Report on

SEDAR 17 Stock Assessment Review

South Atlantic Vermilion Snapper and Spanish Mackerel

October 20 - 24, 2008

Savannah, Georgia

P.L. Cordue

Fisheries consultant

New Zealand

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

14 November 2008

Executive summary

Draft assessments of vermillion snapper and Spanish mackerel were reviewed at the SEDAR 17 Review Workshop, in Savannah, GA, from October 20-24, 2008. This is the report of one member of a four person review panel. It should be read in conjunction with the Consensus Summary Report of each assessment and the CIE reports of two other reviewers. However, this report contains information that was not available to the Review Panel at the time that the Consensus Summary Reports were finalized.

The two stock assessments applied almost identical methods to very similar types of data. Data primarily consisted of fishery dependent abundance indices, length frequencies, and age frequencies. The main assessment method used "statistical catch-at-age" models to determine optimal fits to the data.

A large amount of work was performed by the data and assessment workshop teams. At the Data Workshop, an excellent job was made of bringing together available data and considering its appropriateness for use in stock assessment. At the Assessment Workshop a range of methods were used to interpolate/extrapolate incomplete landings histories and to produce stock assessment results. The industry and the quantity of work are commendable. Unfortunately, these efforts were negated by some lapses in mathematical rigor.

A fatal conceptual error exists in the age-frequency construction method, which means that the age frequencies used in the assessments are either biased or not properly stratified and scaled. This compromises both assessments. Also, the use of ad hoc likelihoods in the statistical catch-at-age models undermines the statistical nature of the models and inhibits the determination of appropriate parameterizations and data weightings.

I find that both assessments are technically unsafe and that point estimates and projections from the assessments should not be used for management purposes. There are two results which are probably robust (even to the technical deficiencies): overfishing is not occurring for Spanish mackerel; and vermillion snapper is not overfished.

My conclusion, for the vermillion snapper assessment, is contrary to that contained in the Consensus Summary Report. In that report, the base model was accepted by the Review Panel. As part of the Review Panel I agreed to those findings. For vermillion snapper, I concluded that most defensible models would probably produce results that would not be too different from the Assessment Workshop base model. I viewed this as a "judgment call" – the base model was technically poor, but it was probably good enough. For Spanish mackerel, because of additional problems, my "judgment call" was that the base model could not be accepted – except for the robust, "not overfishing" conclusion.

However, I later discovered the *error* in the method used to construct the age frequencies. I now conclude that the methodological issues, combined with the error in the construction method for age frequencies, require that both assessments be rejected.

Review Activities

Pre-meeting

Meeting documents and materials were made available in electronic form in advance of the meeting (*see* Appendix 2). I familiarized myself with the background material and read the main assessment documents in detail prior to the meeting.

By email, the Chair requested volunteers to "lead" the assessment reviews (the "lead" being responsible for writing/compiling the first draft of the Consensus Summary Report for the assessment). I volunteered and was assigned Vermillion Snapper by the Chair.

Meeting

The meeting was convened at 1 pm on Monday, October 20, and concluded at lunchtime on Friday, October 24. The review panel (RP) consisted of Gary Shepherd (Chair), Noel Cadigan (Spanish Mackerel lead), Beatriz Roel, and myself. The meeting followed the draft agenda quite closely (*see* Appendix 1), except that Spanish Mackerel was presented first.

The PowerPoint presentations of the assessments were well constructed with an appropriate level of detail. The assessment team representatives (AT) were very professional and helpful during the meeting. They provided informative answers to questions and were responsive to the RP's requests for additional analyses, and in seeking further clarification from people not present at the meeting.

Discussions during the meeting covered a wide range of data and stock assessment issues. At times, discussions were somewhat robust - with differences of opinion within the RP, rather than between the RP and AT. The main issues with regards to the assessment appeared to be covered and conclusions with regard to the acceptability of each assessment were agreed by the RP. However, there was insufficient time to discuss the recommendations made by the Data Workshop (DW) and Assessment Workshop (AW). Also, I discovered a serious error after the meeting while I was preparing this report (see *Post-meeting* section below)

As the vermillion snapper (VS) lead I was responsible for the first draft of the VS Consensus Summary Report (CSR). After discussion amongst the RP it was agreed that the leads would draft material for TOR 1-6 during the meeting (after hours) and the Chair and remaining CIE reviewer would cover TOR 8 (considering each TOR of the Data and Assessment Workshops with regard to whether they were satisfactorily completed or not, and suggesting improvements to methods and outputs).

At the conclusion of the meeting, the draft material had been completed, at least to bullet point form, for TORs 1-6 and 8. Also, there had been some discussion and revision of the wording on most of those TOR.

Post-meeting

The lead authors completed full drafts of the CSRs for their respective assessments and submitted them to the Chair, who distributed them to the full RP for comments. After a relatively short period, and one or two iterations for revision, the wording of the CSRs was accepted by the RP and left with the Chair for final editing.

I then turned my attention to preparation of my CIE report. I followed up a number of "loose ends" with regard to data preparation and assessment methods. In particular, I was able to clarify the detail on the method used to scale age frequencies (using length frequencies). On analyzing the method I found that it was conceptually and fatally flawed, and this changed my conclusions with regard to the VS assessment. I alerted the Chair and the SEDAR Coordinator to the fact that my CIE report would contradict the VS Consensus Summary Report. My concerns were brought to the attention of the AT who prepared a short document in support of the method they had used. Their reply did nothing to alleviate my concerns (see **Length and age frequencies** in Review Findings below). However, the Consensus Summary Reports were not modified.

Review findings

The two stock assessments applied almost identical methods to very similar types of data. Therefore, I consider both stock assessments together under each of the specified terms of reference. My comments apply equally to both assessments unless otherwise stated.

SEDAR 17 Review Workshop Terms of Reference (apply to each stock):

1. ***Evaluate the adequacy, appropriateness, and application of data used in the assessment.***

I take this TOR to apply to the output of the DW (including the methods used to derive the outputs), rather than just "data".

The DW provided data on landings, bycatch, and discard history. They also supplied abundance indices, length and age frequencies, stock definitions and estimates of biological parameters and relationships. These outputs are considered under four headings.

Stock definitions and biological parameters
The first question for a stock assessment is the stock definition. The DWs reviewed previous research and considered available data (genetic, tagging, morphometric and life history parameters, distributional patterns, spawning sites) and drew reasonable conclusions with regard to stock boundaries.

The DWs also recommended estimates of life history parameters (and discard mortality rates) based on previous studies and new analyses. The recommendations appear well founded except in two regards.

