Report on

Joint US-Canada

Technical Review Panel

for the

Pacific Whiting Stock Assessment

February 8 –10, 2010

Seattle, Washington

P.L. Cordue

Fisheries consultant

New Zealand

***For* CIE Independent System for Peer Review**

15 March 2010

Executive summary

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 assessments. One assessment was prepared by the US members of the Stock Assessment Team (STAT) using Stock Synthesis 3 (SS3) and the other by the Canadian member of the STAT using the "TINSS" model. I was a member of the STAR Panel which consisted of two CIE reviewers, a Canadian member, and an SSC representative as the Chair.

There were three main sources of data available to the assessment models. The catch history, commercial catch sampling data (length and age), and acoustic survey data (total biomass, length and age).

The SS3 model had two fisheries (Canadian and US) and fitted to the acoustic biomass indices, the survey length frequencies and age-at-length data, and commercial length frequencies and age-at-length data. Time-varying selectivity was estimated in the commercial fisheries together with time-varying growth. Domed selectivities were estimated for both fisheries and the acoustic survey, together with steepness and a senescent natural mortality (*M*) for fish older than 14 years. Recruitment deviations and nuisance parameters were also estimated (see Stewart and Hamel, 2010, for details).

The TINSS model had simpler population dynamics with just a single fishery and was fitted to the acoustic biomass indices and age frequencies, and commercial age frequencies. Asymptotic selectivities were assumed for the fishery and the survey. The model has primary parameters of F_{MSY}, *MSY*, and *M*; steepness and B_0 then become derived parameters. Recruitment deviates were estimated together with selectivities and nuisance parameters (see Martell, 2010).

The preliminary base models offered by the STAT were not acceptable because of serious data issues. To rectify this it was necessary to remove data which could not be reliably interpreted by the stock assessment models. The 1986 and 2009 acoustic biomass estimates were biased to an unknown extent due to calibration issues in 1986, and the presence of large quantities of Humboldt squid in 2009. Also, the acoustic trawl composition data could not be adequately modelled because it was derived from trawl samples obtained in the absence of any statistical design. Once these data were removed, two defensible base models were available. However, only the TINSS model was able to deliver a full Bayesian MCMC run within the time frame of the review. For this reason, the Panel chose the TINSS model as the preferred base run. It provides a very uncertain assessment of stock status, but, I believe, it constitutes the best science that was available within the timeframe of the review.

There is a lot that can be done to improve future stock assessments. With regard to the acoustic surveys, the relatively simple exercise of constructing an informed prior for the *q* associated with recent surveys (1995-2007) will reduce the uncertainty of stock assessment estimates. The acoustic indices provide the only direct abundance data, but they currently contain very little information because of uncertainty with regard to hake target strength and because the available ancillary information on the acoustic *q*(s) is not being used in the assessment (through properly constructed informed priors).

Background

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 assessments that had been conducted by Stewart and 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).

I was one of two CIE reviewers appointed to the STAR Panel. An SSC representative chaired the meeting and a Canadian reviewer was also on the Panel (*see* Appendix 3 for a list of participants). This report presents my review findings and recommendations in accordance with the Terms of Reference for the review (Appendix 2, annex 2). My views are consistent with those expressed in the STAR Panel report, which contains the unanimously agreed findings and recommendations of the Panel.

Review Activities

Pre-meeting

Meeting documents and materials were made available in electronic form in advance of the meeting (*see* Appendix 1). I familiarized myself with the background material and read the main assessment documents in detail prior to the meeting.

Meeting

The meeting began on schedule and generally followed the agenda during the three days (Appendix 2, annex 3). On the first day there were presentations on the 2009 fishery, the acoustic survey methodology and the 2009 acoustic survey, future plans for acoustic data storage and analysis, and the SS3 preliminary base model. The TINSS model and results were presented the following morning; as were some requested analyses for the SS3 model. The remainder of the meeting was primarily devoted to a series of requests (from the Panel) and presentations (by the STAT) with regard to additional model runs using alternative input data or model assumptions. A request was also made of the acoustic team for clarification of the potential impact of Humboldt squid on the 2009 acoustic biomass index.

My reading of the assessment and background documents before the meeting had given me the impression that there were serious data issues that needed to be dealt with before defensible model runs could be produced. At the end of the first day, I expressed my concerns to the meeting. It was clear to everyone at the meeting that there were issues with the 2009 abundance index due to the substantial quantity of squid observed during the survey. However, I also indicated that I saw problems with the use of the composition data collected during the survey. At issue was the use of opportunistic trawling to sample hake marks (i.e., back-scatter/echo-sign) which could be expected to contain substantial length and/or age structure. In the absence of a justification for how such sampling could yield a consistent time series of composition data I was reluctant to endorse its use. I also noted that the younger age classes of hake, that are vulnerable to the US fishery, were growing appreciably during the fishing season, which compromised the use of the conditional age-at-length data in the SS3 model.

Partway through the second day, after all of the scheduled meeting presentations had been made, the Chair called for a brief recess of the meeting so that the STAR Panel could privately discuss the data issues. The first question we addressed was whether there were adequate data available, in a suitable form, with which to conduct a defensible stock assessment. We agreed that there were. Second, we addressed how the existing data could best be used given the available stock assessment models. We concluded that the 2009 survey biomass estimate was biased to an unknown extent and should not be used in a base model. Also, we concluded that the survey composition data could not be adequately modelled with a constant selectivity and therefore should be excluded from any base model. We also concluded that attempts, in SS3, to fit the length frequency data and conditional age-at-length data were problematic because of fish growth during the fishing season. Therefore, we opted for the use of age frequencies constructed by sampling directly for age (i.e., using the otolith samples only).

When the meeting re-convened, I volunteered to communicate our decision to the meeting. Our decision, especially in regard to the exclusion of the survey composition data, was greeted by a mixture of surprise, shock, and, from some participants, anger. As the messenger of the decision, and having been the first person to raise the issue, much of the anger was directed at me. However, it should be noted that it was a joint and unanimous decision of the STAR Panel.

On the morning of the final day of the meeting, members of the acoustic team made presentations in support of using the survey composition data. They offered two arguments. First, that it was common practice to collect length and age data during acoustic surveys. Second, that the length composition of hake catches changed only slowly with latitude during the survey. The first argument was irrelevant; what matters is whether the data are routinely and *justifiably* used in stock assessment model runs. The second argument was not really supported by the data and also somewhat missed the point that the trawl catches are not necessarily representative of what was in the marks. The STAR Panel's position remained unchanged.

