

Pacific Whiting

The Joint U.S.-Canada STAR Panel Report

Pacific Fishery Management Council
Hotel Deca
Seattle, Washington
February 8-10, 2010

Review Panel Members:

Vidar Wespestad (Chair), SSC representative
Geoff Tingley, Center for Independent Experts
Patrick Cordue, Center for Independent Experts
Tom Carruthers, University of British Columbia

Stock Assessment Team (STAT) Members Present:

Ian Stewart, Northwest Fisheries Science Center
Owen Hamel, Northwest Fisheries Science Center
Steve Martell, University of British Columbia

Advisors:

Jason Cope, GMT representative
Tom Libby, GAP representative
Robyn Forrest, Department of Fisheries and Oceans, British Columbia
Chris Grandin, Department of Fisheries and Oceans, British Columbia
Greg Workman, Department of Fisheries and Oceans, British Columbia
John DeVore, PFMC representative

Overview

During 8-10 February 2010, a joint Canada-U.S. Pacific hake / whiting Stock Assessment Review (STAR) Panel met in Seattle, Washington, to review two draft stock assessment documents that had been prepared by Stewart & Hamel (2010) and Martell (2010). The Panel operated under the U.S. Pacific Fishery Management Council's Terms of Reference for the Groundfish Stock Assessment and Review Process for 2009-2010 (PFMC 2008). As in previous years, the Panel attempted to adhere to the spirit of the Canada-U.S. Treaty on Pacific hake / whiting, with the Panel including a member from Canada. The revised stock assessments and the STAR Panel Report will be forwarded to the Pacific Fishery Management Council (PFMC) and its advisory groups, and to the Canadian Department of Fisheries and Oceans (DFO) managers and the Groundfish Sub-committee of PSARC (Pacific Scientific Advice Review Committee).

The Panel convened at 9AM Monday, February 8, 2010 with a welcome from the Chairman and a round of introductions. Mr. John DeVore of the Pacific Fisheries Management Council opened the meeting with an overview of the STAR process and reviewed the terms of reference. The agenda was reviewed and finalized for the duration of the STAR panel.

After the opening proceedings the STAR panel received an overview of the 2009 hake/whiting fisheries in Canadian waters from Chris Grandin, DFO and from Ian Stewart, NMFS Northwest Fisheries Science Center (NWFSC) on the 2009 fishery in U.S. waters. The presence of Humboldt squid was noted in the 2009 fishery, primarily in northern Washington and southern BC. Squid were not a problem in the fishery off Oregon as reported by fishermen present at the STAR panel.

Following the presentation of fisheries data, Dr. Dezhong Chu of the NWFSC Acoustic staff provided the STAR with an overview of the methodology of acoustic stock estimation techniques and the manner of constructing stock estimates from acoustic signals. This was followed by a presentation of the 2009 survey results by Dr. Rebecca Thomas, also a member of the NWFSC acoustics group. The survey results were the whiting/hake stock in 2009 was estimated to be 1.462 million metric tons (mmt), up from 0.879 mmt in 2007. Dr. Thomas noted that the majority of the stock was located in US waters in 2009 and fish were nearly continuously distributed from California to mid Vancouver Island. The presence of Humboldt squid was noted in the acoustic survey, and their relative abundance/biomass was much higher than that from the 2007 survey.

The presence of squid in 2009 was problematic. Echograms suggested that squid caused the distribution of hake to be altered on survey transects where they were present. In these situations, it appeared that hake were schooled near the bottom with a mixed hake/squid layer above it. Sampling by trawl and camera indicated variability in the distribution of squid and hake in the distributional strata of transects where squid were present making it difficult to determine the actual species mix. On 44 transects of the total of 77 transects where hake were present, there is higher confidence of hake identification, although there may have been some squid on these transects. The estimated biomass from these transects was 0.87 mmt. A large part of the STAR panel discussion focused on the acoustic survey in regards to squid in the 2009 survey and the adequacy and accuracy of acoustic survey trawl sampling of observed echo sign (marks) for length frequency and species composition of the sign/marks.

The panel also received a presentation by Dr. Chu on the ongoing analysis of the time series of the acoustic survey. Currently the data reside in various locations and data forms that preclude direct analysis of all of the data collected since the late 1970s. Dr. Chu's group has an on-going project to prepare a standardized database of the entire data set. The STAR panel agrees with Dr. Chu that this is an important task and supports his efforts to improve the acoustic data base for the U.S.-Canadian hake stock.

The remainder of the first day of the meeting was taken up by an overview of the data sources used for the 2010 whiting/hake assessment by Drs. Ian Stewart and Owen Hamel of the U.S. STAT Team. On the second day, the STAR resumed with presentations of the results of the two stock assessment models under review. Ian Stewart of the U.S. STAT presented the Stock Synthesis model description and results, which was followed by a presentation of the TINSS model description and results by Dr. Steve Martell of the Canadian STAT.

