REPORT ON THE STOCK ASSESSMENT AND REVIEW

OF PACIFIC HAKE (WHITING)

FEBRUARY 2–4, 2004

SEATTLE, WASHINGTON

Prepared by

Patrick Cordue
Fisheries Consultant
New Zealand

for

University of Miami
Independent System for Peer Review

20 February 2004

EXECUTIVE SUMMARY

A draft Pacific hake (whiting) assessment was reviewed at a STAR Panel meeting from February 2–4, 2004, at NMFS Northwest Fisheries Science Center, in Seattle, Washington. I fulfilled the role of an "external" member of the Panel. Background documentation and the draft assessment document were supplied before the meeting in sufficient time to allow full familiarization with the assessment. The meeting was conducted in a constructive and amicable atmosphere and the STAT Team was proficient, professional, and helpful.

The assessment used a single-sex age structured model implemented in AD-Model Builder. Maximum likelihood estimation was used with Bayesian extensions to quantify uncertainty. There were three main data sources: commercial catch and catch at age, acoustic survey biomass indices and proportion at age, and recruitment indices from a SWFSC midwater trawl survey targeting juvenile rockfish. A critical assumption made for all runs presented in the draft assessment was that the acoustic surveys provided indices of *absolute* abundance (equivalent to acoustic catchability, $q = 1$). This is a dubious assumption as it requires that known negative biases in areal and vertical availability be canceled out by positive biases in target identification error and the poorly estimated length to target strength relationship. I was concerned by this assumption and suggested that the alternative of estimating q be more fully explored.

At the end of the second meeting day, the Panel and STAT Team agreed to take forward to the projection stage two model runs: one with $q = 1$ for continuity in management advice; and one with an estimated q (being more scientifically defendable). However, on the final day of the meeting, when the agreed runs estimating q were presented, the estimates of q appeared too low to be credible (being < 0.3). A simple procedure was adopted to bound q using expert opinion with resulting bounds for q of 0.55–1.25. The meeting then agreed to take forward two runs with fixed q to the projection stage: $q = 1$ and $q = 0.6$. These were taken to represent the range of uncertainty with regard to stock status.

The assessment uses relevant data sources in a coherent model framework and provides the best currently available scientific information on the status of the stock. The use of fixed values of q in the two runs is a temporary measure which was forced on the meeting due to time constraints. For the next assessment an informed prior on the acoustic q should be developed and this should be updated for future assessments as more data (particularly on target strength) become available. The dubious assumption of $q = 1$ need not be made again.

BACKGROUND

This report presents my personal view of the assessment and makes some recommendations with regard to future assessments of this stock and some aspects of the STAR Panel process. This report is best read in conjunction with the STAR Panel Report which contains further details of the meeting and other recommendations concerning the assessment.

REVIEW ACTIVITIES

Meeting Preparation

Prior to the meeting, the supplied background material (Appendix 1) was read and other relevant documents and literature (available on the Web) were consulted. Given the reliance of the assessment on the use of absolute abundance estimates from the acoustic time series, particular emphasis was placed on obtaining documentation relating to the survey design (Anon. 1995, 1998) and the length to target strength relationship assumed (Traynor 1996, Kieser et al. 1999).

Meeting Attendance

A brief narrative of the meeting is given below. Details of requested analyses and final model runs are in the STAR Panel Report.

2 February

The meeting was convened at 8am and began with a round of introductions.

The first presentation at the meeting was by Guy Fleischer on the 2003 acoustic survey. This was the first survey undertaken by the newly formed acoustic group at the NWFSC. They had been careful to follow established sampling procedures and survey design. It was also the first survey done entirely on the Canadian research vessel W.E. Ricker. The presentation primarily covered the survey results: distribution of fish, biomass estimates, estimated proportions at age. There were questions and discussion with regard to survey design and the use of the acoustic estimates as absolute abundance. The issue of the "bio-fouling" of the transducer was also raised (barnacle fouling was discovered part way through the survey during a calibration). The barnacles were removed before completing the survey and separate calibration constants were applied to the relevant survey legs. There was some discussion of previous surveys and, in particular, the 1986 survey which had differing pre and post survey calibrations.

The issue of variance calculations for the survey estimates was raised and it was stated that this was an area that was being considered. Routine calculation of estimate precision had not been done. The Panel requested that a CV be estimated for the 2003 total biomass estimate using the formulae of Jolly and Hampton (1990). Although not directly applicable (to the survey design used) the idea was to obtain a "ball park" estimate of the sampling variance (being better than nothing at all).

