Report on

Sablefish and Longnose Skate STAR Panel

May 7-11, 2007

Pacific Fishery Management Council

National Marine Fisheries Service

Hatfield Marine Science Center

Newport

Oregon

P.L. Cordue

Fisheries consultant

New Zealand

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

20 May 2007

Executive summary

A STAR Panel met from May 7-11, 2007 in Newport, Oregon, to review assessments of sablefish and longnose skate. The Panel consisted of two CIE reviewers 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 draft sablefish assessment was performed using a recent version of Stock Synthesis 2 and modeled a single stock in the waters off Oregon, Washington, and California. Five alternative runs were presented. In the proposed base model, nine data sources were used including two environmental variables (used as recruitment indices). All model configurations included available length, age, and biomass data from four bottom trawl surveys of the slope or shelf. Available length and age data from trawl, hook and line, and pot fleets were included. The estimated catch history extended back to 1915 (split into the three methods). All proposed models had the NWFSC slope survey ("fleet 7") q fixed at 1.

The STAR Panel were concerned about many aspects of the proposed base model. In general terms, there were three main concerns. First, the model appeared too complex relative to the expected information content of the available data. Second, the age, length, and length-at-age data from each data source were input into the model as if they were independent (when they clearly were not). Finally, the assumption of q=1 had no firm basis. In specific terms, the STAR Panel had one over-riding concern with the base model. The model did not fit the fleet-7 abundance indices despite the assumption of q=1. The expected values were almost all larger than the observed values (they went "over the top").

The STAR Panel and STAT worked towards a new base model by making progressive changes to the proposed base model. The fleet-7 biomass indices were flawed in their calculation and a modified time series was used. An informative prior was developed for the fleet-7 q to give a sensible basis for choosing fixed values of q (which did prove necessary). The "over the top" problem was fixed by down-weighting the commercial fishery age and length data.

The draft longnose skate assessment modeled a single stock in the waters off Washington, Oregon, and California. A two-sex model was implemented in Stock Synthesis 2 using mainly length data, biomass indices from trawl surveys, and some limited age data. The landings history was input into the model with estimated discard rates and an assumed discard mortality. The model started in 1980 with a non-zero historical fishing mortality (based on estimated average annual catches 1950-1979).

The model was more complex than it needed to be and the STAR Panel recommended a single-sex model with deterministic recruitment. Many other changes were suggested and adopted. There were two important changes. First, an informed prior was developed for the NWFSC shelf-slope survey so that there was a firm basis for choosing fixed values of *q* (which did prove necessary). Second, the characterization of uncertainty in total fishery-induced mortality was captured by explicit catch histories from the start of the fishery in 1915, rather than by the numerous sub components (historical *F*, discard rates, discard mortalities, proportion of longnose in the total skate landings).

The final assessment had a base model with *q* fixed at the median of the prior and used the “best” (or “mid”) catch history. The two sensitivity runs combined both major dimensions of uncertainty to bracket the base model. The “low” *q* was combined with the “low” catch history to provide an “optimistic” run and the “high” *q* was combined with the “high” catch history to provide a “pessimistic” run.

Both stock assessments were technically improved by the STAR Panel process. However, the sablefish assessment retains its general deficiencies. The assessment model and data inputs need to be subjected to a full and extensive review. In the longnose skate assessment the use of landings to construct catch histories should be revisited. It would be preferable to use an effort-based approach. In other regards, the longnose skate assessment can perhaps serve as a template for other stock assessments with similar data availability.

Background

The sablefish and longnose skate STAR Panel met in the Captain R. Barry Fisher Building at Hatfield Marine Science Center, Newport, Oregon, from May 7-11, 2007. This was the first of five 2007 STAR Panels in the biennial meeting schedule. In contrast to 2005, when up to five species were considered by each STAR Panel, assessments for only two species were presented. This allowed more time for presentations, questions, answers, and discussion. However, because each STAT consisted of only a single person, the reduced number of assessments considered had no benefits in terms of the STATs ability to perform additional exploratory runs or analyses.

The two assessments presented an interesting contrast. It was the first assessment of longnose skate which was performed by an enthusiastic young scientist performing her first stock assessment. In contrast, sablefish assessments have been done many times before and the 2007 assessment was performed by a senior scientist with a lot of experience with sablefish assessments.