The specification of growth parameters by the DW, estimated from age-length data, is inappropriate for an assessment which is fitting to length frequency data. It is a modelling decision as to whether growth parameters are estimated inside the model (simultaneously with other parameters) or externally to the model. Arguments can be made in favour of either option, but when length frequency data are being fitted and age-length data are available, I believe it is generally preferable to fit growth internally (and include the age-length data in the fitting process). Certainly, the modelers should be allowed to explore such options.

The steepness parameter of the Beverton-Holt stock recruitment relationship does not appear to have been considered by the DWs. This parameter cannot be estimated accurately within a stock assessment model unless the stock has been driven to low levels (less than 20% of virgin) and there are good data available on relative recruitment and stock size at a range of stock sizes. The life history group is perhaps best placed to advise the AW on a plausible range for steepness (being the expected proportion of virgin recruitment when SSB is at 20% of the virgin level).

Landings, bycatch, and discards

For both stocks, landings histories were extended back to the 1950s. The DWs clearly put a substantial effort into reconstruction of landings histories, both commercial and recreational. Efforts were also made with regard to estimating discards for the various fisheries, and importantly for Spanish Mackerel as a bycatch in the shrimp fishery.

The landings and discards for the commercial fisheries were well constructed by the DWs. My only comment is that the quantification of uncertainty using only CVs for different periods may not be adequate. There may be periods of time when the estimates are likely to be biased (e.g., a negative bias due to under-reporting). As an aid to the AW (in choosing alternative landings stream), the uncertainty associated with landings histories should be quantified in terms of potential bias and a CV (e.g., during period X-Y, the DW suggest a potential bias from –30% to 10% with a CV of 10% - the CV being taken to apply to a specified landings history after bias adjustment). The point here, is that a CV of 40% (as specified during some periods by the DWs) is not made up entirely of observation error around an unbiased estimate of landings – the large CV is being driven by concerns about potential *bias* – and that is how it should be expressed (e.g., as was done by the AW for the saltwater surveys).

The landings and discards for the recreational fisheries were also well constructed by the DWs, to the extent that they chose to do so. The early landings estimates were constructed by the AW based on three imprecise and potentially highly-biased saltwater-angler surveys (SEDAR17-RD13, RD14, and RD15).

For SM, the bycatch in the shrimp fishery is substantial, but difficult to estimate due to sporadic observer coverage (SEDAR17-DW12). Again, the DW left the job of constructing a full history to the AW (*see* SEDAR17-RW02).

Length and age frequencies

Length frequencies were constructed with some care by the DWs. Raw length data were generally stratified by fishery, area/state, and year; and scaled by landing-numbers within strata. However, there seems to have been little systematic investigation of the variability of length by alternative strata (e.g., month, depth). In some cases alternatives were investigated but dismissed – with minimal documentation. There is a need for a more detailed study. For year-round fisheries, month/season is likely to be a necessary stratification (depending on how fast the fish are growing). Also, raw data should probably be scaled first to trip landings and then to higher level strata.

The age frequencies were constructed using a two-step method involving an initial (minimal) stratification and then a correction to compensate for potential differences between the length frequency of the length sample and the aged sample. For example, from SEDAR17-RW02 (SEDAR17-RW01 has similar wording):

> "Weighting was initially by state landings in numbers, and then by length composition as shown in Figures 3.14 -3.18, respectively. This latter weighting is intended to correct for a potential sampling bias of age samples relative to length samples (see Section 3 in SEDAR10 for South Atlantic gag grouper)."

The reference to SEDAR 10 eventually leads to a document which describes and illustrates a potential sampling bias for otoliths. The method used to "weight by length composition" is described in SEDAR17-RW02 (and SEDAR17-RW01) under recreational fisheries:

> "The recreational ages were weighted by the recreational length composition to overcome potential bias in selecting fish to age and to transfer the weighting given to the length composition based on landings to the age composition. The weighting value for each age record was the proportion from the length composition corresponding to the year and length (1 cm bins) of the aged fish. The weighting values were then summed by age and year to determine the age composition of the fishery. Each value was normalized to sum to 1 across years by dividing each value by the sum for that year."

This weighting method is contrary to the usual means of obtaining an age frequency from a representative length frequency and an age-length key. In the usual method, for each length, the proportion of age-at-length is applied over all lengths (or length bins) to obtain the age frequency. In the above, "length-proportion method", there is no need for each length bin to have age records as each age record is simply weighted by the proportion-at-length for the length of the aged fish. This is a simple method, but it is conceptually flawed.

The method has no prospect of doing what it is claimed to do: "to overcome potential bias in selecting fish to age and to transfer the weighting given to the length composition based on landings to the age composition". The use of an age-length key does have that effect, which means that the careful stratification and scaling can be done for the length frequency and non-random age samples can be used to obtain an age frequency. However, the length-proportion method is scaling the age frequency on the basis of observed length proportions – conceptually, it is difficult to see how this could improve the age frequency. The AT were queried on the use of this method after the Review Meeting and said that it was tested by simulation and found to compensate for age sampling that selected too many large fish, and that it had little effect on representative samples. However, they offered no mathematical support for the method.

Mathematically, to disprove a theory, one needs only to provide a single counter example. Let us consider some simple examples – they do not have to be realistic – but simply within the parametric context of the problem.

Consider samples, from just a single year, for a population where fish are either 1 or 2 years old and measure 10 cm at age 1 and 20 cm at age 2 (with no variability in length at age). Suppose that there are equal numbers of fish aged 1 and 2 years and that the sampled length frequency reflects this exactly (so, 50% length 10 cm and 50% length 20 cm). Now take a non-random sample of aged-fish: 80 aged 1 year and 20 aged 2 years. Under the age-length key method: the proportion of age 1 fish at 10 cm is 1; the proportion of age 2 fish at 20 cm is 1; and the estimated age frequency has 50% aged 1 and 50% aged 2. Under the length-proportion method, each age record gets a weight of 0.5 (being the length-proportion whether the aged-fish measures 10 cm or 20 cm); the summed weights for age 1 is 40; the summed weights for age 2 is 10; the estimated age frequency has 80% at age 1 and 20% at age 2. In fact, for any length frequency with equal proportions in each length bin, the length-proportion method does not alter the age frequency (whether it is a random or non-random sample).