The second presentation was by Tom Helser on the stock assessment methods and results (Helser et al. 2004, February draft). He began with the final results, explaining how and why the latest assessment differed from the 2002 assessment (Helser et al. 2002). The latest assessment results were more optimistic than the 2002 base model mainly because of the higher estimate for the strong 1999 cohort. He then explained the latest methods and estimation procedure and how these differed from what had been done in 2002. He covered the new data that had been used and a suite of runs which created an "audit trail" from the previous results to his current preferred model. Several issues were raised for discussion: the use of the acoustic indices as absolute; the estimation of an "acoustic selectivity" which applied to both the trawl proportions at age and the biomass indices (despite the acoustic beam presumably having a different selection pattern to the trawl gear); the use of such low CVs on the biomass indices (0.1 for recent acoustic surveys and 0.2 for early acoustic surveys); the internal consistency of the model with regard to variance assumptions; and the form of the likelihood used ("median unbiased" rather than mean unbiased, see Appendix 2).

The Panel requested that point estimates be obtained for a variety of new model formulations. Some models were to include the estimation of the acoustic catchability (q), and increased CVs on the biomass indices. Also, the dubious 1986 acoustic estimates were to be excluded from some runs.

3 February

The meeting resumed at about 10am to allow time for the numerous model runs to be completed. The results were presented by Tom Helser and Rick Methot. Discussions concentrated on two aspects of the model results: the internal consistency of the variance assumptions, and the change in total likelihood and likelihood components for different model formulations. With regard to the first aspect, two main diagnostics were considered: the standard deviation of the standardized residuals (which should be near to 1 if the variance assumptions are consistent with the residuals), and the median estimated effective sample sizes for the multinomial distributions (which should be near to the assumed effective sample sizes).

There was discussion about the large number of parameters being estimated for the fishing selectivities (192 for the random walks and 6 extra for year and age specific selectivity). Rick indicated that the philosophy being followed was essentially as in a VPA, with the selectivities being flexible enough to allow the model to closely match the estimated catch at age. Whether this is a good idea or not is debatable, but the alternative of trying to identify and model the covariates affecting fishery selectivity is difficult and may not produce better estimates in any case. The Panel did consider a run where the random walks for the selection parameters were essentially eliminated and replaced by three periods coinciding with changes in fleet composition. It appeared that the increase in total likelihood from this run to the run with the random walks was sufficient to justify the use of the random walks (but it was hard to judge as it was unclear how many effective parameters the random walks imply).

At the end of the day the meeting had agreed on four runs to be taken forward to the final day of the meeting. Two runs were to include an assumed $q = 1$, to provide continuity with previous assessments. In one run (Model 1a) variance assumptions were to be tuned so that they were consistent with model output. However, it was know that this would result in the acoustic biomass time series being very poorly fitted. The second run (Model 1b) was to have lower CVs on the acoustic indices to improve the fit to the time series. The two remaining runs (Models 2a and 2b) were to have the acoustic q estimated. They differed in their treatment of the acoustic selectivity.

4 February

The meeting resumed at 8am on the final day. Panel members began writing their assigned sections of the STAR Panel report while STAT Team members were getting the results of the runs ready for presentation (and waiting on preliminary MCMC runs to finish). Tom and Rick first presented results from Models 1a and 1b. The meeting discussed which of the two runs was to be preferred. We had already decided that two final runs would be taken through to the projection and yield stage: one with $q = 1$, and one with estimated q. We concluded that Model 1b was to be preferred to Model 1a. This was not an easy decision. Model 1a had a very bad fit to the acoustic indices and appeared to violate the assumption of independent lognormal errors (as it over-estimated all of the indices from 1977 to 1989). Conversely, Model 1b did not fit the acoustic indices as well as it should, given the assumed CVs (thus also contradicting the statistical assumptions).

The results from estimating q provided the meeting with a further problem. The estimates were both below 0.3 and most meeting members rather strongly suggested that these were implausibly low. However, whether they are implausible depends on the level of bias in the assumption of $q = 1$. The meeting agreed to bound q using four components of error (a simplified version of the method from Cordue 1996). Bounds were agreed for areal availability, vertical availability, target identification error, and error in mean target strength and gave overall bounds for q of 0.55–1.25. Since the estimates of q were now seen to be implausible it was decided not to continue with either Model 2a or 2b. Instead, another run with $q = 0.6$ was produced to provide a credible alternative to the assumption of $q = 1$. These two runs were seen as providing credible lower and upper bounds on reality.