Also, on the final day, the STAT presented the SS3 and TINSS model runs using the Panel's preferred specifications. The Panel again conferred privately and agreed that the TINSS model should be adopted as the Panel's preferred base model. This was primarily because the full Bayesian model run was available for the TINSS model and was not available for the SS3 model. Our decision was explained to the meeting (by our Canadian member) and was greeted with some surprise. We explained that we were happy with both models but we could not endorse a stock assessment which we had not seen (the full Bayesian SS3 model run could not be produced before the end of the meeting because it required a day or more of run-time).

Just before the meeting closed, we agreed amongst ourselves who would draft the different sections of the STAR Panel report.

Post-meeting

I had volunteered to draft a large proportion of the STAR Panel report. I remained in Seattle on the day after the meeting and produced a draft of my sections of the report and circulated it to Panel members by email (before flying home the next day). The Chair compiled a complete draft of the report using the contributions from Panel members and also contributed the "Overview" section. I reviewed the draft and offered a much more concise overview section to the Panel. Via email, we eventually agreed on a final draft which was received by the PFMC representative very close to the deadline. The final version needed more editing and checking but we ran out of time.

There was considerable email correspondence from members of the SSC and others just before the STAR Panel report was finalized. I was copied in on some of the emails. They concerned two matters.

First, there initially appeared to be a challenge to the STAR Panel recommendations by the US members of the STAT, despite their agreement at the meeting that there were no significant points of disagreement between the Panel and the STAT. Subsequently, the US members of the STAT said that they were not challenging the recommendations. However, they drew a distinction with regard to points of disagreement: those that needed to be noted in the STAR Panel report, and those that existed. They said that had the SS3 model been chosen then they would have recorded a disagreement with the Panel's recommendations in terms of input data and model structure. However, because the TINSS model was chosen, they had not recorded any disagreement.

Second, some members of the SSC expressed their concern that so much data had been omitted from the Panel's preferred runs. They thought that STAR Panels should perhaps not be allowed to remove so much data and suggested that data issues might be dealt with in another forum. I agree that there should be much closer scrutiny of the suitability of data for stock assessment purposes than now currently appears to occur. However, it is difficult to imagine a STAR Panel being able to sensibly review a stock assessment where the data are prescribed and the *use* of the data with a *particular* stock assessment model is not subject to review. There are some data which should never be used in a stock assessment model, but not all "good" data are appropriate for use in every model.

Summary of findings

Each of the Terms of Reference are considered below.

1. *Become familiar with the draft Pacific hake/Whiting stock assessment(s) and background materials.*

The background material and draft assessments were provided in a timely manner and the documentation was generally to a high standard. However, the documentation on the 2009 acoustic survey and on the surveys as a time series (e.g., considering issues of comparability) was scant. Of recent surveys, only the 2005 survey appears to have been documented in any detail, and only in draft form (Fleischer undated). During the meeting we were supplied with additional acoustic documentation: a paper on Humboldt squid target strength estimates and a document on acoustic survey protocols.

2. *Comment on the quality of data used in the assessment(s) including data collection and processing.*

There were three main sources of data available to the assessment models. The catch history, commercial catch sampling data (length and age), and acoustic survey data (total biomass, length and age, 1977-2009, every three years until 2001, then every two years).

The protocols for sampling commercial catch were not discussed in detail but I am sure that they conform to acceptable practice. Likewise, the trawling and catch sampling protocols during the acoustic survey are probably acceptable in terms of sampling for species composition and hake length frequencies in support of hake biomass estimation. However, it should be noted that trawls are aimed at obtaining a "sample" of fish (ideally about 300) from the targeted mark and large catches are avoided (Fleischer undated). It appears that many of the trawls are "dips" into the marks, or, at least, are of short duration to avoid making large catches. It seems very likely that this approach will lead to a domed shaped selectivity with regard to length (i.e., larger fish are under-represented in the trawl catches). This could create some problems, but it is a generic issue for most acoustic surveys.

The acoustic survey design is archaic having just a single element of statistical design (a random starting position for the first transect), but it is adequate. Ideally, the survey would be pre-stratified (rather than post stratified) and there would be a variable amount of effort across strata (with more effort in those strata with higher variability in hake biomass). However, the proposed move to analyze the data using geo-statistical methods, and the fact that the time series has continued for so long with the same design, suggests to me that a change in design may not be beneficial (perhaps threatening comparability with the recent surveys).

There are two points in the existing time series which have major quality issues. In 1986, the pre- and post-survey calibration results differed by a factor of approximately 1.5. The pre-survey calibration was used to produce the biomass index, primarily it seems because it gives the lower estimate (no documentation supplied). However, the magnitude of the bias associated with this estimate has not been quantified beyond the probable range of 0-50%. In 2009, the extensive presence of squid in the survey area introduces an unknown bias in the biomass index. Again, this has not been quantified in any way other than to produce estimates with the best guess at partitioning hake and squid biomass (1.462 mmt) and another from marks "confidently" identified as hake (0.87 mmt).

The production of length frequencies and conditional age-at-length data from the acoustic trawl and backscatter data is very problematic. Each transect is assigned a hake length frequency from combining the length samples from a number of hauls which have similar length frequencies and are geographically related (Chu and Thomas, 2010). The scaled length frequencies are then produced by weighting across transects using estimated hake numbers on each transect (and age-at-length data are scaled in a similar way). The post-stratification and the partly subjective method of assigning length distributions to transects is less than ideal. It is probably adequate for producing total hake biomass indices (assuming the slope of the length to target strength relationship is actually correct – it is assumed to be 20; and the selectivity is not too domed shaped). However, it is probably not adequate for producing a reliable overall length frequency (or age-at-length data) for the portion of the population vulnerable to the survey trawling in any given year. Note, this issue is not the main problem with using the composition data in stock assessment runs – it is an additional problem. It is discussed further under TOR 3 below.

The construction of length frequencies, age frequencies, and conditional age-at-length data for the commercial fisheries is not ideal. The only stratification appears to be by country and at-sea or shore-based processing within country (Stewart and Hamel, 2010). These data need to be analyzed in detail to determine an appropriate stratification to produce properly scaled composition data for assessment (e.g., temporal and spatial strata, and vessel type/size could all be important). An output of the analysis will be a suitable definition of sub-fleets/fisheries – where for each sub-fleet it can be expected that the fishery selectivity for each sub-fleet stays relatively constant over time. Sub-fleets may also have a temporal component to their definition (e.g., the surimi fleet only existed for a few years). Catch histories, as well as composition data, need to be prepared for each sub-fleet over the period that they existed in the fishery.