The STAR Panel focused their attention on the modeling approaches and the treatment of data and explored model sensitivity. Secondly, the STAR Panel explored the influence of the 2009

acoustic survey data and age compositions on model results. This was to examine the potential bias of the presence of large numbers of Humboldt squid on results, and in general the question of adequacy of sampling of echo-sign for length-age composition.

The STAR Panel progressed in three stages through the base models and the underlying data. This process began by examining the underlying structure of the base models and then moved toward an examination of the fishery and survey data used in the base model. From there the STAR Panel formulated different configurations of the two base models until arriving at models with defensible input data and minimal complexity.

The preliminary base models provided by the STAT teams were not considered acceptable by the STAR Panel primarily because of data issues. The STAR Panel's preliminary preferred runs specified a number of changes to the input data:

- Remove all acoustic age and length frequency data
- Remove 1986 and 2009 acoustic biomass estimates
- Split acoustic time series into two parts (separate qs): 1977-1992, 1995-2007
- SS3: Remove length frequencies and conditional age-at-length; replace with age frequencies

The acoustic composition data were omitted because it was considered extremely unlikely that the opportunistic sampling of hake layers with mid-water trawl gear could provide a consistent time series (i.e., with a constant selectivity across years). The 2009 biomass estimate was clearly compromised by the significant presence of squid and was not comparable to earlier survey estimates. Also, a smaller survey area was covered in the 1977-1992 surveys and although "expansion factors" had been applied to the survey estimates it was prudent to split the time series into two components. The 1986 survey estimate was potentially biased as the pre- and post-survey calibrations were substantially different. Finally, efforts to fit the length frequencies and conditional age-at-length data in the SS3 model had been less than successful with very poor residual patterns for the length data. There was also the technical issue that fish were growing during the fishing season and this would potentially compromise the use of the conditional age-at-length data.

The STAR Panel considered that the SS3 and TINSS models were equally acceptable to provide a base model run. However, the full MCMC run was only available for the TINSS model. It is primarily for this reason that the Panel adopted the TINSS model, with the Panel's preferred data specification, as the base model.

The final preferred base model was the TINSS model:

- acoustic biomass indices split into two time series: 1977-1992; 1995-2007 (the 1986 and 2009 indices are omitted, as are all composition data); standard deviation in log space assumed constant within each time series: 0.5 and 0.25 respectively
- commercial age frequencies (single fishery; US and Canadian data combined)

The point estimate of 2010 depletion is 37% with a projected OY of 339,000 mt (based on the 40-10 rule using estimated F_{MSY} rather than the proxy of $F_{40\%}$). These are "risk neutral" estimates being the medians of the marginal posterior distributions. A decision table, with alternative catch streams and three states of nature, can be constructed from the base MCMC run

using the central 50% of the posterior distribution, and the two tails each containing 25% probability.

Summary of data and assessment models

The STAR Panel was provided with five basic components to enable and support the review. These were:

- (i) background documentation;
- (ii) overviews of the US and Canadian fisheries;
- (iii) information and data relating to the 2009 acoustic survey;
- (iv) full details of the base Stock Synthesis 3 (SS3) stock assessment model (Stewart and Hamel, 2010); and
- (v) full details of the base TINSS stock assessment model (Martell, 2010).

In addition to the written papers, presentations on the acoustic survey, SS3 model and outputs, and the TINSS model and outputs were made to the Panel.

The quantity and coverage of the background material was adequate and provided in a timely manner in advance of the meeting together with the draft SS3 and TINSS stock assessments.

The only subject not covered by the material made available in advance of the meeting was information about the 2009 acoustic survey. Typically this would not necessarily be required; however, due to the presence of large numbers of squid and the impact of this on the hake biomass estimate, the STAR Panel would have benefited from an earlier awareness of the survey and these associated issues.

The quality of the written material was high and largely comprehensive and all participants fully supported the Panel members in their understanding of the techniques, results and caveats, freely answering calls for clarification and for additional supporting detail. Those responsible for producing and presenting the material are to be congratulated.

Issues of note from the fisheries include the failure to catch the OY in some years due to reaching bycatch limits, the inefficiencies exhibited by new entrants to the fishery, and some market difficulties leading to a temporary reduction in effort. The presence and extent of the Humboldt squid, *Dosidicus gigas*, radiation and their influence on the fishery in 2009 was also noted.

Within the assessment were descriptions of some alternative survey-based approaches to develop hake biomass indices none of which had proved satisfactory.

Suggestions for future reviews of Pacific hake / whiting assessments.

When it is fully implemented, the Pacific Hake / Whiting Agreement between the U.S. and Canada will establish a process for developing and reviewing stock assessments and providing management advice for this important transboundary stock. Given the definite possibility that

the assessment review next year (2010) may again operate under the STAR Terms of Reference, the PFMC's Scientific and Statistical Committee (SSC) should consider altering the STAR Terms of Reference to better accommodate alternative stock assessments developed by Canadian scientists.