There was considerable discussion on the relative weightings to be assigned to the two runs. Divergent opinions eventually forced the compromise to assign equal weightings (though no one thought that they should have vastly different weights). There was some discussion on how the two runs might be merged in a single projection analysis but it was concluded that the two runs should be kept separate and a simple decision table produced. There was also some discussion of appropriate management strategies for the stock given the highly variable recruitment strengths observed.

5 February

I met with Guy Fleischer and his acoustic group at the NWFSC. We had discussions about the previous acoustic surveys and current research directions. This was very helpful in clarifying the methods used in obtaining the acoustic biomass indices and the estimates of proportion at age.

9–13 February

After the meeting the STAR Panel report was completed via email. Comments were primarily editorial. However, it was agreed to include some discussion of management strategies in the report because even the application of conservative strategies can result in point estimates for stock depletion less than the overfishing threshold of 25%. This is simply due to the high recruitment variability of hake.

Conduct of the Meeting

The meeting was held in a constructive and amicable atmosphere. Presentations were professional and discussions informative. The draft assessment document was well written and comprehensive. The meeting covered a lot of ground and reached a good outcome. However, there are two ways in which better results might have been achieved.

First, it would have been very helpful to have a fully documented record of which aspects of the assessment (data and methods) had previously been reviewed (to some extent). On several occasions the meeting was told that "such and such" had been looked at by someone in the past. The extent to which the area had been "reviewed" was not apparent and nor was the extent to which it had been documented. It would aid a STAR Panel if there was a section in the assessment document listing all previous "reviews" of areas relevant to the assessment, including references. Ideally, a set of all relevant documents would be available to the Panel if requested.

For the hake assessment, the complexity of the assessment and the number of data sources, means that no STAR Panel can fully review the assessment in a period of three days. A list of previous "reviews" (examination of methods/data sources with regard to particular choices/preferences) would help in that the Panel would know which areas had not been previously covered. However, even with this help, I do not think that three days is sufficient. As recommended in the STAR Panel report, a full week would seem justified for the hake assessment.

Strengths and Weaknesses in Assessment Methods and Advice

This is extensively discussed in the STAR Panel report. Below, I summarise the issues and explain why the assumption of the acoustic $q = 1$ should be abandoned in future assessments.

Data

The assessment uses data from three main sources: catch and catch at age from the commercial fisheries; biomass indices and proportions at age from the acoustic survey; and recruitment indices from the SWFSC midwater trawl survey targeting pelagic juvenile rockfish. The main points with regard to the data are:

- Estimated catches are available from 1966-2003 by nation and fishery. They are believed to be accurate from 1977-2003. In the earlier period the total catch may have been underestimated.
- There has been extensive sampling of the commercial catch, with catch at age estimates for the U.S. fishery from 1973-2003 and for the Canadian fishery from 1977-2003.

- The acoustic surveys are triennial from 1977 to 2001, with the latest survey in 2003. The surveys from 1977 to 1989 cover a smaller depth range than the later surveys and the 1977 to 1992 surveys do not go as far north as the later surveys.
- Deep water and northern expansion factors were applied to the appropriate surveys in an attempt to make the whole time series consistent.
- The SWFSC midwater trawl survey provides a recruitment index from 1983-2003. This survey covers a small geographic area relative to the distribution of juvenile hake. However, the indices have been shown to have a significant correlation to model estimates of recruitment.
- There are differences between the SWFSC series and a shorter recruitment time series over a wider area (PWCC-NMFS midwater trawl survey).

The only concern about the annual catches is a possible large underestimation before 1977. The effect on assessment results should be explored in sensitivity tests (but it may make little difference). The sampling of the commercial catches appears more than adequate in terms of number of hauls sampled, fish measured, and fish aged (Table 3, Helser et al 2004). There may be issues with regard to how the data are combined/weighted to obtain annual estimates of proportion at age, but the Panel did not have time to explore this. Of more concern may be the effects of ageing error. There was limited discussion of this at the meeting. Although strong cohorts are clearly visible in the data one does wonder how consistent the readings can be across years given changes in personnel and methods (apparently, whole otoliths were read in the earlier part of the time series, before "break and burn" was adopted).