3. Evaluate and comment on analytic methodologies.

The preparation of composition data from the commercial fisheries involves analytic methods which have been commented on above. Finer-scale stratification is needed, and the scaling needs to be done by number (rather than catch weight) in each step of the scaling (currently, it appears that samples are scaled-up by estimated number to trawl/landing numbers, but then by catch weight across strata).

The calculation of biomass indices and composition data from the acoustic surveys has also been commented on above in terms of data preparation. There are others issues in regard to using the length compositions to calculate the numbers and biomass associated with each transect (transects are 10 nautical miles apart; the average densities on each transect are assumed to extend 5 nautical miles to each side of the transect – so each transect does correspond to an absolute estimate of hake numbers and biomass). The average hake backscatter on each transect is converted to density in numbers by dividing by the average back-scattering cross section from the length-target strength relationship (Traynor, 1996). Biomass is then obtained by multiplying by the mean weight from the estimated length-weight relationship (Chu and Thomas, 2010).

The calculation of mean weight and mean back-scattering cross section (ts) both use the length frequency distribution assigned to the transect. Since the trawling is likely to under-estimate the proportion of large fish in any hake mark (see discussion in the previous section), the estimated numbers in each transect is likely over-estimated (because ts is under-estimated), and, to a lesser extent, so is the biomass (the ratio of mean weight to mean ts is proportional (~) to length, since weight ~ $length^3$ and ts ~ $length^2$). This assumes that the slope of the length-target strength relationship really is 20, which is not necessarily a good assumption (see McClatchie et al., 2003). In any case, the point is that the likely domed shaped selectivity of the trawling confounds, to some extent, the biomass estimates. There is perhaps little point in investigating this further until a defensible length-target strength relationship for hake is established.

A related issue is the combining of length samples across trawls to form a single length frequency which is then assigned to various transects. The raw lengths are all weighted equally, whether they were obtained from a trawl on a dense and extensive mark or one of lower density and extent. The samples are not self weighting, because tow duration and the type of tow is changed according to the density of the mark to obtain a "sample" (of ideally about 300 fish). This is another generic problem for an acoustic survey with an objective to estimate representative length frequencies from targeted trawling. The appropriate weighting of length samples depends on the relative numbers of fish in the marks from which they were obtained; and the numbers of fish in a mark can only be accurately estimated from a representative length frequency. This confounding will apparently be alleviated to some extent by the post-stratification which grouped trawls with similar length samples. However, the post-stratification has subjective components and only helps if the trawl really is obtaining representative length samples from the marks.

The stock assessment models are constructed from equations describing population dynamics and statistical assumptions with regard to the data. The structural and statistical assumptions of the models were not presented in any detail. However, the population dynamics in both models are based on well established equations. The distributional assumptions are tested when the models are fitted to the data – which is perhaps the best test of the assumptions. Low-level, technical details of the models have not been reviewed due to time constraints.

4. *Evaluate model assumptions, estimates, and major sources of uncertainty and provide constructive suggestions for improvements if technical deficiencies or additional major sources of uncertainty are identified.*

The preliminary base models brought to the meeting by the STAT are discussed first. Then the STAR Panels preferred models are considered.

STAT preliminary base models

The SS3 model had Canadian and US fisheries and was fitted to the acoustic time series of biomass indices, the survey length frequencies and conditional age-at-length data, and commercial length frequencies and conditional age-at-length. Time-varying selectivity was estimated in the commercial fisheries together with time-varying growth. Domed selectivities were estimated for both fisheries and the acoustic data, together with steepness, a senescent natural mortality (M) for fish older than 14 years, and recruitment deviations (see Stewart and Hamel, 2010 for details).

The TINSS model had simpler population dynamics with just a single fishery and was fitted to the acoustic biomass indices and age frequencies, and commercial age frequencies. Asymptotic selectivities were assumed for the fishery and the survey. The model has primary parameters of F_{MSY}, MSY, and M; steepness and B_0 then become derived parameters. Recruitment deviates were estimated together with selectivities and nuisance parameters(*see* Martell, 2010).

Both models treated the acoustic data inappropriately in a number of respects. First, the 2009 and 1986 biomass indices were used by both models. Given the data quality issues where both indices are known to be biased (but the magnitude of the bias is unknown) these indices should not have been used. The approach, taken by the STAT, of giving these indices a higher CV than the other indices, which were considered to be unbiased, is technically flawed: bias and variance are different concepts.

Second, as noted by Dr. Mark Maunder in a 2009 review (*see* Stewart and Hamel, 2010) the selectivity associated with the acoustic backscatter and the selectivity associated with the survey trawling cannot legitimately be assumed to be the same (as was done in both models). Different processes are involved. The trawling can be expected to have a domed shaped selectivity: not all young fish are there and older (larger) fish are less likely to be caught. The backscatter is more likely to be asymptotic: not all the younger fish are there. It could be argued that it is domed to some extent – perhaps the older/larger fish tend to be near the bottom, perhaps even in the shadow zone – but the backscatter selectivity could be very different from the trawl selectivity.

Third, the early surveys (1977-1992) covered different depth zones and different latitudes to the later surveys. The estimates had been "corrected" by applying "expansion factors" (*see* Helser et al., 2004) and were assumed, in the preliminary models, to be comparable to the other indices. However, the dangers of extrapolation are well known; borrowing data from other years to correct for potential biomass that was not observed is not appropriate. These early surveys, even with "corrections", cannot be considered comparable to the later surveys. The Panel agreed with Dr. Maunder (*see* Stewart and Hamel, 2010) who recommended that the early surveys be put into a separate time series (this is a pragmatic decision given that there wasn't time to do a detailed analysis of which surveys were comparable to each other; in some cases the limited northern extent may not matter if the fish distribution is concentrated in the south).

Finally, the composition data from the acoustic survey were obtained by opportunistic trawling at the discretion of the voyage leader, targeted on marks, aimed at getting a sample of fish, and avoiding large catches (Fleischer, undated). This was done with mid-water trawl gear (predominately), from two different vessels, and any number of different voyage leaders and fishing operators. Sometimes trawls were "dipped" into the marks; sometimes they were perhaps targeted under the mark (to catch the fish as they dive); sometimes tows were long and sometimes they were short. Sometimes marks were in mid-water and sometimes they were on the bottom. Sometimes the tows were over deep water, sometimes shallow water. All of these factors can influence the species composition and the hake length frequency obtained in a trawl catch. No data were presented with regard to the consistency of the methods or in support of the model assumption of constant selectivity over time.