The STAR Panel had only three members. My two colleagues were, Dr Martin Dorn, the SSC representative and Panel Chair and Vivian Haist, my fellow CIE reviewer. This report should be read in conjunction with the STAR Panel reports and Haist's CIE report.

Review Activities

Pre-meeting

Meeting documents and materials were received in electronic form well in advance of the meeting (*see* Appendix 4). I familiarized myself with the background material and current assessments prior to the meeting. Paper copies of the substantial assessment documents were also made available at the meeting, which was helpful.

Meeting

The meeting was convened at 12:30 pm on Monday, May 7, 2007 and closed at about 6 pm Friday, May 11, 2007. I will only give a brief summary of the meetings activities. Details of the requests to the STATs and their responses are contained in the STAR Panel reports. I was the designated rapporteur for sablefish and Haist covered longnose skate.

The first afternoon was devoted to sablefish, with Dr Schirripa presenting his assessment. A full assessment, which was well documented, had been brought to the meeting. The presentation of the assessment proceeded relatively slowly as the Panel asked questions and discussions developed. The main aspects of the assessment were covered in the presentation, but due to a lack of time, part of the STAT's presentation, on environmental variables, was held over until later. A set of requests for the sablefish STAT was drafted before the meeting closed for the day.

The Tuesday morning was devoted to the presentation of the longnose skate assessment by Dr Gertseva. As with sablefish, a full assessment had been completed and documented. The presentation proceeded with fewer questions by the Panel than for sablefish. The sablefish assessment had a complex model and used many data sets. In contrast, the longnose skate assessment was "data poor" with a simpler model. Over lunch, the Panel and STAT agreed on a set of requests.

After the presentations of the assessments it was clear that both assessments were unacceptable to the Panel. However, in both cases the Panel believed that acceptable assessments could be attained before the end of the meeting. The Panel made suggestions for alternative model configurations, and guided the STATs towards technically acceptable assessments.

From Tuesday afternoon, the meeting progressed, generally, with alternating sessions on the two species. The usual pattern was for the STATs to present the results/progress on their current set of requests (for runs and/or analyses), and after discussion of the results, for the Panel and STAT to agree on another set of runs/analyses (for details of the requests and the STATs responses see the STAR Panel reports). One joint session was held where informed priors were developed for the trawl survey proportionality constants (q) that each STAT had fixed equal to 1 (*see* Appendix 2).

Progress towards acceptable assessments was mixed. In the case of longnose skate, the base run and sensitivities were finalized on Friday morning after an exploratory MCMC run failed to converge (so a fall-back position of three fixed-q runs, using the informed prior, was adopted). In the case of sablefish, there were no acceptable runs as of early Friday afternoon. However, the last set of runs provided acceptable diagnostics and a base model and three sensitivity runs were accepted late Friday afternoon.

Post-meeting

As rapporteur for sablefish, I completed a draft STAR Panel report over the weekend and circulated it by email to the other members of the STAR Panel. Likewise, Haist circulated a draft longnose skate report. The reports were revised via a brief email collaboration amongst the Panel (before being circulated to other meeting participants, including the STATs, for checks on factual accuracy). The Chair received comments from the sablefish STAT which were forwarded to me for my comments. The STAT contended that he had fully completed three requested runs which we had said were only partially completed. This is a minor dispute which will hopefully be resolved by a slightly different choice of wording in the final STAR Panel report (the Chair will make some minor changes).

I also drafted a brief document for distribution to STATs that were yet to present their assessments. A few issues had arisen at the first meeting which clearly had potential to arise at every meeting. I provided a draft to Dr Dorn for comment and at his suggestion we sought and received comments from Dr Methot and Dr Stewart on the suggested iterative re-weighting procedure. If other issues arise at subsequent STAR Panel meetings, I may provide some further suggestions to STATs (*see* Appendix 3).

Review findings

For each species, the original and revised stock assessments are discussed below under three sub-headings. First, I summarize the original stock assessment, the concerns that the Panel had with the original assessment, and the changes that were made by the STATs to arrive at acceptable assessments. Second, I discuss the major sources of uncertainty in the revised assessments. Finally, I list the strengths and weaknesses of the revised assessments.

Sablefish

The sections below are worded not too differently from the sablefish STAR Panel report – this is because I wrote the first draft of the sablefish report.

