

Values of natural mortality used in the base models (either fixed or estimated) stayed essentially the same for four of the six assessments, but were increased for black rockfish (south), and decreased for chilipepper (Table 1). The increase in black rockfish (south) was due to the black rockfish (north) assessment and the desire of the mop-up Panel to have consistent values for the two assessments. In my opinion, the values of natural mortality used in the black rockfish (north) assessment were result-driven rather than data-driven. The lower values of natural mortality in the chilipepper assessment were from a new internal-model fit to the data (with a new model and much more data since the last assessment in 1998).

Table 1: Values of natural mortality, steepness, recruitment variability, and start year for catch history used in the 2007 assessments and the previous assessments for six rockfish species.

	M (male)		M (female)		Steepness		σ_R		Catch history	
	prev.	2007	prev.	2007	prev.	2007	prev.	2007	prev.	2007
Sablefish	0.07	0.07	0.07	0.07	0.34	0.48	0.28 ¹	0.6	1900	1900
Black (south)	0.12	0.16	0.20 ²	0.24 ²	0.65	0.60	0.4	0.5	1945	1915
Chilipepper	0.25	0.20	0.22	0.16	1.00	0.57	1.0	1.0	1970	1892
Cowcod	.055	.055	.055	.055	0.50	0.60	0.0	0.0	1916	1900
Darkblotched	0.07	0.07	0.07	0.07	0.95	0.60	0.8	0.8	1928	1928
Canary	0.06	0.06	0.09 ²	0.10 ²	0.33	0.51	0.4	0.5	1916	1916

1: sea level was an environmental covariate

2: “ramp” from male M to this value for older females

Values of steepness used in the Beverton-Holt stock recruit relationship changed markedly for four of the six assessments, and were also altered for the other two (Table 1). Generally, this was the result of new data. There is a great deal of consistency in the values used for the rockfish assessments in 2007 (ranging from 0.48 to 0.6 – see Table 2) which is due to the use of species-specific informed priors from the Dorn meta-analysis (either the median value or estimation using the informed prior).

The values of σ_R changed little between the previous assessments and the 2007 assessments (Table 1, note that the 2005 sablefish value is not comparable with that in

2007 because of a change in how environmental variables were used). In the usual SS2 formulation, σ_R can unfortunately have quite a large effect on results as it is used in a penalty function to ensure that recruitment deviations follow the stock-recruitment relationship (but the user does have the option to reduce the emphasis on the stock-recruit relationship). In the 2007 assessments there was little consistency in how σ_R was determined and there were a wide range of values used (Table 2). In some cases, the value was tuned (input value adjusted until the standard deviation of the recruitment deviations (“output value”) was approximately equal to the input value); in other cases, it was partially tuned (choose a relatively high value – to “let the data speak” – and then fix it at the output value). In the case of cowcod, recruitment was assumed deterministic (due to lack of data).

In 2007, every STAT attempted a full catch history reconstruction rather than relying on an assumed equilibrium catch (see catch history start years in Table 1). This is possibly the best general approach, but for some assessments it may be that better results could be achieved by estimating an initial age-structure in the absence of a full catch history. If a full catch history is constructed, then not only “best guess” estimates are needed, but a catch history envelope within which the effects of alternative catch streams can be tested. This was not well done in 2007.

Table 2: Values of steepness and recruitment variability used in the 2007 stock assessments.

	Steepness	σ_R
Sablefish	0.48	0.6
Skate	0.40	0.0
Black (north)	0.60	0.35
Black (south)	0.60	0.5
Blue	0.58	0.5
Chilipepper	0.57	1.0
Cowcod	0.60	0.0
Darkblotched	0.60	0.8
Canary	0.51	0.5
Arrowtooth	0.90	0.8

As the “continuity” CIE reviewer on the 2007 STAR Panels, where possible, I encouraged each Panel to adopt similar approaches – to achieve some level of “consistency” across assessments. However, different issues arose at each meeting; different STATs and Panelists tended to emphasize different aspects of assessments or be concerned about different technical issues. Also, my experience over the meetings was cumulative rather than “simultaneous”. I took lessons from earlier Panels and applied them to the later ones – the converse was not possible.