It is very difficult to see how such an approach could lead to a consistent time series of composition data. It is very easy to see how it could fail to do so. All that is needed is a change in the proportion of tows that target under the mark as opposed to dipping into the mark (for many species, larger fish are better at avoiding trawl gear as they dive more strongly). Alternatively, or as well as, a change in the proportion of deep-water or shallow-water tows could have an impact. A change in the vertical distribution of the marks would also impact on the selectivity (e.g., more marks near the bottom make the larger fish easier to catch as fish cannot dive under the bottom). Each survey will have its own particular trawling selectivity which will depend on the nature of the marks in that year and the decisions made by the voyage leader(s) and the fishing operator(s).

The commercial catch histories and composition data were treated differently in the two models. The SS3 model assumed two fisheries and fitted to length frequencies and conditional age-at-length data within each fishery. The TINSS model had just a single fishery and only fitted to age frequencies (generated from otolith data independently of the length samples; i.e., not length frequencies with an age-length key, but sampling directly for age).

In the TINSS model, a time-invariant asymptotic fishery selectivity was assumed. This is a very simple assumption which is undoubtedly violated. However, it may be that this simple approach provides a reasonable estimate of a mean selection process which enables annual recruitment strengths to be adequately estimated (through the commercial age frequencies, conditioned by the other data).

The SS3 approach was far more complex. Two fisheries were modelled, each with a time varying domed-shaped selectivity, together with time-varying growth, and senescence natural mortality. The fits to the length frequency data showed very bad residual patterns (see Stewart and Hamel, 2010). This is to be expected, despite all of the time-varying parameters which were estimated, because the model is essentially age based and does not keep track of the length frequency of the population over time.

Predicted length frequencies are produced mid-year in the years when there are observations, assuming that they are normally distributed about a mean length with a know CV. In reality, the length samples were obtained throughout the fishing season as the fish were growing (i.e., the spread of length at age includes growth as well as natural variation). Also, the fishery operates with predominantly length-based selection (which will have changed over time as fleet composition changed) which alters the length distribution of the population and can give rise to spurious patterns of apparent changes in growth (e.g., if larger fish are removed preferentially then samples early in a fisheries history show larger mean size at age and later samples show smaller mean size at age despite the size at age remaining constant). The growth of fish during the season also compromises the use of the conditional age-at-length (because the age-proportions at given length can change dramatically during the season for the 3-5 year old fish caught in the US fishery).

The failure of the SS3 model to adequately fit the length frequencies and the technical problem with fish growth compromising the conditional age-at-length data suggests that it is better to replace these two problematic data sets with the associated age frequencies.

STAR Panel preferred models

For the reasons discussed above, the preliminary base models provided by the STAT were not considered acceptable by the STAR Panel. The 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 *q*s): 1977-1992, 1995-2007, with standard deviations in log space constant within each series: 0.5 and 0.25 respectively.
- SS3: Remove length frequencies and conditional age-at-length; replace with age frequencies

The Panel was requested to specify a single preferred base model. Since the full Bayesian MCMC run with diagnostics was available for the TINSS model and was not for the SS3 model, the Panel chose the TINSS model. However, if the full Bayesian run for the SS3 model has satisfactory diagnostics, then it would be equally acceptable.

A summary of the technical merits and deficiencies of both models is given below:

Technical merits:

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:

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 *q*s 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 was not available to the Panel before the finalization of the review due to time constraints).
- The model does not have informed priors for the acoustic *q*s which limits our ability to judge the plausibility of the estimated size of the stock

Estimates

Point estimates of 2010 depletion, 2010 projected OY, the acoustic *q*s, steepness (*h*), and natural mortality (*M*) for the Panel's preferred model runs are given below with the STAT's preliminary SS3 base model (SS3 update). The preferred base-model point estimates are the TINSS median (from a full Bayesian MCMC run); the modes of the posterior distribution (MPD) are given for comparison with the SS3 runs.

	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
*q*s	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

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 *q*s are not comparable between the SS3 update run (being the STAT team's preliminary SS3 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 *q*s but they were not reported – but are probably very similar. The second *M* for the update run is for senescence.

The relatively tight range of the point estimates across the runs in the table hides the large degree of uncertainty in the assessment. The Panel requested sensitivity runs for both models (MPD for SS3 and MCMC for TINSS). For SS3, the dimension of uncertainty explored was a combination of values of *M* and domed versus asymptotic fishery-selectivity. The range of estimated 2010 depletion for these MPD sensitivity runs was 15-42%. For the TINSS model, the dimension of uncertainty explored was alternative priors on F_{MSY} and *MSY*, which had little effect on estimated 2010 depletion but gave a range on OY of 250,000-400,000 t (for the medians of the posteriors). In the base TINSS model the 95% credibility intervals on 2010 depletion and was 13-79%.

Unresolved problems and major uncertainties

The main problems are the lack of information in the data – this is a data rich, yet information poor, stock assessment.

- 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. The scale of the stock, in absolute terms, is very poorly determined, as is the level of 2010 depletion.
- 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).

5. *Determine whether the science reviewed is considered to be the best scientific information available.*

The Panel's preferred base models use a defensible data set and model assumptions and hence, I believe, constitute the best scientific information available on which to base management decisions. The Panel chose the TINSS model because the full Bayesian run was available and was reviewed. However, if the diagnostics for the SS3 full Bayesian model are adequate then that run is equally acceptable.

6. *Provide specific suggestions for future improvement in any relevant aspects of data collection and treatment, modeling approaches and technical issues.*

This is covered under "Conclusions and Recommendations" below

7. *Provide a brief description on panel review proceedings highlighting pertinent discussions, issues, effectiveness, and recommendations*

This was covered under "Review Activities" above.

Discussion

Data and information

It is critically important in statistics and stock assessment to understand the distinction between data and information. The information content of a data set is not easily measured unless the data are obtained from simple random sampling (e.g., *n* independent and identically distributed random variables). In a fisheries stock assessment, data sets consist of a number of time series, obtained from complex sampling schemes. They are fitted by models which make numerous assumptions with regard to population dynamics and the structural and statistical relationship of the data to the model population. The apparent information content of data, as measured by model outputs, must be interpreted cautiously, as the information content is *conditional* on the model assumptions. Be very wary of statements such as, "the data tell us …". The data will generally tell us little until we make certain structural assumptions; change the structural assumptions and the data may tell us a different story.