Assessment summary

The draft sablefish assessment was performed using a recent version of Stock Synthesis 2 (SS2) and modeled a single stock in the waters off Oregon, Washington, and California. Five alternative runs were presented. In the proposed base model, nine data sources were used including two environmental variables (used as recruitment indices). The proposed runs excluded logbook and pot survey indices which were used in the previous assessment. All model configurations included available length, age, and biomass data from four bottom trawl surveys of the slope or shelf. Available length and age data from trawl, hook and line, and pot fleets were included. The estimated catch history extended back to 1915 (split into the three methods). All proposed models had the NWFSC slope survey (“fleet 7”) q fixed at 1.

The STAR Panel was concerned about many aspects of the proposed base model. In general terms, there were three main concerns. First, the model appeared too complex relative to the expected information content of the available data. Second, the age, length, and length-at-age data from each data source were input into the model as if they were independent (when they clearly were not). Finally, the assumption of q=1 had no firm basis. In specific terms, the STAR Panel had one over-riding concern with the base model. The model did not fit the fleet-7 abundance indices despite the assumption of q=1. The expected values were almost all larger than the observed values (they went “over the top”).

The STAR Panel and STAT worked towards a new base model by making progressive changes to the proposed base model. Minor changes included the incorporation of some discard rate data that were available but had not been used, a tightening of the allowed variability on the annual fishery selectivities, and the exclusion of the zooplankton time series. Iterative re-weighting procedures were also applied to the age and length frequency data sets.

An important change was made to the fleet-7 biomass indices. In some years, in the Conception stratum, all trawl stations were north of Point Conception and the average catch rate had been applied to the whole stratum area – despite catch rates being known to generally be much lower south of Point Conception. A new biomass time series was obtained in which the Conception stratum extended only to Point Conception.

An informative prior was developed for the fleet-7 *q* by considering its individual components and using the opinions of meeting participants (and some data) to bound each component. The median of the prior was obtained by using the "best guesses" for each component. The range on *q* was (0.22, 0.86) with a "best guess" of 0.56. A prior was formed by equating the "bounds" to 99% of the distribution (*see* Appendix 2). The prior was used in a model run but the estimated *q* was well outside the "bounds". The decision was made to fix *q* at the median of the prior.

After the above changes were incorporated the revised base model still exhibited the "over the top" problem. This problem was "solved" by down-weighting the commercial fishery age and length frequencies (by shifting the emphasis level from 1.0 to 0.1). This is a pragmatic approach, which the STAR Panel and STAT agreed was justified given the uneven spatial and temporal coverage of the commercial fishery sampling (and hence the large potential that the data were not representative).

Uncertainty in the base model was represented by three sensitivity runs: a lower *q*, a higher *q*, and a run excluding the remaining environmental time series (sea surface height).

Primary sources of uncertainty

The assessment results are driven by the prior on *q*, from which the three fixed values of *q* were derived. Ideally, *q* would be estimated, using the informed prior, in one or more model runs and the uncertainty associated with each model would be described by the associated posterior distributions of key outputs. This result could not be achieved during the STAR Panel meeting.

Major uncertainties:

- The value of q remains very uncertain.
- The low-q and high-q sensitivity runs are only indicative of potential biases in the base model; they do not span the full range of uncertainty.

- There is uncertainty associated with other fixed and estimated parameters including natural mortality and steepness. The implications of errors in these parameters were not explored during the meeting.

Strengths and weaknesses of current approach

Relative to previous assessments the 2007 assessment was somewhat simplified. Some changes in the model parameterization had been made and the choice of input data was reviewed which resulted in the exclusion of two problematic data sets. I appreciated the sentiments, as did the other members of the STAR Panel. This was a “step in the right direction” but I view the efforts as very small steps relative to what needed to be done.

The revised assessment is much improved from a technical basis with the elimination of three major problems: an assumed fleet-7 q=1 (without an adequate basis for the assumption); an “over the top” fit to fleet-7 biomass indices; and flawed biomass indices for fleet-7.

Its original merits remain but so do its *general* deficiencies. It is the best available assessment and I believe it is adequate to inform management. However, there remains a possibility, that if the general deficiencies were rectified, that the same assessment conclusions would not be reached.