A proper interpretation of the acoustic data is crucial to the assessment. The survey results in 1986 were not used in the final assessment runs because of the different pre and post survey calibrations. The time series is now long enough that this potentially bad data need not be used. The spatial coverage of the surveys is best in the 1995–2003 surveys. The attempt to correct for areas not covered in the earlier surveys is probably as good a result as can be achieved. However, this attempt was made primarily with a view to assuming $q = 1$ in the assessment runs. A more certain method of achieving a consistent relative time series is to use the same survey area in each year (i.e., discard some of the data from the later surveys). This is an unpalatable alternative, but it could at least be explored as a sensitivity run. A further alternative of splitting the time series and estimating twos *q*s could also be explored.

The estimated proportions at age for the acoustic survey are derived from targeted trawls on the marks seen on the acoustic transects. It is the decision of the Chief Scientist as to which marks are trawled on. Dense marks take priority as misidentification of such marks will have the most effect on the hake biomass estimates. Marks may be targeted by either midwater or bottom trawl gear, but the predominant method appears to be midwater trawl. There are two major problems with this data. It is not possible to obtain representative samples of proportion at age without some randomization in the allocation of the trawls. Further, a consistent time series cannot be achieved when subjective decisions govern which marks will be trawled on and which gear will be used. The only way in which it may be possible to properly use this data is a post-stratification based on the factors which affect the structure of the trawl length frequencies (e.g., depth, "mark type"). In years when there is a single dominant cohort this may not make much difference.

There is also the issue of how to weight the samples from individual trawls to obtain an overall stratum length frequency. Apparently, each fish measured is given equal weight, independent of the proportion of the trawl catch sampled or the estimated biomass of the targeted mark. Alternative weighting schemes could provide quite different results in some years and these should be explored.

Biological parameters

Year specific weights at age were used in all years for each fishery and survey because of significant variation in the observed weight at age. A constant and age independent estimate of natural mortality was used. A constant female maturity at age vector was also used. The Panel did not consider the derivation or use of these estimates in any detail. However, each of these factors could affect the assessment results and some sensitivity runs with alternative assumptions should be done in future.

Stock assessment model and estimation procedure

The assessment model uses standard population dynamics equations being a single stock, single sex, age structured, and spatially aggregated model. The Canadian and U.S. fisheries are modeled as distinct year-round fisheries. Fishing selectivity patterns are year specific (constrained by a random walk) to allow for changes in fleet composition and shifts of fish distribution (across the border). The acoustic time series is modeled using a single selectivity pattern which applies to both the biomass indices and the estimated proportions at age. The estimation procedure is essentially maximum likelihood with Bayesian extensions for estimating parameter uncertainty.

The final model runs which were supported by the meeting differed mainly in the assumed value for the acoustic q: Model 1b, $q = 1$; Model 1c, $q = 0.6$. As a compromise both models were given equal relative weight by the STAR Panel and STAT Team. Nobody strongly favoured one run over the other with the choice of weights being between 50:50, 60:40, or 40:60. In previous assessments the assumption of $q = 1$ had always been made although this was questioned in the 2001 assessment (Moore 2002). Therefore, it was something of a "break through" to have a run with $q < 1$ being taken forward.

I supported a higher weight for Model 1c as this essentially used the maximum likelihood estimate of q within the bounds we had specified for q. I also supported Model 1b going forward as it provided some continuity in terms of management advice. However, the scientific basis for assuming $q = 1$ for any part of the acoustic time series is dubious.

For an acoustic survey to provide an absolute biomass index requires that the negative biases are exactly balanced by the positive biases. In the case of the hake surveys, including the recent ones, there are negative biases for the areal availability (not all adult hake are necessarily in the survey area) and the vertical availability (there is under sampling in the shadow zone near the seabed). There is also a potential bias with target identification (assigning acoustic marks to particular species mixes – e.g., 100% hake) but this could be positive or negative. These biases may be small (particularly in recent surveys) and $q = 1$ could be a reasonable approximation if the length to target strength relationship for hake was accurately known. However, the cited reference for this relationship (Traynor 1996) is not compelling about the accuracy of the relationship.