The approach, adopted by some stock assessment scientists, whereby all available data are fitted in the model is based on a belief that "more data gives more information". This is clearly not the case when some of the data are badly biased, or cannot be adequately fitted by the model, or if contradictory data are used (e.g., Schnute and Hilborn, 1993).

The question that should be asked when deciding which data to use in a stock assessment model is whether the inclusion of the data will provide a "better stock assessment". That is, will it provide more accurate estimates that will help managers make better decisions.

Acoustic data

The acoustic surveys span more than 30 years and have consumed a large amount of resources over that time. Therefore, people are very keen that the most be made of the available data. Unfortunately, it is not the case that the information content of data is directly related to the quantity of data or the cost of collecting it. The extensive trawling that was performed during the acoustic surveys was necessary for target identification and to provide length frequencies of hake for biomass estimation. The trawling effort was not wasted. However, the trawl data are not adequate to provide a quantitative time series of composition data suitable for use in a stock assessment model which assumes time-invariant selectivity. The issue for stock assessment is that the changes in selectivity from year to year cannot be adequately modelled with the existing data. It may be, that with sufficient analysis of the nature of the trawling that occurred in each survey, that a method for estimating annual survey selectivity could be developed. However, with annual changes in selectivity, the information content of the data is weakened. It is probably better to let the recruitment-strength signals, in the stock assessment models, be driven by the commercial catch data (where there is a much better chance of modelling changes in selectivity), rather than be contaminated by dubious signals from the acoustic composition data.

This Panel was the first to exclude the acoustic composition data from base model runs. However, there has been increasing concern about these data from recent STAR Panels:

> 2008 STAR Panel report: "It was disconcerting to learn that the acoustic survey biomass estimates are based on very sparse sampling to establish the species, size and age composition of the acoustic signs. While it is accepted that this is typical of acoustic surveys, it would have been reassuring to have been shown some evidence that a single short tow from a long acoustic transect provides a reliable and unbiased estimate of the species, size, and age composition of identified fish aggregations."
> 2009 STAR Panel: "There is concern that some of the input data may be biased. The STAR Panelists suspect, in particular, that the acoustic survey age- and length-compositions may be biased because of a tendency for biological sampling to occur disproportionately on dense aggregations of fish that may not be representative. The raw acoustic survey data need to be analyzed to allow verification that an appropriate stratification was applied. Additionally, there need to be explicit rules regarding how the age and length data are collected in the survey and an explicit recounting of the rules that applied in past surveys. The methods for combining length samples into strata needs further review. A post-stratification scheme that creates more homogeneous strata by pooling tows with catches of similar length structure is not justified. This procedure could bias the estimates of length compositions applied to the acoustic survey tracks, and grossly overestimate the precision in estimated length compositions. "

The acoustic team responded to the criticisms of the 2009 Panel in Stewart and Hamel (2010). I was not convinced by their response which included the unlikely statement: "While the biological sampling is not completely random, the trawls tend to occur at points of the most density, and the trawls are thus representative of about 99% of the hake observed by the acoustic survey."

Commercial length frequencies and conditional age-at length

The approach used in the STAT's preliminary SS3 base model was to fit as much data as possible in a disaggregated form (albeit, not disaggregated by sex – they are proposing to do that next year). This is a reasonable approach if the data are carefully prepared and the model is structured so that it could be expected to fit the data. Unfortunately, the data were not carefully analysed or appropriately stratified and scaled. Also, the fish were growing through the fishing season, so that mid-year predictions of length frequencies and age-at-length could not really be expected to fit the data (which, in reality, includes growth). Finally, the fisheries have predominately length-based selection processes and the model does not track the length distribution of the population (so, again, it was unlikely that the length frequencies would be well fitted). Indeed, the length frequencies were very badly fitted as evidenced by the residual patterns, despite the use of time-varying growth and selectivity (*see* Stewart and Hamel, 2010).

The bad lack of fit shows that the model assumptions were not satisfied and suggests that it is unwise to use the model to provide stock assessment advice.

In the Panel's recommended SS3 base model, the length frequencies and age-at-length data were replaced with age frequencies obtained directly from otolith samples (avoiding the growth issue). The subsequent fits to the model data were reasonable, although the SS3 model was over-parameterized since the time-varying selectivity blocking structure had not been revisited.

Burden of proof

This review has considered technical aspects of data preparation and model interpretation of data. It is shown that there are technical deficiencies with model interpretation of the data that were used in the STAT's preliminary base models. These data were removed in the STAR Panel's preferred models. It is not possible, given the time limits on this review, to show conclusively that the STAR Panel's preferred models constitute better stock assessments. However, there is no doubt that they are technically more defensible. Some may argue that too much data were removed. However, I suggest that quantity is no substitute for quality and that it is for the proponents of "more data always gives more information" to demonstrate that such an approach delivers a better stock assessment in this case.

Critique of the NMFS review process

The STAR process used for the hake review was very similar to other STAR meetings that I have participated in as a CIE reviewer. As is often the case, the STAT's preliminary base models were found to be technically deficient because of data issues or model assumptions. The Panel typically spends much of the meeting constructing acceptable runs rather than reviewing the runs that were offered by the STAT, or, reviewing the technical details of the model equations. This meeting was no exception.

However, it has been my experience that this process does lead to much improved stock assessments. Also, the critically intensive process does provide a spur for incremental improvements in analytical and stock assessment methods.

Conclusions and Recommendations

The preliminary base models offered by the STAT were not acceptable because of serious data issues. To rectify this it was necessary to remove data which could not be reliably interpreted by the stock assessment models. Once this was done, two defensible base models were available. However, only the TINSS model was able to deliver a full Bayesian MCMC run within the time frame of the review. For this reason, the Panel chose the TINSS model as the preferred base run. It provides a very uncertain assessment of stock status, but, I believe, it constitutes the best science that was available within the timeframe of the review.

There is a lot that can be done to improve future stock assessments. With regard to the acoustic surveys, the relatively simple exercise of constructing an informed prior for the q associated with recent surveys (1995-2007) will reduce the uncertainty of stock assessment estimates. A similar recommendation was made by the 2004 STAR Panel and subsequent panels. It is lamentable that there has been no progress reported on this as yet. Likewise, progress on estimating hake target strength has been very limited with the unreliable estimate of Traynor (1996) still being used. The acoustic indices provide the only direct abundance data, but they currently contain very little information because of uncertainty with regard to hake target strength and because the available ancillary information on the acoustic q(s) is not being used in the assessment (through properly constructed informed priors).