Merits:

- Efforts were made to simplify the model and to apply greater discrimination in the use of some data sets.
- The fixed values of q have an informed basis.
- SS2 was used and as such brings the advantages of a standard and well tested package.
- Environmental variables were used as recruitment indices which is technically superior to the previous approach (where they modified the stock recruitment relationship).

Deficiencies:

- The complexity of the model is not justified given the likely information content of the available data.
- The use of combined age and length selectivities makes the interpretation of model results extremely difficult. While the concept is not too difficult, the effect that the use of such a complex parameterization has on model results is very difficult to understand. The parameterization also appears unnecessary given that growth morphs are not being used (and so the complexity is imposed simply to fit problematic length data that should probably not be used in any case.)
- Many of the data sets have not been scrutinized and analyzed nearly enough to justify their inclusion in base model runs.

- The age, length, and length-at-age data are used inappropriately. It may not be uncommon to use “all of the data” in this way, but it is technically incorrect. In the case of sablefish it is also unwise. There is almost no genuine information on recruitment (or biomass) in the length data which is not already contained in the age data.
- It was apparent that the STAT had used ad-hoc methods, at unspecified times in the past, to get the model “working”. This had included fixing selectivity parameters and adding temporary (made-up) data. Some of this data remains in the input files even in the final runs.
- The link between the environmental indices and recruitment remains to be validated (although current results are encouraging).
- The prior on *q* was not derived under optimal conditions (data and/or expertise that may exist was not available at the time).
- A detailed analysis of residual patterns appears not to have been undertaken in recent assessments. E.g., an investigation of sex ratios and whether the patterns are adequately explained by the current model.

Longnose skate

Assessment summary

The draft assessment modeled a single stock in the waters off Washington, Oregon, and California. A two-sex model was implemented in SS2 using mainly length data, biomass indices from trawl surveys, and some limited age data, to estimate virgin biomass, growth parameters and recruitment deviations. The landings history was input into the model with estimated discard rates and an assumed discard mortality. The model started in 1980 with a non-zero historical fishing mortality (based on estimated average annual catches 1950-1979). Uncertainty in the results had been characterized primarily by asymptotic 95% confidence intervals and a likelihood profile on discard mortality rate.

The draft assessment was a good attempt by an enthusiastic young scientist doing her first stock assessment – and the first for longnose skate. The model was more complex than it needed to be and the STAR Panel recommended a single-sex model with deterministic recruitment. I suggested that the simplest and most transparent way to capture the uncertainty associated with the unknown catch history was to construct “low”, “high” and “mid” catch histories from the start of the fishery in 1915 (*see* Appendix 1). Several other suggestions were made by the STAR Panel to which the STAT was most receptive. Steady progress was made towards a base model for which an MCMC run was made. Unfortunately, the chain had clearly not converged and for characterization of uncertainty fixed-*q* runs with low and high catch histories were used.

There were numerous changes made to the draft assessment (*see* the STAR Panel report) and the final assessment provides a much better characterization of the true uncertainty.

There were two important changes. First, there was the move from an assumed q=1 (with no firm basis) to fixed values derived from an informed prior (*see* Appendix 2). Second, the characterization of uncertainty in total fishery-induced mortality was captured by explicit catch histories from the start of the fishery, rather than the numerous sub components (historical F, discard rates, discard mortalities, proportion of longnose in the total skate landings).

The final assessment had a base model with q fixed at the median of the prior and used the "best" (or "mid") catch history. The two sensitivity runs combined both major dimensions of uncertainty to bracket the base model. The "low" q was combined with the "low" catch history to provide an "optimistic" run and the "high" q was combined with the "high" catch history to provide a "pessimistic" run.

Primary sources of uncertainty

The assessment for longnose skate is undoubtedly "data poor". The catch history (fishery-induced mortality) is poorly known, ageing is yet to be validated, the maturity estimates used differ substantially from those of another study, and the assessment results depend strongly on an informed prior for the NWFSC shelf-slope trawl survey q.

Major uncertainties:

- The catch history is very uncertain.
- The value of q remains uncertain
- The low-q-low-catch and high-q-high-catch sensitivity runs are only indicative of potential biases in the base model; they do not span the full range of uncertainty.
- Ageing is yet to be validated.
- There is uncertainty associated with other fixed and estimated parameters including natural mortality and steepness. The implications of errors in these parameters were not fully explored during the meeting.