We see that the length-proportion method does not properly correct for non-random age samples and does not transfer the scaling of the length frequency to the age frequency. Also, it is important to consider what effect the method may have on representative age samples (because it was used on all of the age data). For simplicity, consider a population which has no variation in length at age:

p_j = proportion at age j years
l_j = length of fish at age j years

If the length-proportion method is applied to the population's length frequency and age frequency the estimated proportion at age j years is proportional to p_j^2 (because the length-proportion for a fish aged j years is p_j and the proportion of fish aged j years is also p_j). Therefore, application of this method will distort representative age frequencies – it will tend to reduce the proportion of fish at older ages (as there tend to be fewer fish at older ages – e.g., continue the above example, but suppose there are

60% fish aged 1 year, 30% aged 2 years, and 10% aged 3 years; the length-proportion method scales the true age frequency to give an estimate of 78% aged 1, 20% aged 2, and 2% aged 3). The method will have the effect of increasing the slope of the right-hand limbs of age frequencies, which could lead to over-estimation of total mortality in stock assessment models (depending on how many biased age frequencies are also present).

In summary, for the length-proportion method: the scaling of length frequencies is not transferred to the age data; biased age sampling is not properly corrected; representative age samples are distorted.

The use of this method is a serious issue, but it was not discussed at the Review Meeting. Unfortunately, it was not until I was preparing this report that I fully investigated the issue.

The use of age-length keys (applied to aggregated length bins) might be a possible method for using the biased otolith data, but this may be inappropriate depending on how fast the fish are growing over the period in which the length and age data are collected. Another alternative is to include such data as conditional age-at-length (with an appropriate likelihood). Also, the "random age samples" (to the extent that they can be identified) need to be properly stratified and scaled (just scaling to landings by state is probably inadequate). It may be that many of the age samples are not adequate, in terms of coverage, when proper stratification is considered.

The only measure of sample size provided by the DWs for age and length data were the number of fish aged or measured. It is well known that the number of fish measured in a length frequency can vastly over-estimate the effective multinomial sample size (e.g., Pennington et al. 2002). The number of trips/tows sampled is often a better approximation to the effective sample size than the number of fish measured. Effective sample sizes for age frequencies will depend on the method used (e.g., sampling directly for age or through an age-length key). The best method to determine effective sample sizes (for age or length frequencies) is through a full bootstrap of the raw data (e.g., see Bull and Dunn 2002).

A bootstrap tool is also needed for effective planning of age and length sampling. It is difficult to determine how many fish should be measured and aged for a particular assessment. However, if tools are available, for example, in estimating the achieved precision in an age-frequency, for a given number of measured and aged fish (assuming an age-length key is used), then criteria can be set on the basis of desired precision. That said, some basic rules of thumb can go a long way in determining what level of sampling is needed.

It appears that age samples are barely adequate (in terms of coverage across important stratifying variables) for most fisheries in most years for both SM and VS (e.g., some of highest sample sizes are for VS commercial handline fishery, but there are almost no samples from Georgia, and in the last 3 years when sample sizes from North and South

Carolina are good, the samples from Florida drop away – see Table 3.8 in SEDAR17-RW01; for SM the sample sizes in the commercial gillnet fishery may be adequate for 1996-2006; but the sample sizes for individual states are not shown – see Table 3.9 in SEDAR17-RW02)

Abundance indices

The DWs reviewed many potential abundance indices and provided good documentation on the pros and cons of each potential time series. They also properly considered the suitability of the indices given changes in fishing regulations.
The GLM models used were probably adequate but better GLM models could have been developed and better diagnostics could have been presented. For each GLM there should be a description of all potential explanatory variables and a method adopted for determining the optimal set of variables to use (e.g., forward stepwise). Consideration should also be given to explaining catch rather than catch per unit effort – and using potential effort variables as explanatory variables. Vessel characteristics or indeed vessels as a categorical variable were absent from most GLMs – yet, often, vessel is a crucial explanatory variable (perhaps acting as a proxy for skipper or particular areas fished). When a combined (delta-) model is used, diagnostics and indices should be given for both the binomial model and the positive model. Diagnostics should include the effects of explanatory variables – and consideration of whether they are plausible.

For SM, the AW base model did not use the DW recommended abundance indices, but instead used a "combined" time series. For VS, the AW base model used the recommended indices, but a "combined" time series was used in alternative models. Comments on the use of combined indices and the method used to derive them are given below under TOR 2.

2. ***Evaluate the adequacy, appropriateness, and application of methods used to assess the stock.***

I took this TOR to apply to the output from the AWs.

Four assessment methods were used for each stock: "statistical catch at age" (SCA); non-equilibrium production model (ASPIC); stochastic stock reduction analysis (SRA); and catch curve analysis. Also, the AWs were responsible for constructing the early recreational landings histories (using the salt-water angler surveys in 1960, 1965, and 1970); the bycatch history (SM) for the shrimp trawl fishery; and the "combined" abundance time series.

Combined abundance time series

The AWs constructed combined abundance indices using a Bayesian method applied to the ratio of consecutive indices within each time series (SEDAR17-AW06). It perhaps seems tidier to fit to just a single combined time series rather than having to bother with multiple time series. However, the philosophy is questionable, especially if the individual time series are suggestive of different abundance trends. In that case, it is essential to investigate the consequences of different hypotheses with regard to the reliability of alternative abundance indices (*see* Schnute and Hilborn 1993).

Even if the philosophy is accepted, there are technical and practical issues with the adopted method. The method assumed a multiplicative and normal error structure for the ratio of consecutive indices. The usual assumption, for individual indices, is multiplicative and lognormal – which leads to lognormal errors for the ratio of indices. Also, independent errors were assumed, which is clearly violated when $n – 1$ ratios are formed from consecutive pairs of n indices. The robustness of the method to these violated model assumptions has not yet been studied.

Practical problems also exist. It was not clear what the combined time series was indexing – clearly vulnerable biomass of some sort – but what selectivity was used? (I was told an average of some sort.) Also, what CVs should be used for the indices? The method provided estimates but these started at 0 (for the index of 1 in the first year) and became increasing large for successive indices (*see* Table 1 in SEDAR17-AW06). In the AW SCA models, each year was given the same CV – which is not appropriate.

Early recreational landings history

Slightly different methods were used for SM and VS but both relied on the imprecise and potentially highly biased saltwater angler surveys of 1960, 1965, and 1970 (SEDAR17-RD13, RD14, RD15). Both methods essentially interpolated between an assumed landing in 1950 and the survey estimates adjusted for bias (75% of the estimates was used in the base models). In the case of SM, the recreational catch in 1950 was taken as the average of the three surveys; and for VS the 1950 catch was taken to be zero. The high SM estimates pre 1960 were corroborated by other early surveys/reports. The methods adopted by the AW are reasonable. They identified the high potential for bias and performed sensitivity runs at proportions of 50%, 100%, and 125% (base = 75% of estimates).