The process for future assessments of Pacific hake should ensure that the STAT has adequate time to conduct the assessment. Late arrival of data and a compressed schedule to resolve the assessment can result in a rushed assessment that can lead to incorrect results. A different assessment and review process is needed given the expectation that this situation will re-occur with late-season fishing in both countries. For example, a partial release of catch quota could be made to accommodate the early season, with a later release based on a new assessment that is completed in March or April.

Simulation evaluation

A simple but useful exercise when proposing a relatively complex assessment model is to simulate data from the same model using known parameter values. The model can then be refitted to these data to evaluate whether such parameters can be reliably estimated ('search for parsimony': TOR Appendix B, Section D 4 a). It is unlikely that there is sufficient information to separately estimate the parameters of the SS3 base case model reliably, particularly if subject to realistic levels of observation error (to an extent this is supported by the literature; e.g. Thompson 1994). An advantage of the TINSS model was the ability to demonstrate that parameters could be accurately estimated from simulated data.

Requests by the STAR Panel and Responses by the STAT

The first set of requests was to the US STAT team and applied to their preliminary base model.

1. A plot of MCMC posterior parameter correlation among: descending limb selectivity parameters, natural mortality rate, senescence, time varying growth parameters, B_0 .

Rationale: To better understand parameter confounding and the surface of the joint posterior.

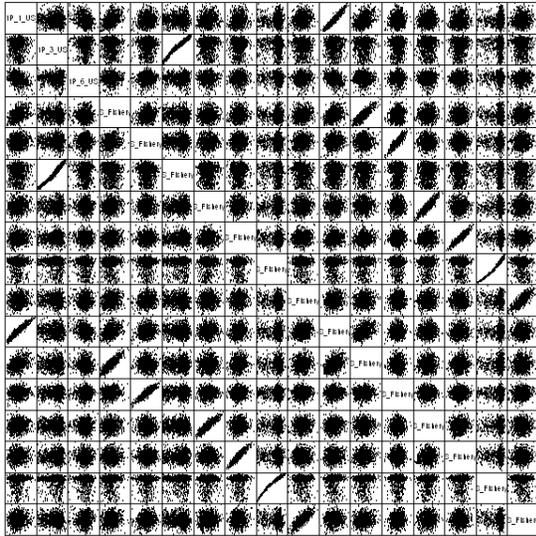


Figure 1.

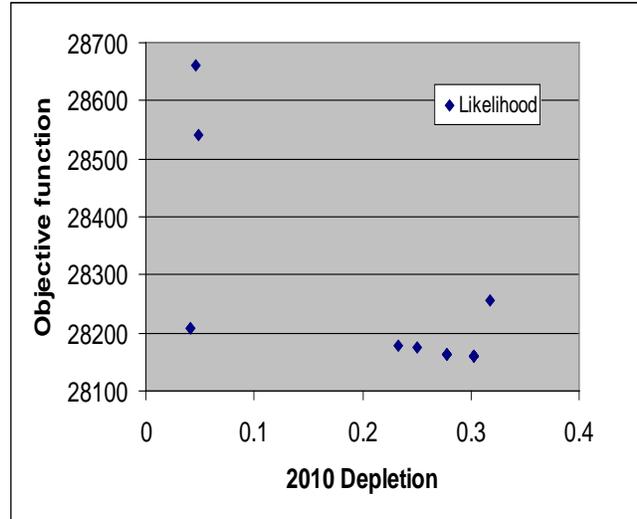


Figure 2.

Since they are related issues, a discussion of this STAT response is included below in the response to request 2.

2. Jittered runs (multiple random starting values) of the MLE run returning the normalized depletion estimates of a number of runs (10+). For each run could the STAT team also report the value of the objective function and the number of runs that did not satisfy convergence criteria.

Rational: This was requested to understand whether SS3 can reliably arrive at similar model predictions from different initial values

Response: The STAT team produced the plots displayed above.

Discussion: In general the SS3 assessment model appears overly complex leading to undesirable model properties. The SS assessment seeks to simultaneously estimate time varying growth, senescence, descending limb parameters of several dome shaped selectivities, initial biomass and recruitment anomalies that are all to some extent theoretically confounded (e.g. Thompson, 1994). This is illustrated by Figure 1 above that describes strong parameter cross correlation and redundancy in model complexity.

The poorly defined objective surface is reflected by the inability of AUTODIFF to converge reliably. In 29 different jittered runs the base case SS3 model did not converge in 10 and found subjectively ‘unrealistic’ values in all but 5 of the remaining model runs. Of the five model runs that were deemed credible, the range of estimated depletions ranged from around 2.3 to 3.2 (Figure 2). This large range is concerning since it implies that the estimation method cannot consistently find a global minimum (this is relevant to the TOR Appendix B Section D 4 f). It follows that many MLE runs must be undertaken before one can have confidence over whether a true global minimum as been found. This raises the issue of whether individual sensitivity runs are representative of genuine sensitivities or simply the product of poor convergence from a particular set of starting values.

These concerns was less relevant to the TINSS model since most sensitivity analyses were conducted on the MCMC run (we had assurance from the Canadian STAT team that the TINSS MLE model converged on the same parameter values and management recommendations irrespective of starting values).