Strengths and weaknesses of current approach

The approach suggested by the Panel and adopted by the STAT for the longnose skate assessment can perhaps, in many ways, serve as a template for stock assessments with similar data availability. My main concern with the current assessment is the use of landings data to obtain alternative catch histories. A better approach would be to develop alternative catch histories based on an analysis of historical bottom trawl effort (in conjunction with estimates of longnose catch rates, discard rates, and discard mortality rates).

Merits:

- An acceptable assessment was obtained in what is a very data-poor situation.
- A wide range of uncertainty was explored
- The fixed values of q have an informed basis.

- SS2 was used and as such brings the advantages of a standard and well tested package.

Deficiencies:

- The alternative catch histories were based on landings rather than bottom trawl effort and as such contain much more annual variability than is plausible.
- Ageing is yet to be validated for longnose skate.
- The prior on *q* was not derived under optimal conditions (data and/or expertise that may exist was not available at the time).

Conclusions and Recommendations

Both stock assessments were technically improved by the STAR Panel process. However, the sablefish assessment retains its general deficiencies. The assessment model and data inputs need to be subjected to a full and extensive review. In the longnose skate assessment the use of landings to construct catch histories should be revisited. It would be preferable to use an effort-based approach. In other regards, the longnose skate assessment can perhaps serve as a template for other stock assessments with similar data availability.

I support the research recommendations given in the STAR Panel reports. In brief, my recommendations are:

Sablefish

- A full review of the assessment:
    - Model complexity should be simplified to be compatible with the expected information content of the data.
    - Personnel with specialist experience and skills should critically review each data source.
    - The existing age frequencies (and model fits) should be critically examined to see if cohorts (at relatively young ages) are being tracked reliably.
- The exercise for deriving the prior on *q* should be redone.
    - All potentially relevant data sources should be made available to a selected group of participants with appropriate skills and experience.
    - Ideally, priors would be formed for all of the trawl surveys used in the assessment.
    - The sablefish *q*-priors could be derived at a more general workshop covering several species.
- The apparent link between environmental variables and sablefish recruitment needs to be validated in a full cross validation study (*see* Francis 2006).

Longnose skate

- Under alternative assumptions construct “best”, “low”, and “high” catch histories using:

    - Commercial bottom trawl effort
    - CPUE for longnose skate
    - Discard rate estimates
    - Discard mortality assumptions
- Investigate the AFSC triennial survey time series with regard to “rogue” years for longnose skate and other species (a multi-species analysis of the trawl survey results is required, not another analysis of gear and survey protocols).
- The exercise for deriving the prior on *q* should be redone.
    - All potentially relevant data sources should be made available to a selected group of participants with appropriate skills and experience.
    - Ideally, priors would be formed for all of the trawl surveys used in the assessment.
    - The longnose skate *q*-priors could be derived at a more general workshop covering several species.
- Ageing (validation) studies and maturation rate studies.

References

Sullivan, K.J. et al. 2006. Report from the fishery assessment plenary May 2006: stock assessments and yield estimates. (Draft document available form the N.Z. Ministry of Fisheries.)

Francis, R. I. C. C. 2006. Measuring the strength of environment-recruitment relationships: the importance of including predictor screening within cross-validations. ICES Journal of Marine Science, 63: 594-599.

Appendix 1: Longenose skate catch history multipliers

The main uncertainty associated with the longnose skate assessment is the "catch" history. In the context of "discard" fisheries, the term "catch" is ambiguous. Here it refers to the retained catch plus the discard mortalities.

The Panel recommended to the STAT that three alternative catch histories be constructed: "low", "high", and "mid" or "best. The "best" catch history was constructed from the STAT's estimated landings series, estimated discard rates, and assumed discard mortality. The "low" and "high" catch histories were also constructed from the landings estimates, but used different assumptions with regard to three components of uncertainty: proportion of longnose skate in the total skate landings, the discard rate, and the discard mortality. The catch histories were constructed "outside the model" (i.e., the discard rate in the model was set equal to 0) to allow easy comparisons of assumed catch histories. Another option is to supply the model with alternative landing histories, discard rates, and discard mortalities. However, the inputs would then be very difficult to compare and it would still be necessary to compute the implied catch histories for comparison (so the calculations would have to be done anyway).

To go from an estimated longnose-skate landing (which was derived from a total skate landing) to a longnose-skate catch for alternative parameters is more complex than one might imagine.