Spanish mackerel bycatch in the shrimp trawl fishery

The AW were faced with the difficult task of extrapolating a bycatch history for the shrimp trawl fishery from 1950 to 2007 with estimates from only 8 years in the latter part of the period (SEDAR17-AW07). The method used a "hockey stick" relationship between the SM bycatch and the logarithm of annual shrimp landings. The relationship is dubious and is driven by two observations which had very high bycatch and also had the highest annual shrimp landings (*see* Figure 3 in SEDAR17-AW07). The bycatch history derived using this method is extremely spiky due to the occurrence of annual shrimp landings near the junction of the hockey stick (estimated landings are low to the left of the junction and high to the right). A smoothed time series was used in a sensitivity run.

The approach to this problem was not ideal. The original estimates of bycatch come from a GLM using a delta-lognormal model (SEDAR17-DW12) which suffers from the same problems as mentioned for the abundance time series (not the best GLM models or diagnostics). The eight model estimates are then bundled into a dubious relationship with annual shrimp landings – which results in an entirely implausibly spiky time series. I think that a much simpler approach is called for: some "ballpark estimates" with sensitivity runs (each run having a relatively smooth bycatch history).

ASPIC, SRA, and catch curves

It seems a good idea to investigate a range of estimation methods at different levels of complexity that use different subsets of the available data. However, the ASPIC models could never integrate all available data sources and would therefore always be questionable as a stand-alone method (and therefore be of limited value as a sensitivity analysis – if the results are different to the base model, they are dismissed). The same argument applies to the SRA and catch curve results – but these methods were not even implemented properly.

The catch curve results included estimates from "tracking cohorts" – and this was claimed to be preferable to using the right-hand limb of annual samples (SEDAR17-AW04, AW05). Although cohorts can be tracked over consecutive years of catch sampling data and a regression performed on the proportions (of the cohort within each year), the estimates are of dubious value. The problem is that the annual *proportions* are not directly proportional to the annual cohort *numbers* because they are confounded due to the presence of the other cohorts. Strong assumptions would be required to make the method valid. On the contrary, the usual catch curve analysis of the right-hand limb of an age frequency is fairly robust to violations in the required assumptions of constant selectivity and constant recruitment (Dunn et al. 1999).

The SRA was performed on ratios of consecutive indices (SEDAR17-AW08). The error structure of the ratios was assumed to be normal with all errors mutually independent. Errors for indices are usually considered lognormal (hence ratios would also have lognormal errors) and consecutive ratios clearly have dependent errors. Violation of these two assumptions makes the method dubious unless robustness to the violated assumptions can be demonstrated.

A better method of testing the sensitivity of assessment results to model structure is to perform *defensible* assessments using alternative model structures. For a good example of this see Francis (2004) where three base models are presented (all in the same integrated framework but with alternative population dynamics and data assumptions).

SCA models

The use of a quasi-likelihood with user specified weights for groups of likelihood components undermines the statistical nature of the model. Unless the weights are all set to 1, it is not appropriate to calculate standardized residuals or to perform model comparisons on the basis of AIC or other statistical criteria.

For example, it is not clear that any cohorts can be seen tracking through in the age frequency data. It may be that there is no justification for estimating annual recruitment deviations (over the period covered by data). It is not uncommon to compare models with and without recruitment deviations estimated – and to only accept the model with the extra parameters if the parameters are justified by a sufficient decrease in negative log-likelihood (AIC = 2 log-likelihood units per parameter).

Standardized residuals are an essential diagnostic in relationship to the question: is the achieved fit compatible with the assumed CVs and multinomial sample sizes? Also, the "natural weight" for each likelihood component can be determined by calculating standardized residuals (which should be approximately N[0,1]) and iteratively adjusting CVs and multinomial sample sizes until the standard deviation of the standardized residuals is approximately equal to 1 for each data set (e.g., see Bull et al. 2002; the total CVs are taken as the "sum" of observation error and "process" error).

Alternative approaches exist for adjusting CVs and sample sizes so that the input variance assumptions match the variance of observed residuals (e.g., see Methot 2007). This matching of "input and output variances" is a pre-requisite to having an internally consistent model. At times, one may wish to move away from "natural" weights. A good example of this is shown in a New Zealand stock assessment where the natural weights lead to a misfit of a reliable abundance time series and the abundance indices were "up-weighted" (Francis 2004).

The use of ad-hoc estimators needs a strong justification. There need to be some advantages gained from their use – they need to have been developed to deal with some outstanding issues. The estimators used in the SM and VS assessments do not appear to fall into this category.

The method is also less than ideal when it comes to the likelihoods used (though this matters little when arbitrary weights are applied). For example, the likelihood for lognormal abundance indices includes a small constant to avoid taking the log of zero. This should not matter, but one wonders how a zero index could be included in an assumed lognormal time series – or how predicted values could equal zero. Also, the likelihood is given for lognormal indices but there is no mention of the assumed relationship between an abundance index and its associated biomass. "Lognormal CPUE" is ambiguous. Below are two relationships between biomass and CPUE with lognormal errors:

$$X_i = qB_i\varepsilon_i \quad \varepsilon_i \sim \mathrm{LN}(\frac{-\sigma^2}{2}, \sigma^2), \quad \text{so that } \mathrm{E}(\varepsilon_i) = 1$$

$$\log(X_i) = \log(qB_i) + \varepsilon_i \quad \varepsilon_i \sim \mathrm{N}(0, \sigma^2)$$

where X_i is a CPUE index, B_i is the vulnerable biomass, q the proportionality constant, and ε_i the error. Each relationship gives rise to a different likelihood. The documentation should specify the assumed relationship so that reviewers can check the derived likelihood – rather than the reviewer having to deduce the relationship from the stated likelihood.

The structure of the population dynamics model is questionable in SM because there are consecutive seasonal commercial fisheries (Florida in winter; northern states in summer and early fall). However, in the model, all fisheries are assumed to be year-round and to act simultaneously – contrary to reality. This could affect the fit to CPUE indices in particular.

The models have relatively few parameters that need estimation. However, the implementation of the models requires that all *F*s are estimated (as free parameters), rather than being directly calculated. Use of the Pope approximation, or an iterative solution to the Baranov equations, would reduce the free parameters in the models enormously with a large reduction in run time. The proportionality constants could also be calculated directly (on each function iteration) rather than estimated as free parameters.

The issue of fitting landings histories exactly was a focus of discussion for some members of the RP. The AWs followed the recommendation of a previous SEDAR to account for landings uncertainty by specifying alternative landings histories in a range of sensitivity runs – and fitting landings exactly in each run. I support this approach as there appear to be no good alternatives. In special circumstances it may be possible to estimate a portion of a landings history – but for the SM and VS assessments there is clearly no information in the available data for the model to reliably estimate early landings.