There was, of course, much greater consistency across assessments within each meeting than between meetings. Assessments considered together, by the same panel, naturally had the same technical issues considered. Across meetings, this was much more difficult to do – because of a change in personnel and assessments. For example, the Newport meeting was dominated by trawl survey *qs* – crucial, because the indices were being used

as absolute indices. In contrast, the mop-up meeting began by focusing on fits to pre-recruit indices (because one assessment showed an exact fit, while the other didn't) and finished with a focus on relative weighting of data sets. Also, noteworthy was the Portland meeting where the black rockfish STATs were very much focused on the estimates of exploitation rate obtained from tag experiments or surveys (and wanting the assessment results to be consistent with these estimates – despite the estimates not being provided to the model – and not having quantified the biases in the tag estimators).

Reporting of uncertainty

In the 2007 assessments, uncertainty in catch history was often incorporated into a two-dimensional specification of uncertainty (Table 3). Its use often had little impact on estimates of depletion, but contributed strong contrast in estimates of yield. In the previous assessments catch history only featured in the 2003 black rockfish (south) assessment (for which I was on the Panel). Other components of uncertainty used in 2007 were natural mortality, steepness, data weights (the relative weights of “conflicting” data sets), and trawl survey qs (Table 3).

Table 3: The components of uncertainty used in the 2007 assessments and the corresponding previous assessments (where available).

	2007	Previous
Sablefish	trawl survey q	trawl survey q , steepness (but tiny steps)
Skate	catch history, trawl survey q	
Black (north)	M	
Black (south)	catch history, M	catch history
Blue	catch history, M	
Chilipepper	steepness	data weights
Cowcod	data weights, M	steepness
Darkblotched	M	M
Canary	steepness	selectivity (but really a steepness proxy)
Arrowtooth	catch history, M	

The choice of the “main dimension of uncertainty” was often a rather hurried procedure in 2007. Often, the base model was not determined until Friday and “low” and “high” runs were relatively quick “add-ons”. In the cases of arrowtooth flounder and blue rockfish the results of “low” and “high” runs were not seen by the Panel until after the meeting. In the case of arrowtooth flounder, I don't recall ever seeing the results of the runs. Also, apparently there was a problem with the results being too extreme and the dimension of uncertainty was changed – but I was never informed of the details.

The choice of “low” and “high” runs (which represent model uncertainty) is very subjective. The guideline to Panels/STATs was to choose a base model representing 50% “probability” and to have low and high runs each representing 25% probability. In the case of single dimensions of uncertainty, where fixed values of parameters were used and for which an informed prior was available, this guideline was followed by using the

median value for the base model and the mean of the lower and upper quartiles for the low and high runs. This is a semi-objective approach but it only applies to single dimensions of uncertainty. There is no general method by which the guideline can be objectively followed. Even the use of full Bayesian methods will be problematic because “model uncertainty” (across models) and “statistical uncertainty” (within a given model) both need to be considered in a full assessment.

In all of the 2007 assessments “statistical uncertainty” was dealt with using confidence intervals calculated using the inverted Hessian and making asymptotic multivariate normal assumptions. This is a poor substitute for using Bayesian posterior distributions. I was told that in some previous assessments that comparisons had been made between the point estimates and approximate confidence intervals and the full Bayesian distributions. The results were found to be similar. This is comforting for those particular assessments. However, mathematically the estimators (point and interval) are different and I am unaware of a theorem which provides general conditions under which they produce similar results. I know that in some assessments the estimators have very different properties (and provide very different estimates).

In some of the 2005 assessments I am aware that uncertainty was “captured” by using (low and high) quantiles of parameters obtained from approximate confidence intervals of the base model. This is a very poor approach as the breadth of the uncertainty is driven by the assumptions made in the base model. It is no less subjective and far less instructive than choosing a dimension or two of sensitivity and using them to illustrate the robustness (or otherwise) of the base model results.

Strengths and weaknesses of the STAR Panel process

The “current (assessment) approaches” on the West Coast, as seen in the 2007 assessments are singular: the use of an integrated statistical model (SS2) on all/most available data. The assessments consist of point estimates obtained by minimizing the total negative log-likelihood including prior probabilities. Approximate confidence intervals are calculated, but full Bayesian posterior distributions are not. The strengths and weaknesses of this approach cannot really be isolated from the STAR Panel process. The process is such that fully Bayesian assessments cannot be conducted. Also, STATs are aware that their draft assessment will be changed during the STAR Panel meeting. A common response is to present an incomplete assessment – they are perhaps not willing to do the work of a full assessment when they know much of it will need to be redone.