For a given year, let,

e = estimated longnose skate landing
b = the proportion of longnose skate in the total skate landing (that was used to get *e* from a total skate landing)
p = proportion of longnose skate in the total skate landing to be applied
d = discard rate to be applied
m = discard mortality rate to be applied.

The catch that "came on-board", assuming the alternative parameters, is:

$$\frac{e}{b}\frac{p}{(1-d)}.$$

The discarded catch that survived is:

$$d\,\frac{ep}{b(1-d)}(1-m).$$

Therefore, the total "extracted" biomass, being the required "catch", is:

$$e\frac{p}{b}\left[1+\frac{dm}{1-d}\right]$$

The above equation was used to derive multipliers for the STAT's estimated landings to provide the three catch histories. Three time periods were identified where the catch history had intrinsically different levels of uncertainty: ≤ 1980, 1981-1994, and ≥ 1995. In the first time period, a constant value of $b = 0.62$ had been used, and there were no observer samples for discard rates during the period. For 1981 onwards there were annual values for b (estimated from observer sampling or from trawl survey results) and there were some observer estimates of discard rate (only one for 1981-1994: $d = 0.93$). The period from 1995 onwards was supported by explicit estimates of longnose skate landings and more than one discard rate estimate (giving a best estimate of $d = 0.53$). The parameter values used in each period, for each catch history, are given below.

	Longnose proportion (≤ 1980)	Discard rate (≤ 1980)	Discard rate (1981-1994)	Discard mortality (all years)
Low	0.50	0.85	0.91	0.3
Mid	0.62	0.93	0.93	0.5
High	0.75	0.97	0.95	0.7

The multipliers for each period, of the (existing) longnose skate landings, to obtain the three catch histories were:

	≤ 1980 $b = 0.62$, apply p, d, m	1981-1994 $p = b$, apply d, m	≥ 1995 $p = b$, $d = 0.53$, apply m
Low	2.18	4.03	1.34
Mid	7.64	7.64	1.56
High	28.59	14.30	1.79

Appendix 2: Construction of informed priors for trawl survey *q*s

The STAR Panel Chair suggested to the STATs (for both sablefish and skate) that it could be beneficial to construct informed priors for the trawl surveys where each of them had fixed q=1. A joint session was held for this since the proposed method was identical for both species. I led this session given my extensive experience in New Zealand at eliciting informed priors for trawl (and acoustic) survey *q*s using expert opinion and available data.

The general approach described below has been used in New Zealand for several years in one form or another (e.g., see hoki and orange roughy assessments in the 2006 Plenary report, Sullivan et al. 2006).

The approach requires that the trawl survey q is split into three components: areal availability (the proportion of stock biomass in the trawl survey area), vertical availability (the proportion of biomass in the water column that is available to the trawl after vertical herding), and vulnerability (the proportion of biomass between the wings (assuming wing-spread estimates) that is retained in the cod-end). During discussions, areal availability was split into two components: depth and latitude (essentially being what proportion of biomass was south of the southern survey-area boundary).

Sablefish

Discussions were held on each of the four components for sablefish, with regard to what was thought to be fully selected fish (being about 53 cm long and perhaps 3-6 years old). The objective with regard to each component was to agree a “lower bound”, an “upper bound”, and a “best guess”. By default, the best guess was the mid-point of the bounds. It was noted that data were available to help with some components (e.g. proportion of biomass south of Point Conception) and finalization of the bounds and best guesses were delayed until the data became available.

The final bounds and best guesses for each component were:

	Depth	Latitude	Vertical av.	Vulnerability
Low	0.85	0.82	0.8	0.4
High	0.98	0.88	1.0	1.0
Best	Mid point	0.85	Mid point	0.8

NWFSC slope trawl survey data from 2003-2006 were used to determine the latitude values. Other values were chosen by consensus (in particular, for the bounds, on the basis that everyone was willing to accept that the “true” value was within the specified bounds).

The consequent bounds on q and the best guess are: (0.22, 0.86) and 0.56. The best guess was equated to the median of a lognormal distribution and the bounds to 99% of that distribution. This gave a normal prior on log (q): mean = -0.58, sd = 0.184.