I support the recommendations from the STAR Panel report, which are substantively the same as those below:

- A detailed spatial and temporal 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).
- 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 *q*s 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.

References

Helser, T.E; Methot, R.D.; Fleischer, G.W. 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.

McClatchie, S.; Macaulay, G.; Coombs, R. 2003. A requiem for the use of 20 log10 Length for acoustic target strength with special reference to deep-sea fishes. *ICES J. Mar. Sci., 60*: 419 - 428.

Schnute, J.T.; Hilborn, R. 1993. Analysis of contradictory data sources in fish stock assessment. *Can. J. Fish. Aquat. Sci. 50*: 1916–1923.

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: Bibliography of supplied material

Anon. 2004. Draft USA – Canada Pacific hake agreement. 16 p.

Anon. 2007. Pacific whiting Magnuson-Stevens fishery conservation and management reauthorization act (p.l. 109-479). 6 p.

Chu, D.; Thomas, R. 2010. Integrated acoustic and trawl survey: design, method, and analysis. Draft February 2010. 20 p.

Fleischer, G.W. undated. The 2005 integrated acoustic and trawl survey of Pacific hake, *Merluccius productus*, in U.S. and Canadian waters off the Pacific coast. Draft. 59 p.

Hamel, O.; Stewart, I. 2009. Stock Assessment of Pacific hake, *Merluccius productus*, (a.k.a. Whiting) in U.S. and Canadian waters in 2009. 246 p.

Helser, T.; Stewart, I.; Hamel, O. 2008. Stock Assessment of Pacific hake, *Merluccius productus*, (a.k.a Whiting) in U.S. and Canadian waters in 2008. 128 p.

Martell, S. 2009. Assessment and management advice for Pacific hake in U.S. and Canadian waters in 2009. DFO Can. Sci. Advis. Sec. Res. Doc. 2009/021. iv + 54 p.

Martell, S. 2008. Assessment and management advice for Pacific hake in U.S. and Canadian waters in 2008. 56 p.

Martell, S. 2010. Assessment and Management advice for Pacific hake in U.S. and Canadian waters in 2010. Draft January 26, 2010. 91 p.

Martell, S.; Pine III, W.; Walters, C. 2008. Parameterizing age-structured models from a fisheries management perspective. *Can. J. Fish. Aquat. Sci. 65*: 1586-1600.

Methot, R.D. 2010. User Manual for Stock Synthesis Model Version 3.10. Updated Jan 12, 2010. 163 p.

PFMC 2008. Terms of reference for the groundfish stock assessment and review process for 2009-2010. 35 p.

Sampson, D.; et al. 2008. Report of the 2008 U.S. / Canada Pacific Hake (Whiting) Stock Assessment Review (STAR) Panel. 33 p.

Sampson, D.; Haddon, M.; Cadigan, N.; 2008. Addendum to the Report of the 2008 U.S. / Canada Pacific Hake (Whiting) Stock Assessment Review (STAR) Panel - STAR Panel Response to the Minority Report. 6 p.

Sampson, D.; et al. 2009. Pacific Whiting The Joint U.S.-Canada STAR Panel Report. 22 p.

Sinclair, A.F.; Grandin, C.J. 2008. Canadian Fishery Distribution, Index Analysis, and Virtual population Analysis of Pacific Hake, 2008. 61 p.

Sinclair, A.; Martell, S.; Grandin, C.; Fargo, J. 2008. Minority report to the 2008 Pacific hake STAR Panel report. 5 p.

Stewart, I.; Hamel, O. 2010. Stock Assessment of Pacific hake, *Merluccius productus*, (a.k.a. Whiting) in U.S. and Canadian waters in 2009. Draft *01/25/2010.* 290 p.

Appendix 2: Statement of Work for Patrick Cordue

External Independent Peer Review by the Center for Independent Experts

Joint US-Canada Technical Review Panel for the Pacific Whiting Stock Assessment

Scope of Work and CIE Process: The National Marine Fisheries Service's (NMFS) Office of Science and Technology coordinates and manages a contract providing external expertise through the Center for Independent Experts (CIE) to conduct independent peer reviews of NMFS scientific projects. The Statement of Work (SoW) described herein was established by the NMFS Project Contact and Contracting Officer's Technical Representative (COTR), and reviewed by CIE for compliance with their policy for providing independent expertise that can provide impartial and independent peer review without conflicts of interest. CIE reviewers are selected by the CIE Steering Committee and CIE Coordination Team to conduct the independent peer review of NMFS science in compliance the predetermined Terms of Reference (ToRs) of the peer review. Each CIE reviewer is contracted to deliver an independent peer review report to be approved by the CIE Steering Committee and the report is to be formatted with content requirements as specified in **Annex 1**. This SoW describes the work tasks and deliverables of the CIE reviewer for conducting an independent peer review of the following NMFS project. Further information on the CIE process can be obtained from www.ciereviews.com.

Project Description: The Pacific hake (or whiting, *Merluccius productus*) stock assessment will provide the basis for the management of the largest groundfish fisheries off the West Coast of the U.S. and British Columbia. In 2008, Pacific whiting fishery accounted for 89% of the landed catch and 52% of the associated ex-vessel value in the U,S. groundfish fishery. In addition, the treaty between the U.S. and Canada which establishes an annual assessment and management process is expected to be ratified sometime soon. The technical review will take place during a formal, public, multiple-day meeting of fishery stock assessment experts. Participation of external, independent reviewer is an essential part of the review process. The Terms of Reference (ToRs) of the peer review are attached in **Annex 2**. The tentative agenda of the panel review meeting is attached in **Annex 3**.

Requirements for CIE Reviewers: Two CIE reviewers shall conduct an impartial and independent peer review in accordance with the SoW and ToRs herein. CIE reviewers shall have working knowledge and recent experience in the application of fish population dynamics, with experience in the integrated analysis modeling approach, using age-and size-structured models, use of MCMC to develop confidence intervals, and use of Generalized Linear Models in stock assessment models. Each CIE reviewer's duties shall not exceed a maximum of 14 days to complete all work tasks of the peer review described herein.

Location of Peer Review: Each CIE reviewer shall conduct an independent peer review during the panel review meeting tentatively scheduled in Seattle, Washington during February 8- 10, 2010.