There is no doubt that the assessments which leave a STAR Panel meeting are technically better than those that enter the meeting. However, the quality of the assessments, on an absolute scale, is a moot question. If judged by New Zealand standards, I doubt that any of the 2007 assessments would have been accepted, because a full Bayesian assessment is generally required. The 2007 West Coast assessments are all based on MPD estimates (Mode of joint Posterior Distribution). In New Zealand, sensitivity analysis will make

much use of MPD estimates, but management decisions will almost always be based on medians of (marginal) posterior distributions (and probabilities calculated from full MCMC projections).

The STAR Panel process does not lend itself to Bayesian assessment because the production of full posterior distributions (using MCMCs) can be very time consuming. STAR Panels operate as workshops (despite the terms of reference) and, typically, many runs are done during a meeting with, often, wholesale changes to models. There simply isn't time for MCMCs. That said, often model uncertainty is more important than uncertainty within a given model, so the omission of MCMC runs is not necessarily all bad.

Of course, capturing model uncertainty at the assessment stage is futile if it is then ignored in projections or a rebuilding analysis. In the case of cowcod, the most important aspect of uncertainty was the conflict between data sets. This was captured by the STAT and the STAR Panel in the final assessment with runs that combined alternative values of natural mortality and data weights. However, only natural mortality was used in the rebuilding analysis – and, hence, the major uncertainty was completely ignored.

Another problem I had with the process this year was the lack of feedback and follow-up to STAR Panel members when problems were detected by the SSC or changes were made to assessments and/or reports. At the Santa Cruz meeting, purely by chance, I heard that there had been a problem with the Newport meeting assessments (from a month earlier). I eventually confirmed that there was a problem and found a solution for it (that can be applied to future assessments). Also, the STAR Panel report for the rejected blue rockfish assessment was modified by the USA based members of the Panel without consultation with the CIE reviewers. My attendance at the mop-up meeting was the only reason I discovered this occurrence. I was also concerned to see that an appendix, written by the blue rockfish STAT, had been included in the STAR Panel report. At my initiative, the full Panel has since agreed a revised STAR Panel report.

The biggest problem for the STAR Panel process is that assessments can be and are substantially modified on the basis of suggestions from Panel members. In 2007, there were substantial changes to many of the assessments. Who, then, did the review? Unfortunately, the changes were so substantial that the Panel obtained part “ownership” of the assessments. They cannot be said to have reviewed an assessment which they helped create. Of course, there is the SSC who have a final look at each assessment. But they do not do a substantive review, perhaps spending half an hour considering each assessment. Perhaps longer, when they entertain and accept changes to an assessment (as in the case of black rockfish (north)). But, then, even if the assessment had been properly reviewed, they have changed it. The SSC should not be making any changes to reviewed assessments – they, surely, should operate on an “accept” or “reject” basis.

The 2007 process was not without its strengths:

- The inclusion of CIE reviewers provides the Panels with a good mix of local and international expertise.
- The inclusion of a single CIE reviewer in all Panels improved the consistency of decisions across Panels (but there is a personal cost for the reviewer in such a large time commitment).
- With only two assessments at each meeting there was adequate time to address all of the *major* assessment issues (this was not the case in 2005, when there were too many assessments at each meeting).
- The workshop atmosphere lends itself to open discussion which can lead to the development of new ideas and useful generic recommendations.
- The discussions and exploration of assessments do lead to improved assessments.

But, there are some serious problems:

- Modifications to many of the assessments were so substantial that Panels obtained part ‘ownership’ of the assessments. Therefore, the “final” assessments were not fully and *independently* reviewed.
- In some cases, the assessments recommended for management use by the SSC differed from those accepted by the STAR Panels. The SSC does not fully review assessments but did modify assessments accepted by STAR Panels. Therefore the *final* assessments were not fully reviewed.
- A crucial dimension of uncertainty incorporated into a final assessment (cowcod) was ignored in the rebuilding analysis (which essentially overrode the accepted assessment and replaced it with an assessment that was rejected by the STAR Panel).
- The process does not allow the use of the full Bayesian method, which is accepted (by many) as the best approach currently available.