From the abstract of Traynor (1996), writing relative to pollock: "For Pacific whiting, TS measurements have not been made over an adequate range of fish lengths to determine the appropriate relationship for use in scaling echo-integration surveys. However, results presented in this paper and elsewhere suggest a smaller TS (about 2dB for a given length) for 40–60cm fish.". In the conclusion, the author expresses similar sentiments and recommends: "Every effort should be made to obtain both samples for swimbladder morphology studies as well as *in situ* TS measurements for Pacific whiting under a variety of behavior situations, especially for small whiting, so that an appropriate TS to length relationship can be developed." Given these statements the relationship is, at best, only tentative.

At worst, there may be considerable bias in mean backscattering cross sections derived from the relationship. Certainly there are problems with the seven data points used in the regression to obtain the length to target strength relationship:

- Two of the points are from Neal J. Williamson (pers. comm.) – possibly fine, but their derivation is not documented.
- The two new points given are actually from a single recording, but with targets from different depth ranges (50–100m, and 100–150m). There is more than a 2dB difference between the two points.
- The shallow targets could be contaminated with another species as they include a mode at –50dB which is not present in the deep targets.
- The deep targets are at an extreme range for differentiating single targets.

- The single target algorithm used to record the data is known to be biased (Soule *et al.* 1995).
- There was only a single trawl on the mark and this did not adequately sample the full depth range of the targets ("Trawl depth: 94–116 m", perhaps being the headrope depth; there was approximately a 21m vertical mouth opening).
- All recordings were from low density night time marks (but the derived relationship is applied to high density day time marks).
- A slope of 20 is assumed in the relationship (perhaps necessary given the limited length range).

The method of accumulating *in situ* data points for a linear regression to estimate a length to target strength relationship is widely used but suffers from potential selection bias. The problem is that points are obtained sequentially. There is the possibility of omitting target strength recordings which do not fit with the previously adopted points (it is easy to rationalize why a particular recording has "not worked"). Also, if more than one species (or length class) is present then the assignment of target strength modes to particular species (or length classes) is done in light of the existing points. More objective methods are perhaps to be preferred. Cordue et al (2001) applied a modelling method which used all the available target strength recordings aimed at a particular target species. These were contaminated by a range of species as evidenced by the trawl data, but rather than subjectively assigning particular species to particular target strength modes, a model was fitted to the data to estimate length to target strength relationships for a number of species. Other methods using model fitting have also been developed. For example, Coombs and Barr (in press) develop length to target strength relationships by fitting a swimbladder model to *in situ* single target distributions.

I believe that much more work needs to be done with regard to estimating the length to target strength relationship for hake. I was pleased to hear that some swimbladder modeling is now being done and that new technology should allow *in situ* recordings in a wider range of circumstances, including higher density marks during the day. When this new work has been brought together with existing data an accurate length to target strength relationship may be established. At that stage, there might be a case for assuming $q = 1$. In the interim, the fashionable (and certainly better) approach is to develop a Bayesian prior for the acoustic q as recommended in the STAR Panel report.

Another issue with regard to the use of the acoustic data in the assessment model is that of the single acoustic selectivity estimated. In reality there are a minimum of three selection patterns, one for each gear type: acoustic beam, midwater trawl, bottom trawl. I would expect the selection patterns of the three gear types to be quite different and the acoustic beam to be relatively non-selective (though this does depend on target identification methods, and certainly it will be less selective of 2 year old hake as they are often found "mixed with plankton layers"). It is debatable whether it is better to apply a trawl selection pattern to the acoustic biomass indices, or to decouple the trawl selection pattern from the acoustic biomass and simply assume a pattern for the acoustic indices. I prefer the latter approach (rather than the former that was used), with some sensitivity tests on the selection pattern assumed for the acoustic biomass indices.

SUMMARY OF FINDINGS

The draft assessment document (Helser et al. 2004) was well written and comprehensive. The meeting was conducted efficiently and effectively with as good a review as was possible in the available time. The two assessment runs taken forward for use in projections and the decision table capture a defendable range of uncertainty with regard to the status of the stock. I support the equal weightings assigned to the two runs by the meeting on the basis that one run provides continuity for management purposes and the other provides a credible alternative.

CONCLUSIONS AND RECOMMENDATIONS

My conclusions with regard to the assessment are:

- The assessment was much improved by the STAR Panel process.
- The assessment is data rich in terms of quantity, but the quality of some of the data needs to be more carefully considered.
- The hake assessment provides the best currently available scientific information on the status of the stock.

I have recommendations in a number of areas given below.