Statement of Tasks: Each CIE reviewers shall complete the following tasks in accordance with the SoW and Schedule of Milestones and Deliverables herein.

Prior to the Peer Review: Upon completion of the CIE reviewer selection by the CIE Steering Committee, the CIE shall provide the CIE reviewer information (full name, title, affiliation, country, address, email) to the COTR, who forwards this information to the NMFS Project Contact no later the date specified in the Schedule of Milestones and Deliverables. The CIE is responsible for providing the SoW and ToRs to the CIE reviewers. The NMFS Project Contact is responsible for providing the CIE reviewers with the background documents, reports, foreign national security clearance, and other information concerning pertinent meeting arrangements. The NMFS Project Contact is also responsible for providing the Chair a copy of the SoW in advance of the panel review meeting. Any changes to the SoW or ToRs must be made through the COTR prior to the commencement of the peer review.

Foreign National Security Clearance: When CIE reviewers participate during a panel review meeting at a government facility, the NMFS Project Contact is responsible for obtaining the Foreign National Security Clearance approval for CIE reviewers who are non-US citizens. For this reason, the CIE reviewers shall provide requested information (e.g., first and last name, contact information, gender, birth date, passport number, country of passport, travel dates, country of citizenship, country of current residence, and home country) to the NMFS Project Contact for the purpose of their security clearance, and this information shall be submitted at least 30 days before the peer review in accordance with the NOAA Deemed Export Technology Control Program NAO 207-12 regulations available at the Deemed Exports NAO website: http://deemedexports.noaa.gov/sponsor.html).

Pre-review Background Documents: Two weeks before the peer review, the NMFS Project Contact will send (by electronic mail or make available at an FTP site) to the CIE reviewers the necessary background information and reports for the peer review. In the case where the documents need to be mailed, the NMFS Project Contact will consult with the CIE Lead Coordinator on where to send documents. CIE reviewers are responsible only for the pre-review documents that are delivered to the reviewer in accordance to the SoW scheduled deadlines specified herein. The CIE reviewers shall read all documents in preparation for the peer review.

Panel Review Meeting: Each CIE reviewer shall conduct the independent peer review in accordance with the SoW and ToRs, and shall not serve in any other role unless specified herein. **Modifications to the SoW and ToRs can not be made during the peer review, and any SoW or ToRs modifications prior to the peer review shall be approved by the COTR and CIE Lead Coordinator.** Each CIE reviewer shall actively participate in a professional and respectful manner as a member of the meeting review panel, and their peer review tasks shall be focused on the ToRs as specified herein. The NMFS Project Contact is responsible for any facility arrangements (e.g., conference room for panel review meetings or teleconference arrangements). The NMFS Project Contact is responsible for ensuring that the Chair understands the contractual role of the CIE reviewers as specified herein. The CIE Lead Coordinator can contact the Project Contact to confirm any peer review arrangements, including the meeting facility arrangements.

<u>Contract Deliverables - Independent CIE Peer Review Reports</u>: Each CIE reviewer shall complete an independent peer review report in accordance with the SoW. Each CIE reviewer shall complete the independent peer review according to required format and content as described in Annex 1. Each CIE reviewer shall complete the independent peer review addressing each ToR as described in Annex 2.

<u>Other Tasks – Contribution to Summary Report</u>: Each CIE reviewer may assist the Chair of the panel review meeting with contributions to the Summary Report, based on the terms of reference of the review. Each CIE reviewer is not required to reach a consensus, and should provide a brief summary of the reviewer's views on the summary of findings and conclusions reached by the review panel in accordance with the ToRs.

Specific Tasks for CIE Reviewers: The following chronological list of tasks shall be completed by each CIE reviewer in a timely manner as specified in the **Schedule of Milestones and Deliverables**.

1) Conduct necessary pre-review preparations, including the review of background material and reports provided by the NMFS Project Contact in advance of the peer review.
2) Participate during a panel review meeting in Seattle, Washington during 8-10 February 2010 as specified herein, and conduct an independent peer review in accordance with the ToRs (**Annex 2**).
3) No later than REPORT SUBMISSION DATE, each CIE reviewer shall submit an independent peer review report addressed to the "Center for Independent Experts," and sent to Mr. Manoj Shivlani, CIE Lead Coordinator, via email to shivlanim@bellsouth.net, and CIE Regional Coordinator, via email to David Die ddie@rsmas.miami.edu. Each CIE report shall be written using the format and content requirements specified in Annex 1, and address each ToR in **Annex 2**.

Schedule of Milestones and Deliverables: CIE shall complete the tasks and deliverables described in this SoW in accordance with the following schedule.

The schedule below is tentative, and the NMFS Project Contact will confirm the dates of the panel review meeting by 15 October 2009.

4 January 2009	CIE sends reviewer contact information to the COTR, who then sends this to the NMFS Project Contact
25 January 2010	NMFS Project Contact sends the CIE Reviewers the pre-review documents
8-10 February 2010	Each reviewer participates and conducts an independent peer review during the panel review meeting
24 February 2010	CIE reviewers submit draft CIE independent peer review reports to the CIE Lead Coordinator and CIE Regional Coordinator
10 March 2010	CIE submits CIE independent peer review reports to the COTR
17 March 2010	The COTR distributes the final CIE reports to the NMFS Project Contact and regional Center Director

Modifications to the Statement of Work: Requests to modify this SoW must be approved by the Contracting Officer at least 15 working days prior to making any permanent substitutions. The Contracting Officer will notify the COTR within 10 working days after receipt of all required information of the decision on substitutions. The COTR can approve changes to the milestone dates, list of pre-review documents, and ToRs within the SoW as long as the role and ability of the CIE reviewers to complete the deliverable in accordance with the SoW is not adversely impacted. The SoW and ToRs shall not be changed once the peer review has begun.

Acceptance of Deliverables: Upon review and acceptance of the CIE independent peer review reports by the CIE Lead Coordinator, Regional Coordinator, and Steering Committee, these reports shall be sent to the COTR for final approval as contract deliverables based on compliance with the SoW and ToRs. As specified in the Schedule of Milestones and Deliverables, the CIE shall send via e-mail the contract deliverables (CIE independent peer review reports) to the COTR (William Michaels, via William.Michaels@noaa.gov).