The normal prior on log (q) was subsequently used to provide three qs for model runs with nominal weights of 25%, 50%, and 25%. A random sample of size 10,000 was generated from the normal distribution and the mean of the samples below the 25th percentile (of the normal distribution) was exponentiated to provide the “low q”. Similarly, the mean of the samples above the 75th percentile was exponentiated to provide the “high q”. The median of the prior was used in the base model.

The low, base, and high qs were: 0.445, 0.560, 0.712.

Longnose skate

Similar discussions were held for skate with regard to the NWFSC shelf-slope survey. Areal coverage was thought to be very good with just a small allowance made for some fully selected skate that were perhaps shallower than the trawl survey. Similarly, vertical availability was considered to be high, although some skate would clearly go under the ground rope. The largest bounds were place on vulnerability. It is known that flatfish can be herded by trawl gear, and it is possible that this could also occur for skate; or perhaps not. “Best guess” estimates were set at the mid-point of the range for individual factors, except vulnerability which was given a best guess of 1.

	Depth	Latitude	Vertical av.	Vulnerability
Low	0.95	1.00	0.75	0.75
High	1.00	1.00	0.95	1.50
Best	Mid point	1.00	Mid point	1.00

The consequent bounds on *q* and the best guess are: (0.53, 1.43) and 0.83. These values were used to construct a lognormal prior using the same method as for sablefish. This gave a normal prior on log (q): mean = -0.19, sd = 0.187.

Similarly, fixed *q* values were determined for runs with the nominal 25%, 50%, and 25% weights. The low, base, and high qs were: 0.654, 0.830, 1.046.

Appendix 3: Document distributed to STATs yet to present

The following document was distributed to STATs after the first meeting in an attempt to pre-empt some issues which had the potential to repeatedly arise at future meetings. The attempted humor in “E.” and “F” was not without serious intent. While the phrases given in “E.” should never be provided as justifications for decisions made during an assessment, they were all heard during the first meeting.

Suggestions for STATs (1)

Patrick Cordue
16 May 2007

As the person with the “honor” of serving on all of the STAR Panels this year I thought it would be useful to provide some suggestions to STATs which are yet to present their assessments to STAR Panels. The first STAR Panel meeting has just concluded. Below are some lessons learned from this first 2007 meeting and some general advice to STATs which has been gleaned from more reviews than I care to remember. The following are merely suggestions – many special cases arise for individual assessments and STATs may have good reason for taking somewhat different approaches.

A. Iterative re-weighting

It is generally agreed “in theory”, for a model run, that it is desirable for the “input variance assumptions” to be consistent with resultant residuals. Various methods of iterative re-weighting of data sets are available to achieve this outcome. While there is no agreement as to which method is best, STATs should attempt to obtain “consistent” effective sample sizes for age and length frequencies and “consistent” CVs for biomass indices (but only for biomass time series of “sufficient length”, perhaps, at least 10 points). Balancing of the input variance assumptions with the residuals is not the “be all and end all” – but it does provide a good starting point. Subsequent re-weighting is perhaps best done by down-weighting “bad” data sets, rather than up-weighting “good” data sets.

It is recommended to use the built-in SS2 output to do the iterative re-weighting. For each age or length frequency data set, the mean of the ratios of effective sample sizes to input sample sizes is output for each run. Both the arithmetic and harmonic means are provided as output. The user could also calculate the ratio of means (rather than the mean of ratios). In any case, however they are calculated, the sample size scalars should be entered into the “variance adjustments” section of the control file until scalars on successive runs are relatively constant for each data set. For biomass indices, a similar adjustment is made, but the adjustment is additive (and is the difference between the input and output RMSE values). I am aware that there is an option in SS2 to estimate “process error” for biomass indices, but I am told that it has not been fully tested for effectiveness.

For age and length frequencies the suggested iteration procedure may produce a “flip-flop”, being an alternation between two different vectors of sample-size scalars. If this happens, there are at least two choices: continue iterating for ever, or use the average of the two vectors (flip-flopping may continue to occur, but eventually the magnitude of the flip-flop will be inconsequential).

Methods which operate at an annual level within data sets and may subsequently change the relative proportion of samples across years are not considered appropriate (e.g., if the sample sizes were (1000, 500, 100) and re-weighting gave effective sample sizes of (100, 100, 100) there would be cause for concern).