Data reviews:

- The quality of all data sources should be formally assessed:
  - the sampling error (and distribution) of data should be estimated (e.g., CVs on biomass indices)
  - data which cannot be adequately modeled statistically (because of severe or impenetrable systematic errors) should be identified and removed from the estimation models.
- Priority should be given to the acoustic survey data, considering questions such as:
  - are there adequate trawls in each year to estimate proportion at age?
  - can the problems of targeted trawls be ameliorated by post stratification?
  - can a consistent relative time series be developed using core strata?
- The question of ageing error should be formally reviewed, including:
  - documentation of methods and readers used
  - documentation of known problems
  - documentation of within reader and between reader variability
  - recommendations on which data should be used in the assessment
  - recommendations on which otoliths should be re-aged (if any).

Acoustic target strength:

- At a timely point in the future (e.g., when swimbladder modelling results are available) the length to target strength relationship should be re-estimated for hake.
- Alternative methods should be used to explore the level of uncertainty with regard to the relationship.
- Estimates of imprecision in mean backscattering cross section should be incorporated into the variance estimates for the acoustic biomass indices.

Stock assessment modelling:

- A prior on the acoustic q should be developed as soon as possible.
- The prior should be updated as more data are gathered with regard to areal availability, vertical availability, target identification, and mean target strength (the assumption of $q = 1$ need never be made again).
- Realistic CVs should be used for all data points (at least in initial weightings) – in particular, different CVs across years reflecting the level of sampling in those years.
- Alternative formulations of the fishing selectivities should be explored aiming for parsimony in the number of parameters estimated.

STAR Panel process:

For important, data rich assessments:

- The STAR Panel meetings should be five working days.
- From time to time, independent reviews of important aspects of the assessment should be done.
- As part of the assessment document there should be a full record of relevant past reviews.

REFERENCES

(see Appendix 1 for further references)

Anon. 1995. Preliminary cruise results NOAA ship Miller Freeman Cruise No. 95–10: Echo integration-trawl survey of Pacific whiting along the west coasts of the United States and Canada. Draft document RACE Division, Alaska Fisheries Science Center, NMFS, Seattle, WA.

Anon. 1998. Preliminary cruise results NOAA ship Miller Freeman Cruise No. 98–10: Echo integration-trawl survey of Pacific whiting along the west coasts of the United States and Canada. Draft document RACE Division, Alaska Fisheries Science Center, NMFS, Seattle, WA.

Coombs, R.F. and Barr, R. (in press). Acoustic remote sensing of swimbladder orientation and species mix in the oreo population on the Chatham Rise. *J. Acoust. Soc. Am. 115 (4)*: April 2004.

Cordue, P.L. 1996. A model based method for bounding virgin biomass using a catch history, relative abundance indices, and ancillary information. New Zealand Fisheries Assessment Research Document 96/8. 48 p

Cordue, P.L., Coombs, R.F., and Macaulay, G.J. 2001. A least squares method of estimating length to target strength relationships from *in situ* target strength distributions and length frequencies. *J. Acoust. Soc. Am. 109*: 155–163.

Jolly, G.M. and Hampton, I. 1990. A stratified random transect design for acoustic surveys of fish stocks. *Can. J. Fish. Aquat. Sci. 47*: 1282–1291.

Kieser, R., Saunders, M.W. and Cooke, K. 1999. Review of hydroacoustic methodology and Pacific hake biomass estimates for the Strait of Georgia 1981 to 1998. CSAS Research Document 99/15. 53 p.

Soule, M., Barange, M. and Hampton, I. 1995. Evidence of bias in estimates of target strength obtained with a split-beam echo sounder. *ICES Journal of Marine Science, 52*: 139–144.

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.

APPENDIX 1: MATERIAL PROVIDED

Anon. 2002. Report of the Joint Canada – USA Review Panel on the Stock Assessment of the Coastal Pacific Hake/Whiting Stock Off the West Coast of North America, National Marine Fisheries Service, Montlake Laboratory, Seattle, WA. U.S.A., 20-22 February 2002. 15 p.

Anon. 2004. Terms of reference for the groundfish stock assessment review process during 2005–2006. 19 p.

Helser, T.E., M. W. Dorn, M.W. Saunders, C.D. Wilson, M.A. Guttormsen, K. Cooke, and M.E. Wilkins. 2002. Status of Pacific hake/whiting stock in U.S. and Canadian waters in 2001. 37 p.