There were also some other procedural problems:

- A STAR Panel Chair cannot effectively chair a meeting if they also act as a reviewer. Some chairs were 95% chair and 5% reviewer, but others were 80% reviewer and 20% chair. The latter is not appropriate.
- Expecting reviewers to act as rapporteurs as per the terms of reference (which call for the keeping of minutes) inhibits a reviewer’s ability to participate in discussions. After the first meeting, I explicitly refused to take detailed minutes and encouraged that this aspect of the rapporteurs’ role to be assigned to non-Panel meeting participants.
- The STAR Panel terms of reference state that the final assessment document will be reviewed by the Panel. This never happens as the Panel rarely consults after the STAR Panel reports are agreed (by email after the meeting). The Chair is the only Panel member who takes any interest in the assessment document after the meeting is concluded.

- Panel members were not informed of problems identified with assessments at the SSC level (as a participant at all meetings I, at least, should have been advised).
- Panel members were not informed of changes made to what they believed were final STAR Panel reports (this is clearly not acceptable – all Panel members are joint authors).
- The SSC groundfish sub-committee meeting (the “mop up” meeting) was not an appropriate forum for reviewing assessments. Sub-committee members came and went at the meeting as they had other commitments. At times there were too many participants; at other times there were not enough. Also, the time to do the reviews was restricted because of the other business the sub-committee had to deal with.

Recommendations

The procedural problems can fairly easily be addressed with some explicit instructions to Panel chairs and some modification of the terms of reference. My suggestions in this regard are:

- A STAR Panel chair should primarily act as a chair and only participate as a reviewer, intermittently, when absolutely necessary.
- If detailed minutes are required then they should be taken by non-Panel participants.
- The terms of reference for STAR Panels should be reviewed and revised to reflect the realities of the limited temporal availability of Panels.
- If the SSC identifies a problem with an assessment which was caused by a STAR Panel then the Panel members should be informed of the nature of the problem.
- A STAR Panel report once finalized by a Panel must not be modified (except editorially) without the agreement of the full Panel.
- A STAR Panel meeting should not be conducted as part of an SSC groundfish sub-committee meeting.

The serious problems with the STAR Panel process are much harder to address and recommendations on how to address them would be well beyond my terms of reference for this report.

Appendix 1: Statement of work

March 22, 2007

General

In FY 2007, five Stock Assessment Review (STAR) Panels are scheduled to each review one or two assessments of West Coast groundfish stocks as indicated in the specific schedule below. In the past, STAR Panels have reviewed two or three species during a 5-day meeting. In order for Panels to complete reviews of all scheduled assessments, any assessments with unresolved issues requiring time-intensive analyses and discussion will be postponed until a “wrap-up Panel” tentatively scheduled for October 1 – 5, 2007.

Panel	Assessments Reviewed	Location	Dates	Status
STAR Panel 1	Long-nose skate, sablefish	Newport, OR	May 7-11, 2007	Definite
STAR Panel 2	Black rockfish, blue rockfish	Portland, OR	May 21-25, 2007	Definite
STAR Panel 3	Bocaccio, chilipepper rockfish	Santa Cruz, CA	June 25-29, 2007	Definite
STAR Panel 4	Darkblotched rockfish	Seattle, WA	July 16-20, 2007	Definite
STAR Panel 5	Canary rockfish, arrowtooth flounder	Seattle, WA	July 30-August 3, 2007	Definite
STAR Wrap-Up Panel	TBD	Seattle, WA	October 1-5, 2007	Dependent on other STAR Panel reviews

A CIE expert participating in all panels will greatly improve the consistency among the many STAR Panels as well as improve the quality and credibility of the assessments. These assessments will provide the basis for the management of groundfish on the West Coast.

The CIE expert participating in all panels should have experience in population dynamics and stock assessment of groundfish. The expert should have specific experience in the integrated analysis type of modeling approach, using ADMB, age-and size-structured models, use of MCMC to develop confidence intervals, and use of Generalized Linear Models to process survey and logbook data for use in assessment models.

A second CIE review panelist and several NMFS selected panelists will also participate in each of the STAR Panels. 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.