Historically the AIC model selection criterion has been used to select the SS3 base case model structure. Given the parameter redundancy highlighted above, the current base case may not be a suitable starting point for the search for a parsimonious model.

3. Model runs with the removal of the 2009 acoustic survey data for comparison with base case and other runs. Biomass out, composition data in; or all data out.

Rational: To further examine the sensitivity of estimated depletion to the weight on the 2009 survey data (removal being equivalent to an infinite CV).

Response: Estimated 2010 depletion was reduced from 31% (base model, CV=0.5) to 25% with the exclusion of the 2009 biomass index. Including or excluding the 2009 composition data made very little difference. There was an existing run with a CV of 0.25 on the 2009 index which had depletion estimated at 43%.

Discussion: This emphasized the importance of the weight given to the 2009 biomass index – a point that had already been noted by both STAT teams.

4. Provide details of the scaling procedures used to produce commercial length frequencies and conditional age at length.

Rational: Clarification was required due to some slight ambiguities in the documentation.

Response: The relevant equations were presented. Samples were scaled-up by number within each haul/landing, and then scaled by catch weight within fleet (shore-based or at-sea).

Discussion: The scaling at each stage should be by number. Also, spatial and temporal strata, and perhaps finer scale fleet strata, should be used. (See Research Recommendations.)

5. Provide expansion factors applied to the acoustic estimates in each year.

Rational: To confirm the years in which indices had been adjusted and by how much.

Response: Extracts of documents were presented detailing the history of adjustments which had been made to the indices from 1977 to 1992 inclusive. There had been an initial adjustment for a change in assumed target strength (from -35 dB per kg to the

Traynor (1996) relationship), combined with an adjustment for area differences (Dorn et al. 1996). The area expansion factors ranged from 1.47 to 1.78. Adjustments were later revised using a more complex method (Helser et al. 2004).

Discussion: Substantial expansion factors were applied to the 1977 to 1992 indices. This suggests that the acoustic time series does not have a consistent proportionality constant (q) across all years, and indicates that it should be modeled as two time series with separate q s.

6. Propose the number of q s to use and the years to which each q applies.

Rational: The simple split indicated by which surveys had been expanded may not be the only basis by which to assign surveys to alternate q s.

Response: The STAT team declined to offer a suggestion.

Discussion: This was an optimistic request – it would require a very detailed analysis of spatial distributions and other factors to arrive at an alternative method of splitting the time series.

7. Likelihood profile across R_0 for all likelihood components including penalties.

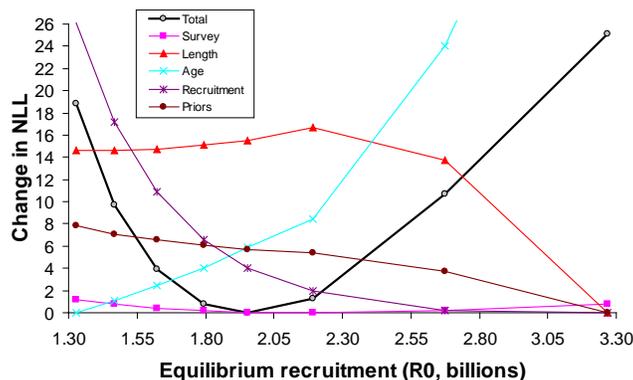
Rational: The examination of likelihood profiles across R_0 for individual components reveals which data are consistent with low or high biomass and shows the contribution that each data type makes to the total likelihood.

Response: The main contributions to the total likelihood came from the age data and the *penalties* on recruitment deviations. R_0 was determined by a trade-off between these two components, with the age data preferring low biomass which was associated with high penalties on the recruitment deviations.

Discussion: It is a common feature of these types of models that the survey biomass index makes little relative contribution to the total likelihood which is dominated by age/length data. It is not so common to see a relatively arbitrary penalty function playing such a central role in determining the estimate of R_0 . This needs further investigation (see Research Recommendations).

8. Likelihood profile all likelihood including penalties.

Rational: As



across q (0.6-1.3) for components

for the previous

request, but looking at another parameter of interest.

Response/Discussion: The survey biomass indices were central to the estimate of q as would be expected given that R_0 was largely determined by the other data (and penalties).

Given the sensitivity of assessment results to the weight placed on the 2009 acoustic biomass index a request was made to the acoustic team to provide a further summary of data with regard to the squid problem:

Acoustics request: 2009 acoustic survey, how extensive were the squid: spatial distribution; number of transects with identified squid or mixed echo sign; biomass estimates split by mark type (hake, mixed hake & squid)

Rational: A decision needed to be taken on whether the 2009 acoustic index was comparable to other points in the recent time series.