B. SS2 minimizer and verbosity setting

The main item of interest output to the screen during a minimization is the derivative table – one needs to ensure that the maximum gradient is very small when the minimizer stops – and it is often instructive to see which parameters are causing a problem and/or if any parameters always have zero derivatives. It is therefore suggested that a verbosity setting of 0 or 1 is used (so that derivative tables are not lost in all the other output)

Also note, just because a positive definite Hessian is achieved does not mean that a local minimum has been found – let alone a global minimum. Always take note of the total likelihood and check that convergence has occurred by using different starting points (there is an SS2 “jitter” feature).

C. SS2 safe/optimized version

There are two main options for compiling ADMB models and hence also two main options for compiling SS2: a “safe” option (debug mode, where run-time errors such as exceeding array bounds are checked for); and an “optimized” option.

It appears that both “safe” and “optimized’ versions of SS2 have been distributed. STATs should ensure that during development of their models that they use a “safe” version. “Production” runs can, of course, be done with the optimized code.

D. Displaying output

The Excel and R interfaces to the SS2 files are very useful. STATs should be able to use both interfaces. STAR Panels may prefer that R is used to display the fits to indices. The main function ss2.output() should be saved in an R workspace so that it doesn't have to be read into R each time.

And, not so seriously:

E. You know the STAR Panel is giving you a hard time when…

In desperation, you say one of the following,

1. But we have always done it that way
2. Previous STAR Panels accepted it
3. It gives a good result

F. You know that a STAR Panel member is a bit tired when…

1. They mistake a seagull's tail feathers for a penguin's head (tired *and* short-sighted!)
2. They forget your name and/or your stock assessment species and/or what day of the week it is and/or what city they are in.

Appendix 4: Bibliography of supplied material

I. Current Draft Stock Assessments

A. Status of the Sablefish Resource off the Continental U.S. Pacific Coast in 2007, Michael J. Schirripa. *DRAFT.*

B. Status of the Longnose Skate (*Raja rhina*) off the continental US Pacific Coast in 2007, Vladlena V. Gertseva and Michael J. Schirripa, *DRAFT.*

II. Background Materials

A. 2006 Workshop Summary Reports

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

B. Previous Sablefish Assessment and STAR Panel Report

1. Status of the Sablefish Resource off the Continental U.S. Pacific Coasts in 2005. Michael J. Schirripa and J.J Colbert, November 1, 2005.
2. Sablefish STAR Panel Meeting Report, June 20-24, 2005.
3. Summary Minutes from the Scientific and Statistical Committee September 19-21, 2005 Meeting in Portland, Oregon. Sablefish assessment discussed on page 9.

C. Manuscripts

1. Kosro, P. Michael, William T. Peterson, Barbara M. Hickey, R. Kipp Shearman, and Stephen D. Pierce. 2006. Physical versus biological spring transition: 2005. Geophysical Research Letters, Vol. 33.
2. Mackas, D.L., W. T. Peterson, M. D. Ohman, and B. E. Lavaniegos. 2006. Zooplankton anomalies in the California Current system before and during the warm ocean conditions of 2005. Geophysical Research Letters, Vol. 33.
3. Schirripa, M.J. and J.J. Colbert. 2006. Interannual changes in sablefish (*Anoplopoma fimbria*) recruitment in relation to oceanographic conditions within the California Current System. Fisheries Oceanography 15:1, 25-36, 2006.

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

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

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

III. Meeting Materials

A. Draft Agenda
B. Meeting Location Information
C. HMSC Driving Directions & Map
D. Panel Participants

Appendix 5: Statement of work

Consulting Agreement between the University of Miami and Patrick Cordue

Statement of Work

April 21, 2007

General

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

The stock assessments will report the status of the longnose skate and sablefish 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 sablefish and longnose skate 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 assessment models.

The Pacific Fishery Management Council's Scientific and Statistical Committee requests that "all review panelists should be experienced stock assessment scientists, i.e., individuals who have done actual stock assessments using current methods. Panelists should be knowledgeable about the specific modeling approaches being reviewed, which in most cases will be statistical age- and/or length-structured assessment models" (SSC's Terms of Reference for Stock Assessments and STAR Panel Process for 2007-2008)

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

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

Specifics

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

The consultant's tasks consist of the following:

1) Become familiar with the draft stock assessments and background materials.
2) Actively participate in the STAR Panel to be held in Newport, Oregon from May 7-11, 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 May 25, 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 June 8, 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.