Documents to be provided to the CIE consultant prior to each of the STAR Panel meetings include the following:

- Current draft stock assessment reports
- Most recent previous stock assessments and STAR Panel reports
- Groundfish Stock Assessment and Review Process Terms of Reference
- An electronic copy of the data, the parameters, and the model used for the assessments (if requested by reviewer).

Terms of reference and specific questions regarding each of the STAR Panels will be developed and distributed to the CIE expert prior to the meeting. The CIE consultant is required to participate in each STAR Panel and address the specific terms of reference, as well as develop a summary report following all STAR Panels. The deadline for the final summary report will be dependent on scheduling of the final tentative STAR wrap-up Panel as indicated in the table below.

Specific

The consultant's duties shall not exceed a maximum total of 85 days: participation in five (5) five-day STAR Panels, and a potential sixth final wrap-up Panel, with several days prior to the meeting for document review and several days following the meeting to complete the written report (see Annex 1 for details on report content). The reports developed by the CIE consultant on each STAR Panel should be based on the consultant's findings, and should not represent consensus of all panelists participating in each review.

The CIE expert will participate in all of the following STAR Panels (dates may be subject to future changes):

- May 7-11, 2007 in Newport, OR (long-nose skate, sablefish)
- May 21-25, 2007 in Portland, OR (black rockfish, blue rockfish)
- June 25-29, 2007 in Santa Cruz, CA (Boccaccio, chilipepper rockfish)
- July 16-20, 2007 in Seattle, WA (darkblotched rockfish)
- July 30-August 3, 2007 in Seattle, WA (canary rockfish, arrowtooth flounder)
- October 1-5, 2007 Wrap-Up (if needed)

The consultant's tasks consist of the following:

1. Become familiar with the draft stock assessments and background materials for each of the STAR Panels.

2. Actively participate in each of the STAR Panels as indicated in the STAR Panel schedule.
3. Contribute to Panel discussions and reports as a member of the STAR Panels, specifically addressing the terms of reference and questions posed for each STAR Panel.
4. Complete a report following each STAR Panel and submit to the CIE according to the submission schedule below.
5. Complete one final summary report after the last STAR Panel addressing the following general questions:
 - Comment on the approaches and findings of STAR Panels that may or may not be consistent with each other or previous Panels.
 - Comment on the general reporting of uncertainty in the stock assessments.
 - Comment on the general strengths and weaknesses of current approaches along the West Coast.
6. Provide all STAR Panel reports to the CIE 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 Shivilani, via e-mail to mshivilani@rsmas.miami.edu.

STAR Panel	STAR Panel Meeting Dates	Submission Deadlines
STAR Panel 1	May 7-11, 2007 STAR Panel	Consultant summary report due to CIE: May 25, 2007 CIE summary report due to NMFS: June 8, 2007
STAR Panel 2	May 21-25, 2007	Consultant summary report due to CIE: June 8, 2007 CIE summary report due to NMFS: June 22 2007
STAR Panel 3	June 25-29, 2007	Consultant summary report due to CIE: July 13, 2007 CIE summary report due to NMFS: July 27, 2007
STAR Panel 4	July 16-20, 2007	Consultant summary report due to CIE: August 3, 2007 CIE summary report due to NMFS: August 17, 2007
STAR Panel 5	July 30-August 3, 2007	Consultant summary report due to CIE: August 17, 2007 CIE summary report due to NMFS: August 31, 2007
<i>Potential 6th Wrap-Up STAR Panel:</i>		
Wrap-Up STAR Panel	October 1-5, 2007	Dates for submission of reports to CIE and NMFS will be determined upon scheduling of the wrap-up STAR panel.

Submission and Acceptance of Reviewer's Report

The CIE shall provide via e-mail all deliverables listed in the schedule above to Dr. Lisa Desfosse (Lisa.Desfosse@noaa.gov) for review by NOAA Fisheries and approval by the COTR, Dr. Stephen K. Brown, by the deadlines listed in the schedule. Approval shall be based on compliance with this statement of work and the terms of reference for the specific STAR Panel. The COTR shall notify the CIE via e-mail regarding acceptance of the reports by the deadlines listed in the schedule above. The COTR will then transmit the report to the NWFSC contact, as soon as possible.