Response: Preliminary information on the spatial distribution of assigned squid backscatter was presented and compared with the assigned hake backscatter. There was clearly a large spatial overlap in the distributions. The “rule of tentacle” which had been used to assign backscatter to hake, when there was a potential mixing of squid and hake, was used on 33 out of 77 transects in US waters. These transects contributed 41% of the total estimated hake biomass. The biomass associated with “definite” hake marks was 870,000 t (compared to the total estimate of 1,470,000 t)

Discussion: There is clear potential for large bias in the 2009 index because of misclassification of acoustic layers and marks. Also, the relative target strengths of squid and hake are very uncertain, as are their relative selectivities to the trawl gear, which makes the partitioning of backscatter between the two species on the basis of trawl catches very problematic.

The second set of requests were for both STAT teams with regard to their preliminary base models

1. Runs with 1986 acoustic survey in or out; runs with alternative calibration used.

Rational: In 1986, the pre-survey and post-survey calibrations differed by 1.7 dB (a factor of about 1.5). As it has an indeterminate bias, the index should be excluded from assessment runs, but the Panel wanted to know if its inclusion made any difference to the results.

Response/discussion: There was little difference to the results in either model.

2. Runs with no acoustic data (compositional data and biomass removed for all years).

Rational: To determine if the results were sensitive to the inclusion of the acoustic survey data.

Response/discussion: The exclusion of the data made little difference to the SS3 model results (as could be expected by the results of the likelihood profile on R_0 – see above). However, in the TINSS model the exclusion of the data resulted in lower estimated spawning biomass and a dramatic reduction in the depletion estimate (about 25% compared to 61% in the base model).

3. Runs with fishery selectivities moved further to the left (informed prior on age at 50% selectivity: weight at 3 years)

Rational: In the SS3 model, estimated selectivities seem somewhat unlikely with full selection not occurring until 8-10 years of age. It was wondered if similar fits to the data could be achieved with full selection at younger ages. In the TINSS model full selection was already further to the left at about 6 years.

Response/discussion: The informed priors were over-ridden by the data and results were little changed. This issue could do with further investigation.

For US STAT team:

4. Summary of commercial catch by season within year (by sector if possible).

Rational: There was concern that the growth of fish during the fishing season could compromise the use of the conditional age-at-length data (in the SS3 model).

Response/discussion: Data provided by the US STAT team and from other sources suggested that this was an area of concern, particularly, but not exclusively, in the last three years. It is primarily an issue for the US data as younger fish, which grow faster, are caught in this fishery.

For Canadian STAT team:

5. Implied prior on derived variables e.g., depletion

Rational: To understand the prior inference regarding management reference points.

Response/discussion: Removing the likelihood function reveals that prior assumptions lead to less pessimistic depletion estimates and that the model is not updated strongly by inclusion of the data.

6. MCMC integration results for the three parameter Thompson selectivity curve (posterior density plots for M and the gamma parameter).

Rational: The dome shaped Thompson selectivity curve seeks to estimate a third parameter that controls the slope of the descending limb. This request served to investigate whether dome shaped selectivity and lower natural mortality rate could provide a better explanation of the data.

Response/discussion: Parameters estimates were in keeping with asymptotic selectivity and high natural mortality rate. Dome shaped selectivity was not supported by the TINSS model fitted to the aggregated Canadian and US commercial data.

The third set of requests were for both STAT teams:

1. Model run for:

STAR Panel's preliminary preferred run:

- Remove all acoustic age and length frequency data
- Remove 1986 and 2009 acoustic biomass estimates
- Split acoustic time series into two parts (separate qs): 1977-1992, 1995-2007
- SS3: Remove length frequencies and conditional age-at-length; replace with age frequencies
- SS3: Use MPD growth estimates from preliminary base model
- Acoustic selectivity: e.g., 50% 2+ biomass; and try alternative assumptions: 20% at age 2, 80% 3+;

Time permitting: explore sensitivities (e.g., acoustic age and length frequencies included; SS3: low M, high M; TINSS: alternative priors on C*, F*, high and low central tendency, high and low variability).

TINSS: MCMC run for preliminary preferred model with diagnostics (time permitting).

Rational: The STAR Panel wished to see results for runs from each model which used only fully defensible data. Sensitivities were requested to explore dimensions of uncertainty appropriate to each model. MCMC results were not requested for the SS3 model as they would not have been available in time (this model can takes days to produce MCMC results).

Response/discussion:

In the TINSS model, in comparison with the original base, biomass estimates were reduced and the 2010 depletion estimate was much lower (34%, compared to 61%). There was strong sensitivity to assumed priors with regard to OY, especially in terms of the assumed variance of the MSY prior; lower variance removing the long tail resulting in a much lower point estimate (being the median of the posterior).

The US STAT team presented results for the requested run but also offered an alternative run on which they based their sensitivity runs. Their variation was to estimate a single M over all ages rather than fix M at 0.23 and estimate a senescence value (for 14-15+). For both runs, in comparison to their base model, much higher biomass was estimated (about a factor of 2.5) but with similar, although higher, 2010 depletion: 40% compared to the base model estimate of 31%.

2. Runs with the “minimum” 2009 hake estimate (0.87mmt): preliminary preferred model; STAT team preliminary base models.

Rational: To determine the effect of including a “confident”, yet probably biased low, estimate of hake biomass for the 2009 survey.