The ADMB code for each model is well structured and well commented. The automatic generation of R-objects from a model run is also well done. There has been a substantial investment in the existing ADMB models (previous code has no doubt been extensively reused – and tailored to each new stock assessment). However, use of assessment-specific models rather than general packages does need some justification. There is learning curve with any package, but once that cost is paid, the use of a package can not only speed up the production of stock assessments but also increase the reliability of the assessments. A purpose-written model is really only needed if there are no packages which can accommodate critical population or estimation requirements. The SM and VS models could have been accommodated in at least two currently available packages. Serious consideration should be given to moving to packages such as SS2 or CASAL (Bull et al. 2002, Methot 2007).

The base models for SM and VS both suffer from the methodological issues discussed above (and the very poorly constructed age frequencies). They also both have sensitivities to the poorly known parameters of natural mortality and steepness. The SM assessment has additional problems: the substantial but very uncertain bycatch in the shrimp fishery and early recreational catch; and the use of the combined abundance time series. During the review the RP requested several sensitivity runs for both assessments (covering natural mortality, steepness, landings history, and log-likelihood multipliers for the recruitment indices). It was found that some of the conclusions of the base models with regard to stock status were robust. For VS the conclusion that the stock was not overfished was robust to everything tested except low natural mortality.

For SM, the conclusion that overfishing was not occurring was robust to all sensitivities tested. Also, for VS, estimated F in the terminal year was fairly robust to the tested sensitivities.

The RP concluded that the VS base model could be accepted, but warned that the conclusion of the base model that overfishing was occurring was sensitive to model assumptions. The SM assessment was mainly rejected – but the robustness of the “not overfishing” conclusion was noted.

As part of the RP I agreed to these findings. For VS, I concluded that most defensible base models would fall within the range of sensitivities that had been explored – and would therefore *probably* be not too different from the AW base model. I viewed this as a “judgment call” – the AW base model was technically poor, but it was probably good enough. For SM, because of the additional problems, my “judgment call” was that the AW assessment could not be accepted – except for the robust, “not overfishing” conclusion.

However, after discovering *exactly* how the age frequencies were constructed, I cannot accept the AW base model for VS. The method used, in an attempt to correct for biased otolith sampling, is invalid. Properly constructed age frequencies could be quite different from those that were used and could impact substantially on assessment results. I do not know that a defensible base model, using properly stratified and scaled age frequencies, would result in substantially different stock assessment results. However, given all of the methodological issues, and the *error* in the construction of age frequencies, both assessments should be rejected on technical grounds.

3. *Recommend appropriate estimates of stock abundance, biomass, and exploitation.*

The AW assessments are not sufficiently technically sound for any of the point estimates to be recommended.

4. *Evaluate the methods used to estimate population benchmarks and management parameters (e.g., MSY, Fmsy, Bmsy, MSST, MFMT, or their proxies); provide estimated values for management benchmarks, a range of ABC, and declarations of stock status.*

The method of Shepherd (1982) was used to estimate MSY-based reference points. However, this method requires reliable knowledge of the stock-recruitment relationship – which is not the case for either assessment. It would be preferable to use proxies (e.g., based on $F_{40\%}$). I do not support the benchmark estimates from either assessment.

The stock status is robust in two regards:

- VS: the stock is not overfished;
- SM: overfishing is not occurring.

5. *Evaluate the adequacy, appropriateness, and application of the methods used to project future population status; recommend appropriate estimates of future stock condition (e.g., exploitation, abundance, biomass).*

The projection method uses estimated numbers at age as a starting point and projects forward using stochastic recruitment. However, the average projection trajectory is defined to be deterministic (to ensure that the average trajectory is consistent with the deterministic benchmarks). This is an adequate approach for short term projections, but it would be better if the estimation uncertainty in current numbers at age was included.

No specific projections are recommended.

6. *Evaluate the adequacy, appropriateness, and application of methods used to characterize uncertainty in estimated parameters. Provide measures of uncertainty for estimated parameters. Ensure that the implications of uncertainty in technical conclusions are clearly stated.*

The AW made a genuine effort to quantify uncertainty, in terms of: parameter estimates, through a partial bootstrap; robustness, with sensitivity runs; model structure by using alternative methods. However, the methods fall short of best practice.

The parametric bootstrap on recruitment deviations will not capture parameter uncertainty adequately. Consider the effect of increased CVs on abundance indices, or reduced sample sizes for length or age frequencies – how would these changes flow through into larger confidence intervals for estimated parameters? It seems unlikely that they would – the variability in recruitment deviations is largely governed by the best fit to the data rather than the variability of the data. Therefore, there is only a tenuous link between the variance of the input data and the confidence intervals of the estimated parameters. To make matters worse, the CVs are all capped at a maximum of 30% (SEDAR17-RW01, RW02).

There are two main choices for capturing uncertainty appropriately. The best currently available method is the use of full Bayesian methods (e.g., Francis 2004). The main alternative is to use properly constructed likelihoods in conjunction with approximate confidence intervals from likelihood profiles or bootstraps. Sensitivity runs are essential whatever the method.

7. *Ensure that stock assessment results are clearly and accurately presented in the Stock Assessment Report and Advisory Report and that reported results are consistent with Review Panel recommendations.*

At the time of writing this report, only the SM Assessment Report and Advisory Report have been completed. I submitted comments on both reports to the Chair and the SEDAR Coordinator. I was concerned that, although the SM assessment had been mainly rejected by the RP, numerous results were reported in the Advisory Report.

Also, the length of the Addendum (some 100 pages) to the draft stock assessment report appeared excessive given the "partial acceptance" of the assessment.

8. *Evaluate the SEDAR Process. Identify any Terms of Reference which were inadequately addressed by the Data or Assessment Workshops; identify any additional information or assistance which will improve Review Workshops; suggest improvements or identify aspects requiring clarification.*

The DW and AW produced a large quantity of work and adequately addressed most of their terms of reference. Deficiencies have already been noted earlier in this report: invalid construction of age frequencies and inadequate information on length/age frequency sample size (trips/number-of-samples needed as well as number of fish); inappropriate stock assessment models due to quasi likelihoods; inappropriate benchmarks due to poorly defined stock-recruitment relationships.

The TOR for the DW and AW are comprehensive in terms of what is needed to perform a stock assessment. However, they are somewhat lacking in what is required to review a stock assessment. There appears to be no requirement for executive summaries to be produced for any aspect of the data preparation or assessment. The DW and AW reports could have been greatly improved with the inclusion of executive summaries aimed at reviewers who may be unfamiliar with the particular fisheries and data sets.