Helser, T.E., R.D. Methot, and G.W. Fleischer. 2004. Stock assessment of Pacific hake (whiting) in U.S. and Canadian waters in 2003. Draft document February 2004. 112 p.

Moore, R. 2002. Minority report re Anon. 2002. 2 p.

APPENDIX 2: LIKELIHOOD AND LOGNORMAL ERRORS

The presentation of the lognormal error structures in the draft assessment document (Helser et al. 2004) is not ideal as the error structures are not explicitly presented in the equations. There is a statement of the "predicted values from the model" followed by the log-likelihood. This does allow the explicit error structure to be deduced, but it would be better to present the formal equation. When this is done, it is actually seen that the likelihood being used is inconsistent with the usual notion of an index.

To illustrate, consider a biomass index X_i:

$$X_i = qB_i\varepsilon_i$$

where B_i is the biomass (in year i), q is the proportionality constant, and ε_i is the error (in year i). Suppose that the errors are lognormal: $\log(\varepsilon_i) \sim \mathrm{N}(\mu_i, \sigma_i^2)$. It then follows that,

$$\log(X_i) \sim N(\log(qB_i) + \mu_i, \sigma_i^2)$$

and the negative log-likelihood (ignoring constants) is

$$\frac{1}{2}\sum_i \left[\log(\sigma_i^2) + \frac{(\log(X_i) - \log(qB_i) - \mu_i)^2}{\sigma_i^2} \right]$$

If the variances are assumed known, then the first term in the square brackets in the above equation can be ignored. The likelihoods presented in the draft assessment document are consistent with the assumption, in every year, that $\mu_i = 0$. However, under this assumption it follows that:

$$\mathrm{E}(X_i) = qB_i\mathrm{E}(\varepsilon_i) = qB_i e^{\frac{\sigma_i^2}{2}} = qB_i\sqrt{cv_i^2 + 1}$$

where cv_i is the specified c.v. in year i.

When the CVs are relatively small (< 0.35), there is a very small bias in the indices. However, by definition, they are no longer indices in the usual sense. The assumption in the assessment document is consistent with "median" unbiased indices, in that there is a 50% probability that an index will be above or below the true value (qB_i). This would be acceptable if the random variables in question could be expected to have this property. However, this does not appear to be the case. Certainly, the acoustic biomass index is derived from a survey where the transects within a stratum are assumed to be random with regard to fish distribution. If the transect densities are weighted by transect length then the expected value of each estimated stratum biomass is equal to the true stratum biomass (i.e., "mean" unbiased, not "median" unbiased). A similar argument could be made about the two-year old indices (CVs = 0.5) and the total catch (although, with CVs = 0.05, the small bias is irrelevant to the results).

APPENDIX 3: STATEMENT OF WORK

Consulting Agreement Between The University of Miami and Patrick Cordue

16 January 2004

General

The consultant will participate in the Stock Assessment and Review (STAR) Panel of the Pacific Fishery Management Council (PFMC) from February 2-6, 2004, in Seattle, Washington. This assessment will provide the basis for management of the Pacific hake.

The consultant should have expertise in age-structured statistical catch at age modeling and experience with AD Model Builder.

The consultant's duties shall not exceed a maximum total of 14 days: Several days prior to the meeting for document review; the five-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 will be provided with the following documents:

Current draft Pacific hake stock assessments report;
Most recent previous Pacific hake stock assessment;
An electronic copy of the data, the parameters, and the model used for the assessment (if requested by reviewer).

Specific

1) Become familiar with the draft Pacific hake stock assessment and background materials;

2) Participate in the STAR Panel to be held in Seattle, Washington, from February 2-6, 2004;

3) Understand 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) No later than February 20, 2004, submit a written report[1] consisting of the findings, analysis, and conclusions, addressed to the "University of Miami Independent System for Peer Review," and sent to Dr. David Die, via email to ddie@rsmas.miami.edu, and to Mr. Manoj Shivlani, via email to mshivlani@rsmas.miami.edu.

[1] The written report will undergo an internal CIE review before it is considered final. After completion, the CIE will create a PDF version of the written report that will be submitted to NMFS and the consultant.

ANNEX I: REPORT GENERATION AND PROCEDURAL ITEMS

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

2. The main body of the report should consist of a background, description of review activities, summary of findings, conclusions/recommendations, and references.

3. The report should also include as separate appendices the bibliography of all materials provided and a copy of the statement of work.