Response/discussion: The inclusion of the estimate had little effect in any of the runs. However, the direction of the changes were interesting for the SS3 models, with the inclusion of the estimate in the STAR Panel preferred model giving a lower estimate of 2010 female spawning biomass, and its inclusion in the STAT teams preliminary base model giving a small increase.

The fourth set of requests called for a slight change to the STAR Panel’s preferred models and requested model specific sensitivity runs:

STAR Panel preferred base models: as in request 3 with higher CVs on earlier acoustic times series (e.g., .5 for 1977-1992; 0.25 for 1995-2007). SS3: $M=0.23$, Canadian fishery selectivity asymptotic.

Sensitivity runs to preferred base models:

SS3: Two bracketing runs: estimate M with asymptotic Canadian fishery selectivity; estimate M with free selectivities; alternative runs as necessary to fully bracket uncertainty (e.g., in depletion and OY)

TINSS: Two bracketing runs using alternative priors for F^* and/or C^* changing the median values and/or variances.

SS3: a plot of the STAT team preliminary base model posterior and prior density with regards to the slope of the fishery selectivity descending limb parameter (s) including the MLE estimate (in the inverse logit space 0-1).

SS3: likelihood profile for individual components across R_0 for the STAR Panel preferred model

Rational: As pointed out by the US STAT team, the earlier acoustic time series was clearly more uncertain than the recent series (so higher CVs for the earlier series) The dimension of uncertainty for SS3 was chosen to be M and the trade-off between M and domed or asymptotic fishery selectivities. The US fishery doesn’t have access to all of

the older fish, so the US fishery selectivity was allowed to remain domed in all runs. The other SS3 requests concern technical issues which lead to research recommendations. The dimension of uncertainty for the TINSS model were the two crucial assumed priors.

Response/discussion:

In the SS3 model the change to asymptotic selectivity in the Canadian fishery resulted in a large decline in estimated biomass, back to a level similar to that of the STAT team’s preliminary base model. Also, estimated 2010 depletion was somewhat lower at 32% (compared to 40% in the STAR Panel’s preliminary preferred model) which was almost the same as the STAT team’s preliminary base model estimate of 31%. The range of estimated 2010 depletion for sensitivity runs was 15-42%.

The slight modifications to the TINSS model made little difference to the results. The sensitivity runs also showed the same effects: changing the priors had little effect on estimated biomass or 2010 depletion (range 33-38%), but dramatically affected estimated OY (range 250,000-400,000 t).

The final request to the STAT teams was for a summary table of estimates across a number of runs:

1. Summary table for three runs: STAR Panel preferred runs, SS3 and TINSS; SS3 preliminary base model.

Include estimates for: 2010 depletion, OY, acoustic qs , steepness (h), M .

SS3: MPD estimates

TINSS: Median of posterior; MPD (mode of posterior)

Response: The table is given below.

	TINSS	TINSS	SS3	SS3 (update)
Metric	MPD	Median	MPD	MPD
2010 Depletion	29%	37%	32%	31%
2010 OY mt	220.000	339.000	235.000	225.000
qs	0.454/0.467	0.39	0.59/0.68	0.94
h	0.538	0.519	0.86	0.88
M	0.273	0.286	0.23	0.23/0.62

Discussion: A great deal of care is needed in interpreting this table. The OYs for the TINSS run are based on F_{MSY} , but the SS3 runs use $F_{40\%}$. If OY was estimated based on $F_{40\%}$ in the TINSS run, the estimates would be much higher; $F_{40\%}$ is a very aggressive policy in the TINSS parameter space and is not a good proxy for F_{MSY} . Also, the acoustic qs are not comparable

between the SS3 update run (being the STAT team's preliminary base model) and the other runs. The update-run has a domed selectivity for the acoustic survey whereas the other runs assume 50% selection at age 2 and 100% selection at ages 3 and older. A single q is given for the TINSS median; there are actually two qs but they are not very different. The second M for the update run is for senescence.

Description of base model and alternative models used to bracket uncertainty

The STAR Panel considered that the SS3 and TINSS models were equally acceptable to provide a base model run. However, the full MCMC run was only available for the TINSS model. It is primarily for this reason that the Panel adopted the TINSS model, with the Panel's preferred data specification, as the base model.

TINSS model:

- acoustic biomass indices split into two time series: 1977-1992; 1995-2007 (the 1986 and 2009 indices are omitted, as are all composition data); sd in log space assumed constant within each time series: 0.5 and 0.25 respectively
- commercial age frequencies (single fishery; US and Canadian data combined)

Alternative models used to bracket uncertainty.

The full MCMC run for the base model is used to describe and bracket uncertainty. This is achieved by basing the decision table on the central 50% of the posterior distribution and the left and right hand tails (each holding 25% of the posterior distribution).

This single model does not encompass the full range of uncertainty, but it does, by itself, describe such an uncertain assessment of the status of the stock that the addition of further uncertainty would not be useful to managers.

Comments on the assessments

Comments on the data

There may be useful information within the commercial catch and effort data that are currently unused in these assessments.