9. *Review the research recommendations provided by the Data and Assessment workshops and make any additional recommendations warranted. Clearly indicate the research and monitoring needs that may appreciably improve the reliability of future assessments. Recommend an appropriate interval for the next assessment.*

The numerous research recommendations from the DW and AW were not explicitly discussed at the RW. I briefly reviewed the recommendations and I am in broad agreement with the suggestions. However, there is a clear need for the recommendations to be prioritized.

An appropriate interval for the next assessments depends on the requirements of fisheries managers. From a scientific viewpoint, both assessments should be redone as soon as proper age frequencies can be constructed and appropriate SCA models developed (such models are already available in packages such as SS2 and CASAL).

Conclusions and Recommendations

A large quantity of work was performed by the DW and AW. The DW did an excellent job of bringing together available data and considering its appropriateness for use in stock assessment. The AW applied a range of methods to interpolate/extrapolate incomplete landings histories and to produce stock assessment results. The industry and the quantity of work are laudable. Unfortunately, these efforts were negated by some lapses in mathematical rigor.

The fundamental conceptual error in the method used to construct age frequencies means that the age frequencies used in the assessments are either biased or not properly stratified and scaled. This compromises both assessments. The use of ad hoc likelihoods in the SCA models undermines the statistical nature of the models and inhibits the determination of appropriate parameterizations and data weightings.

I find that both assessments are technically unsafe and that point estimates and projections from the assessments should not be used for management purposes. There are two results which are probably robust (even to the technical deficiencies): overfishing is not occurring for Spanish mackerel; and vermillion snapper is not overfished.

I will limit myself to just two recommendations for all future assessments:

- age data be carefully screened to identify random and non-random sampling:
  - random samples be used to construct age frequencies using fully validated methods;
  - non-random samples be used as conditional age-at-length data (within the model) or as age-length keys (if appropriate);
- SCA models be used to integrate available data sources:
  - defensible likelihoods be used without any scaling of likelihood components;
  - a weighting scheme be adopted which ensures, as a first step, that the input variance assumptions are consistent with the variances of the residuals (up-weighting of abundance indices may be necessary as a second step);
  - confidence or credibility intervals be calculated in a manner that ensures they are properly related to input variance assumptions.

References

Also see Appendix 2 for workshop documents cited in the text.

Bull, B.; Dunn, A. 2002: Catch-at-age: User manual v1.06.2002/09/12. NIWA Internal Report 114. 23 p. NIWA.

Bull, B.; Francis, R.I.C.C., Dunn, A.; Gilbert, D.J. 2002: CASAL (C++ algorithmic stock assessment laboratory): CASAL User Manual v2.01.2003/08/01. *NIWA Technical Report 124*. 223 p.

Dunn, A., Francis, R. I. C. C., Doonan, I. J. 1999: The sensitivity of some catch curve estimators of mortality to stochastic noise, error, and selectivity. N.Z. Fisheries Assessment Research Document 99/15. 23 p.

Francis, R.I.C.C. 2004: Assessment of hoki *(Macruronus novaezelandiae)* in 2003. *New Zealand Fisheries Assessment Report 2004/15*. 95 p.

Methot, R. D. 2007: User manual for the integrated assessment program stock synthesis II (SS2): Version 2.00c (March, 2007). 84p.

Pennington, M.; Burmeister, L.; Hjellvik, V. 2002: Assessing the precision of frequency distributions estimated from trawl-survey samples. *Fishery Bulletin 100*:74–80.

Schnute, J.T.; Hilborn, R. 1993: Analysis of Contradictory Data Sources in Fish Stock Assessment *Can. J. Fish. Aquat. Sci. 50*: 1916–1923.

Shepherd, J. G. 1982: A versatile new stock-recruitment relationship for fisheries, and the construction of sustainable yield curves. *Journal du Conseil pour l'Exploration de la Mer 40*: 67–75.

Appendix 1: Statement of Work

Attachment A: Statement of Work for Patrick Cordue

External Independent Peer Review by the Center for Independent Experts

SEDAR 17 Stock Assessment Review

South Atlantic Vermilion Snapper and Spanish Mackerel

October 20 - 24, 2008

Savannah, Georgia

SEDAR Overview:

South East Data, Assessment, and Review (SEDAR) is a process for fisheries stock assessment development and review conducted by the South Atlantic, Gulf of Mexico, and Caribbean Fishery Management Councils; NOAA Fisheries Southeast Fisheries Science Center (SEFSC) and Southeast Regional Office (SERO); and the Atlantic and Gulf States Marine Fisheries Commissions. SEDAR is organized around three workshops: data, assessment, and review. Input data are compiled during the data workshop, population models are developed during the assessment workshop, and an independent peer review of the data, assessment models, and results is provided by the review workshop. SEDAR documents include working papers prepared for each workshop, supporting reference documents, and a SEDAR stock assessment report. The SEDAR stock assessment report consists of a data report produced by the data workshop, a stock assessment report produced by the assessment workshop, and a peer review consensus report prepared by the review workshop.

SEDAR is a public process conducted by the Fishery Management Councils in the Southeast US. All workshops, including the review, are open to the public and noticed in the Federal Register. All documents prepared for SEDAR are freely distributed to the public upon request and posted to the publicly accessible SEDAR website. Verbal public comment during SEDAR workshops is taken on an 'as needed' basis; the workshop chair is allowed discretion to recognize the public and solicit comment as appropriate during panel deliberations. Written comments are accepted in accordance with existing Council operating procedures. The names of all participants, including those on the review panel, are revealed.

The review workshop provides an independent peer review of SEDAR stock assessments. The term review is applied broadly, as the review panel may request additional analyses, error corrections and sensitivity runs of the assessment models provided by the assessment workshop panel. The review panel is ultimately responsible for ensuring that the best possible assessment is provided through the SEDAR process. The review panel task is specified in terms of reference (ToR).

The SEDAR 17 review panel will be composed of three Center for Independent Experts (CIE)-appointed reviewers, one reviewer appointed by the South Atlantic Council, and a chair appointed by the SEFSC director. Council staff, Council members, and Council AP and SSC members will attend as observers. Members of the public may attend SEDAR review workshops.

Overview of CIE Peer Review Process:

The Office of Science and Technology implements measures to strengthen the National Marine Fisheries Service's (NMFS) Science Quality Assurance Program (SQAP) to ensure the best available science for fisheries management. For this reason, the NMFS Office of Science and Technology coordinates and manages a contract for obtaining external expertise through the Center for Independent Experts (CIE) to conduct independent peer reviews of stock assessments and various scientific research projects. The primary objective of the CIE peer review is to provide an impartial review, evaluation, and recommendations in accordance to the Statement of Work (SoW), including the Terms of Reference (ToR) herein, to ensure the best available science is utilized for the National Marine Fisheries Service management decisions.