Applicable Performance Standards: The contract is successfully completed when the COTR provides final approval of the contract deliverables. The acceptance of the contract deliverables shall be based on three performance standards:
(1) each CIE report shall completed with the format and content in accordance with **Annex 1**,
(2) each CIE report shall address each ToR as specified in **Annex 2**,

(3) the CIE reports shall be delivered in a timely manner as specified in the schedule of milestones and deliverables.

Distribution of Approved Deliverables: Upon acceptance by the COTR, the CIE Lead Coordinator shall send via e-mail the final CIE reports in *.PDF format to the COTR. The COTR will distribute the CIE reports to the NMFS Project Contact and Center Director.

Key Personnel:

William Michaels, Contracting Officer's Technical Representative (COTR)
NMFS Office of Science and Technology
1315 East West Hwy, SSMC3, F/ST4, Silver Spring, MD 20910
William.Michaels@noaa.gov Phone: 301-713-2363 ext 136

Manoj Shivlani, CIE Lead Coordinator
Northern Taiga Ventures, Inc.
10600 SW 131st Court, Miami, FL 33186
shivlanim@bellsouth.net Phone: 305-383-4229

NMFS Project Contact:

Stacey Miller
NWFSC/FRAM Division
2032 SE OSU Drive, Newport OR 97365
Stacey.Miller@noaa.gov Phone: 206-437-5670

Elizabeth Clarke
NWFSC/FRAM Division
2725 Montlake Blvd. E, Seattle WA 98112
Elizabeth.Clarke@noaa.gov Phone: 206-860-5616

Annex 1: Format and Contents of CIE Independent Peer Review Report

1. The CIE independent report shall be prefaced with an Executive Summary providing a concise summary of the findings and recommendations, and specify whether the science reviewed is the best scientific information available.

2. The main body of the reviewer report shall consist of a Background, Description of the Individual Reviewer's Role in the Review Activities, Summary of Findings for each ToR in which the weaknesses and strengths are described, and Conclusions and Recommendations in accordance with the ToRs.

   a. Reviewers should describe in their own words the review activities completed during the panel review meeting, including providing a brief summary of findings, of the science, conclusions, and recommendations.

   b. Reviewers should discuss their independent views on each ToR even if these were consistent with those of other panelists, and especially where there were divergent views.

   c. Reviewers should elaborate on any points raised in the Summary Report that they feel might require further clarification.

   d. Reviewers shall provide a critique of the NMFS review process, including suggestions for improvements of both process and products.

   e. The CIE independent report shall be a stand-alone document for others to understand the weaknesses and strengths of the science reviewed, regardless of whether or not they read the summary report. The CIE independent report shall be an independent peer review of each ToRs, and shall not simply repeat the contents of the summary report.

3. The reviewer report shall include the following appendices:

   Appendix 1: Bibliography of materials provided for review
   Appendix 2: A copy of the CIE Statement of Work
   Appendix 3: Panel Membership or other pertinent information from the panel review meeting.

Annex 2: Terms of Reference for the Peer Review

Joint US-Canada Technical Review Panel for the Pacific Whiting Stock Assessment

1. Become familiar with the draft Pacific hake/Whiting stock assessment(s) and background materials.
2. Comment on the quality of data used in the assessment(s) including data collection and processing.
3. Evaluate and comment on analytic methodologies.
4. Evaluate model assumptions, estimates, and major sources of uncertainty and provide constructive suggestions for improvements if technical deficiencies or additional major sources of uncertainty are identified.
5. Determine whether the science reviewed is considered to be the best scientific information available.
6. Provide specific suggestions for future improvement in any relevant aspects of data collection and treatment, modeling approaches and technical issues.
7. Provide a brief description on panel review proceedings highlighting pertinent discussions, issues, effectiveness, and recommendations

Note – CIE reviewers typically address scientific subjects, hence ToRs usually do not involve CIE reviewers with regulatory and management issues unless this expertise is specifically requested in the SoW.

The NMFS Project Contact will provide the Terms of Reference by 6 January 2010.

Annex 3: Tentative Agenda

Joint US-Canada Technical Review Panel for the Pacific Hake / Whiting Stock Assessment

February 8-10, 2010
Hotel Decca
4507 Brooklyn Avenue NE
Seattle, WA 98105

Monday, February 8, 2010

9:00 a.m.	Welcome and Introductions (Stacey Miller or Jim Hastie, NMFS).
	Review the Status of the Pacific hake / Whiting Treaty
9:15 a.m.	Review the Meeting Agenda (Panel Chair, SSC rep.).
	Review Terms of Reference for Assessments and Review Meeting
	Assignment of reporting duties
9:45 a.m.	Data Presentations
	- Overview of the 2009 Hake/Whiting Fisheries
	o Canadian Waters (Chris Grandin, DFO)
	o U.S. Waters (Ian Stewart, NMFS)
10:15 a.m.	Coffee Break
10: 45 a.m.	Data Presentations Continued
	- Acoustic Survey: Design and Analysis (NMFS)
12:00 p.m.	Lunch (on your own)
1:00 p.m.	Data Presentations Continued
	- Acoustic Survey: 2009 Results (NMFS)
2:00 p.m.	Data Presentations Continued
	- Acoustic Survey: On-going Analyses (NMFS)
3:00 p.m.	Coffee Break
3:30 p.m.	Overview of the Data Sources for the 2010 Assessment (Ian Stewart and Owen Hamel, NMFS)
5:30 p.m.	Adjourn for the day.

Joint US-Canada Technical Review Panel for the Pacific Hake / Whiting Stock Assessment

February 8-10, 2010
Hotel Decca
4507 Brooklyn Avenue NE
Seattle, WA 98105

Tuesday, February 9, 2010

9:00 a.m.	STAT Model Presentations - Stock Synthesis Model Description and Results (Owen Hamel and Ian Stewart, NMFS) - TINSS Model Description and Results (Steve Martell, UBC)
12:00 p.m.	Lunch On Your Own
1:00 p.m.	Q&A session with the STATs - Panel develops list of model runs / analyses for the STAT(s).
5:30 p.m.	Adjourn for day.

Wednesday, February 10, 2010

9:00 a.m.	STAT presentation(s) of requested model runs/analyses.
10:00 a.m.	Panel Discussion - Finalize base case model results, discuss structure of decision table and reporting of uncertainty
12:00 p.m.	Lunch On Your Own.
1:00 p.m.	Panel Drafts STAR Report. - Agree to process for completing final STAR report by Council Briefing Book deadline (Feb. 17th for mailed BB).
5:30 p.m.	Review Panel Adjourn.

Appendix 3: Panel membership and STAR meeting participants

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