The acoustic survey is currently generating the only usable biomass index to support the assessment.

Preliminary base runs

Neither of the preliminary base runs presented by STAT teams were considered acceptable by the STAR Panel. There were a number of serious problems with the input data in both models.

The most serious data problems were with the treatment and use of the acoustic survey data. It became clear during the meeting that the 2009 survey results were badly compromised by the unusual occurrence of large quantities of squid in the survey area. The biomass index for 2009 was clearly not comparable with other points in the time series (and had to be omitted in later runs). Also, use of the survey compositional data assumes that the samples from trawl catches provide representative length/age samples of the hake that are vulnerable to the acoustic beam.

There is a selectivity pattern associated with the trawl gear . There is a different selectivity pattern associated with the acoustic beam (that is not comparable to the catch data as used in the base models). However, within the assessment models the trawl selectivity pattern must necessarily be age-based and assumed constant across the time series. Minor deviations from year to year are not an issue, but potential changes in selectivity from year to year may be large and currently cannot be quantified. At issue are: that the trawling is targeted on marks at the discretion of the voyage leader; the stated aim of the trawling is to obtain a “sample” of fish (and to avoid large catches and specific by-catch species); different fishing strategies will have been used by the various personnel doing the fishing over the years (on the two vessels) – e.g., a dip into a layer, or targeting the headline below the layer; and the length compositions are post-stratified and assigned to transects (for scaling up) on an ad hoc basis.

It was noted that acoustic surveys conducted elsewhere in the world also collect length and age data. However, these data do not appear to then be used in assessment models but are used only with the specific survey data.

There is undoubtedly some length structure within the hake layers and marks; the length composition in a trawl catch will not only depend on what fish are in the mark but also on how the mark is fished. The absence of any statistical design aimed at providing consistent representative sampling for length or age structure makes it very difficult to justify the use of the survey composition data in stock assessment model runs.

It was also clear, from the scale of the expansion factors that had been applied to the early indices in the acoustic time series, that they could not be considered comparable to the later indices. As a matter of good practice, they needed to be split into a separate time series (even though, within the early series not all surveys are necessarily comparable). Finally, with regard to the acoustic data, the biomass estimate of the 1986 survey was suspect because of a large difference between the pre- and post-survey calibration results.

The commercial catch composition data were not without problems either. In both models there are concerns that catch compositions have not been stratified and scaled in the most appropriate manner (see Research Recommendations). However, the main issue is that of the growth of young fish during the fishing season. In the SS3 preliminary base model, a single mid-year prediction of proportions at age for a given length are made to fit the conditional age-at-length data. However, in the fishery, which may extend for 7-8 months, including summer when perhaps most growth occurs, the proportions of age-at-length can be dramatically different at the

start of fishing season compared to the end of the fishing season (for some of the younger age classes that are vulnerable to the US fishery). In the TINSS model, the age frequencies were derived from the age samples independently of the length frequencies (so do not suffer from this problem).

Technical merits:

The Panel chose the TINSS model as the base, but both models are equally acceptable and so are considered here.

Data used in both models:

- The most defensible data set that was available in the timeframe of the review

TINSS:

- A reasonably well tested model as it has been used for a number of years and has been peer reviewed on each occasion.
- Has the advantage of relative simplicity in terms of population dynamics.
- Explicitly accounts for observation and process error
- Integrates major aspects of uncertainty through Bayesian estimation.

SS3:

- Developed using a well tested and documented package
- Has separate US and Canadian fisheries and associated selectivities
- Attempts to account for changes in fishery selectivity over time in both fisheries

Technical deficiencies:

As in the above section, both models are considered:

TINSS:

- Some of the technical aspects of the model are not well understood by many stock assessment scientists (because it is a relatively unusual model in the stock assessment context); hence the level of peer review it has received may not be as in-depth as it could be.
- Similarly, the suite of suitable model diagnostics is not as well-developed as for a “standard” observation error model (such as SS3).
- The age frequencies may not be properly weighted because of stratification issues and the aggregation into a single fishery.
- There is no mechanism to compensate for possible changes in fishery selectivity.
- The model does not have informed priors for the acoustic qs which limits our ability to judge the plausibility of the estimated size of the stock

SS3:

- The model may be over-parameterized due to the extensive blocking structure which attempts to compensate for possible changes in fishery selectivities.
- Some of the supposedly un-informative priors on selectivity parameters may actually be highly informative
- The age frequencies may not be properly weighted because of stratification issues.
- The model reviewed by the Panel does not integrate uncertainty through Bayesian estimation (the Bayesian run is not available to the Panel before the finalization of this report due to time constraints).
- The model does not have informed priors for the acoustic qs which limits our ability to judge the plausibility of the estimated size of the stock

Explanation of areas of disagreement regarding STAR Panel recommendations

Among STAR Panel members (including GAP and GMT representatives)

None

Between the STAR Panel and STAT Team

There were no significant disagreements between the STAR Panel and STAT Team.