The NMFS Office of Science and Technology serves as the liaison with the NMFS Project Contact to establish the SoW which includes the expertise requirements, ToR, statement of tasks for the CIE reviewers, and description of deliverable milestones with dates. The CIE, comprised of a Coordination Team and Steering Committee, reviews the SoW to ensure it meets the CIE standards and selects the most qualified CIE reviewers according to the expertise requirements in the SoW. The CIE selection process also requires that CIE reviewers can conduct an impartial and unbiased peer review without the influence from government managers, the fishing industry, or any other interest group resulting in conflict of interest concerns. Each CIE reviewer is required by the CIE selection process to complete a Lack of Conflict of Interest Statement ensuring no advocacy or funding concerns exist that may adversely affect the perception of impartiality of the CIE peer review. The CIE reviewers conduct the peer review, often participating as a member in a panel review or as a desk review, in accordance with the ToR producing a CIE independent peer review report as a deliverable. At times, the ToR may require a CIE reviewer to produce a CIE summary report. The Office of Science and Technology serves as the COTR for the CIE contract with the responsibilities to review and approve the deliverables for compliance with the SoW and ToR. When the deliverables are approved by the COTR, the Office of Science and Technology has the responsibility for the distribution of the CIE reports to the Project Contact.

CIE Reviewer Requirements:

The CIE shall provide three CIE reviewers to conduct independent peer reviews in accordance with the Statement of Tasks, Schedule of Milestones and Deliverables, and SEDAR ToR herein.- Each CIE reviewer's duties shall not exceed a maximum of 14 days for pre-review preparations, conducting the peer review at the SEDAR 17 panel review meeting, completion of the CIE independent peer review reports in accordance with the ToR, and assurance that final review comments and edits are provided to the chair. The CIE reviewers shall participate as technical reviewers on the SEDAR 17 review panel that will consider assessments of South Atlantic vermilion snapper and South Atlantic Spanish mackerel, and these stocks are assessed within the jurisdiction of the South Atlantic Fishery Management Council and the states of North Carolina, South Carolina, Georgia, and Florida. The CIE reviewers shall have expertise in stock assessment, statistics, fisheries science, and marine biology to complete their primary task of conducting an impartial and independent CIE peer review report in accordance with the ToR to determine if the best available science is utilized for fisheries management. The CIE reviewers shall not provide comments on fisheries management decisions.

Statement of Tasks for CIE Reviewers:

The CIE reviewers shall complete the following tasks and responsibilities as described in the SoW and Schedule herein.

1. CIE shall provide the CIE reviewers' contact information (name, affiliation, address, email, and phone) to the Office of Science and Technology COTR no later than the date as specified in the SoW, and the COTR will forward this information to the Project Contact.

2. Approximately two weeks before the peer review, the Project Contact will send the CIE reviewers the necessary documents for the peer review, including supplementary documents for background information. The CIE reviewers shall read the pre-review documents in preparation for the peer review to gain an in-depth understanding of the stock assessment, the resources and information considered in the assessment, and responsibilities as reviewers. Meeting materials will be forwarded electronically to review panel members and made available through the internet (http://www.sefsc.noaa.gov/sedar/), and printed copies of any documents are available by request. The names of reviewers will be included in workshop briefing materials. The list of pre-review documents may be updated prior to the panel review meeting.

3. Each CIE reviewer shall participate on the SEDAR 17 workshop panel (refer to attached agenda) to conduct an impartial and independent peer review with the purpose of determining whether the best available science was utilized. CIE reviewers shall conduct an independent peer review and participate in panel discussions on assessment methods, data, validity, results, uncertainties, recommendations, and conclusions as guided by the terms of reference.

4. Each CIE reviewer shall produce an independent peer review report addressing each of the ToR 1-9 specified herein. The CIE independent peer review report shall be completed in accordance with the Schedule of Milestones and Deliverables specified herein. These reports shall be submitted to the CIE regional coordinator, Dr. David Sampson, via email to David.Sampson@oregonstate.edu, and to CIE lead coordinator, Mr. Manoj Shivlani, via email to shivlanim@bellsouth.net. See Annex II for complete details on the independent peer review report outline.

5. The CIE reviewers will also participate in development of a peer review consensus report for each assessment reviewed, in accordance with ToR 10 and as described in Annex I. CIE reviewers may be asked to serve as an assessment leader during the review to facilitate preparing first drafts of review summary reports. Following the review workshop, CIE reviewers will assist the chair in the development of the peer review consensus reports.

The review workshop will take place at the Hampton Inn and Suites, Savannah Historic District, 201 Martin Luther King Boulevard, Savannah, GA, from 1:00 p.m. Monday, October 20, 2008 through 1:00 p.m. Friday, October 24, 2008. The Project Contact is responsible for the facility arrangements.
Please contact Dale Theiling (SEDAR Coordinator); (843) 571-4366, Dale.Theiling@safmc.net) or John Carmichael, (Science and Statistics Program Manager); (843) 571-4366, John.Carmichael@safmc.net) for additional details.

Hotel arrangements:

Hampton Inn and Suites, Savannah Historic District
201 Martin Luther King Boulevard
Savannah, GA 31401
(912) 721-1600
"SEDAR" Group rate: $ 111.24; rate is guaranteed through September 8, 2008.

SEDAR Review Workshop Panel Tasks:

The SEDAR 17 review workshop panel will evaluate assessments of South Atlantic vermilion snapper and South Atlantic Spanish mackerel. During the evaluation the panel will consider data, assessment methods, and model results. The evaluation will be guided by terms of reference that are specified in advance. The review workshop panel will document its findings regarding each assessment in a peer review consensus report (Annex I). (Note that the consensus report is a SEDAR product, not a CIE product.) CIE reviewers shall participate on the SEDAR 17 workshop panel, conduct independent peer reviews, and produce CIE independent peer review reports to provide distinct, independent analyses of the technical issues and of the SEDAR process (refer to Statement of Tasks for CIE Reviewers). Each CIE reviewer shall contribute to a SEDAR consensus report in accordance with Annex I that will be compiled by the review panel

Chair, and shall produce a CIE independent peer review report in accordance with Annex II.