Unresolved problems and major uncertainties

- It is not clear how best to assess this stock, either in terms of the appropriate level of model complexity, or in terms of the level of data aggregation (but, this is a generic problem for many stock assessments).
- The available input data are inadequate to provide a precise assessment of stock status. In particular, the scale of the stock, in absolute terms, is very poorly determined.
- The stratification and scaling of the age samples may be inappropriate.
- The split of the acoustic surveys into two time series may need revision in terms of which years belong in which series (or if more than two series are needed).

Management, data, or fishery issues raised by the GMT or GAP representatives during the STAR Panel.

The GMT representative noted that there were differences in how management advice is formulated using the two different models. The GMT may have to rely on the assessment author to provide needed GMT input in the absence of people familiar with the TINSS model. The

issue of management advice linked to specific models may be something the SSC may wish to take up.

Recommendations for future research and data collection.

- A detailed analysis of catch, effort, length, and age data by sex, going as far back as possible, and split by fleet, and vessel type, is needed to help understand the commercial data which go into the stock assessment models. In particular, this would enable, (i) defensible length and age frequencies to be constructed by fleet (not just shore-based and at-sea within country), which in turn may enable the modeling of the fisheries data with constant selectivities over time within fleet (or, at least, lead to a reduction in the need for time-varying selectivities); and (ii) abundance indices (i.e. one or more fleet-based CPUE indices) to be explored to provide an alternative (or an addition) to the acoustic survey biomass (should the squid remain in the region and continue to make survey-based hake biomass unreliable; also, having alternative or additional indices would strengthen the ability of the modelers to adequately assess the hake stock) . This should also include additional spatial data describing the tribal and shore-based fisheries.
- Analysis from all data sources (commercial and acoustic survey) aimed at understanding the spatial, vertical, and temporal patterns of hake distribution (by length, age, and sex).
- Fund research into the appropriateness of attempting to produce biomass estimates at length, age, and sex, from acoustic surveys of semi-demersal species such as hake and pollock, including in the presence of possible confounding species such as Humboldt squid and lingcod. Once the work has been done (by statistician(s) with practical fisheries experience, in conjunction with acousticians) convene a workshop to discuss and review the findings. Ideally this should also address the issue of adequately sampling to ground-truth the acoustic estimates, including, for example, duration of trawl sampling, using a commercial trawler to sample, using another (additional) gear type to sample.
- Place a very high priority on obtaining a defensible length to target strength relationship for hake.
- Place a high priority on obtaining a defensible length to target strength relationship for Humboldt squid and assessing available techniques to acoustically distinguish between hake and squid biomass in the field.
- Construct informed priors for the acoustic qs associated with the existing time series (this will ensure that future model runs stay in sensible space, or alternatively, that the estimates will be a revealing diagnostic).
- Provide an option in SS3 to disable or severely limit the penalty on recruitment deviations while maintaining internal consistency in the definition of B_0 .

Acknowledgements

The STAR wishes to extend its thanks to Dr. Jim Hastie and Ms. Stacey Miller of the NWFSC for logistical and Staff support and Mr. John DeVore of the Pacific Fishery Management Council for guidance on terms of reference and recording of the Panel. The STAR panel also thanks Dr. Dezhang Chu and his staff for their contributions and also Mr. Chris Grandin and Mr. Greg Workman of the Department of Fisheries and Oceans, British Columbia for presentation of Canadian fishery and research data.

The Panel would like to extend a note of commendation to the STAT members: Ian Stewart, and Owen Hamel of Northwest Fisheries Science Center and Steve Martell, University of British Columbia. They performed exceptionally and their modeling skills and data knowledge made it possible to explore the models and data sets almost to their entirety. The Panel notes that any inability of the STAT to perform related to inherent limitations in the models, computing power, or a lack of access to data.

References

- Dorn, M. W. 1996. Status of the coastal Pacific whiting resource in 1996. In Pacific Fishery Management Council, Appendix Volume I: Status of the Pacific Coast groundfish fishery through 1996 and recommended acceptable biological catches in 1997, p. A1-A77. Pacific Fishery Management Council, Portland, OR.
- Helser, T.E, R.D. Methot, and G. W. Fleischer. 2004. Pacific whiting assessment update for 2003. In Pacific Fishery Management Council, Status of the Pacific Coast groundfish fishery through 2001 and recommended acceptable biological catches in 2002. Pacific Fishery Management Council, Portland, OR.
- Martell, S. (2010). Assessment and Management advice for Pacific hake in U.S. and Canadian waters in 2010.
- Pacific Fishery Management Council (2008). Terms of Reference for the Groundfish Stock Assessment and Review Process for 2009-2010.
- Stewart I.J. and O.S. Hamel (2010). Stock Assessment of Pacific Hake, *Merluccius productus*, (a.k.a Whiting) in U.S. and Canadian Waters in 2010.
- Traynor, J.J. 1996. Target-strength measurements of walleye pollock (*Theragra chalcogramma*) and Pacific whiting (*Merluccius productus*). ICES Journal of Marine Science 53:253-258.