Terms of Reference:

SEDAR 17 Review Workshop Terms of Reference (apply to each stock):

1. Evaluate the adequacy, appropriateness, and application of data used in the assessment*.
2. Evaluate the adequacy, appropriateness, and application of methods used to assess the stock*.
3. Recommend appropriate estimates of stock abundance, biomass, and exploitation*.
4. Evaluate the methods used to estimate population benchmarks and management parameters (*e.g., MSY, Fmsy, Bmsy, MSST, MFMT, or their proxies*); provide estimated values for management benchmarks, a range of ABC, and declarations of stock status*.
5. Evaluate the adequacy, appropriateness, and application of the methods used to project future population status; recommend appropriate estimates of future stock condition* (e.g., exploitation, abundance, biomass).
6. Evaluate the adequacy, appropriateness, and application of methods used to characterize uncertainty in estimated parameters. Provide measures of uncertainty for estimated parameters*. Ensure that the implications of uncertainty in technical conclusions are clearly stated.
7. Ensure that stock assessment results are clearly and accurately presented in the Stock Assessment Report and Advisory Report and that reported results are consistent with Review Panel recommendations**.
8. Evaluate the SEDAR Process. Identify any Terms of Reference which were inadequately addressed by the Data or Assessment Workshops; identify any additional information or assistance which will improve Review Workshops; suggest improvements or identify aspects requiring clarification.
9. Review the research recommendations provided by the Data and Assessment workshops and make any additional recommendations warranted. Clearly indicate the research and monitoring needs that may appreciably improve the reliability of future assessments. Recommend an appropriate interval for the next assessment.
10. Prepare a Peer Review Consensus Summary summarizing the Panel's evaluation of the stock assessment and addressing each Term of Reference. Develop a list of tasks to be completed following the workshop. Complete and submit the Consensus Report within 3 weeks of workshop conclusion.

* The review panel may request additional sensitivity analyses, evaluation of alternative assumptions, and correction of errors identified in the assessments provided by the assessment workshop panel; the review panel may not request a new assessment. Additional details regarding the latitude given the review panel to deviate from assessments provided by the

assessment workshop panel are provided in the *SEDAR Guidelines* and the *SEDAR Review Panel Overview and Instructions*.

** The panel shall ensure that corrected estimates are provided by addenda to the assessment report in the event corrections are made in the assessment, alternative model configurations are recommended, or additional analyses are prepared as a result of review panel findings regarding the TORs above.

These Terms of Reference may be modified prior to the Review Workshop. If so, final terms of reference will be provided to the reviewers with the workshop briefing materials.

SEDAR Review Workshop Panel Supplementary Instructions

The review panel chair is responsible for reviewing documents prior to the workshop, conducting the workshop in an orderly fashion, compiling and editing the peer review consensus report for each species assessed and submitting it to the SEDAR Coordinator by a deadline determined by the SEDAR Steering Committee and specified in the Schedule of Deliverables. The review panel chair will work with SEDAR staff to complete the SEDAR summary report. The review panel chair may participate in panel deliberations and contribute to report preparation.

Review panel members are responsible for: (1) reviewing documents prior to the workshop, (2) participating in workshop discussions addressing the terms of reference, (3) preparing assessment summaries and consensus reports during the workshop, and (4) finalizing SEDAR documents within three weeks of the conclusion of the workshop. Each reviewer appointed by the CIE is responsible for preparing an independent CIE peer review report.

The chair and SEDAR coordinator will work with the appointed reviewers to assign tasks during the workshop. For example, the chair may appoint one panelist to serve as assessment leader for each assessment covered by the review, with the leader responsible for providing an initial draft consensus report text for consideration by the panel. Reviewers may alternatively be assigned particular terms of reference to address initially. Regardless of how initial drafting is accomplished, all panelists are expected to participate in discussion of all terms of reference and contribute to all aspects of the review.

The review panel's primary responsibility is to determine if assessment results are based on sound science, appropriate methods, and appropriate data. During the course of the review, the panel is allowed limited flexibility to deviate from the assessment provided by the assessment workshop. This flexibility may include: (1) modifying the assessment configuration and assumptions, (2) requesting a reasonable number of sensitivity runs, (3) requesting additional details and results of the existing assessments, and (4) requesting correction of any errors identified. However, the allowance for flexibility is limited, and the review panel is not authorized to conduct an alternative assessment or to request an alternative assessment from the technical staff present. The review panel is responsible for applying its collective judgment in determining whether proposed changes and corrections to the presented assessment are sufficient to constitute an alternative assessment. The review panel chair will coordinate with the SEDAR coordinator and technical staff present to determine which requests can be accomplished and to prioritize desired analyses.

Any changes in assessment results stemming from modifications or corrections solicited by the review panel will be documented in an addendum to the assessment report. If updated estimates are not available for review by the conclusion of the workshop, the review panel shall consult with technical staff present and the SEDAR coordinator to develop an acceptable process for reviewing the final results within the time allotted for completion of the project.

The review panel should not provide advice addressing specific management actions. Such advice will be provided by existing Council committees, such as the Science and Statistical Committee and advisory panels, following completion of the assessment. The review panel is free to point out items of concern regarding past or present management actions that relate to population conditions or data collection efforts.

If the review panel finds an assessment deficient to the extent that technical staff present cannot resolve the deficiencies during the course of the workshop, or the panel deems that desired modifications would result in a new assessment, then the review panel shall provide in writing the required remedial measures, including an appropriate approach for correcting and subsequently reviewing the assessment.

Workshop Final Reports:

The SEDAR coordinator will send copies of the final review panel consensus report and the complete SEDAR stock assessment report for each stock assessed to Mr. Manoj Shivlani at the CIE.

Submission and Acceptance of CIE Reports:

Upon review and acceptance of the CIE reports by the CIE Coordination and Steering Committees, CIE shall send via e-mail the CIE reports to the COTR (William Michaels William.Michaels@noaa.gov at the NMFS Office of Science and Technology by the date in the Schedule of Deliverables. The COTR will review the CIE reports to ensure compliance with the SoW and ToR herein, and have the responsibility of approval and acceptance of the deliverables. Upon notification of acceptance, CIE shall send via e-mail the final CIE report in *.PDF format to the COTR. The COTR at the Office of Science and Technology have the responsibility for the distribution of the final CIE reports to the project contacts.

The COTR shall provide the final CIE reviewer reports to:

SEFSC Acting Director: Bonnie Ponwith, NMFS Southeast Fisheries Science Center, 75 Virginia Beach Drive, Miami, FL 33149 (email, Bonnie.Ponwith@NOAA.gov)

SEDAR Coordinator: Dale Theiling, SAFMC, 4055 Faber Place Drive, Suite 201, North Charleston, SC 29405 (email, Dale.Theiling@safmc.net). (SEDAR shall provide the final CIE Reviewer Reports to the SEDAR Steering Committee and Executive Directors of those Councils having jurisdiction over the included stocks.)

Schedule of Milestones and Deliverables:

September 15, 2008:	CIE will provide the CIE reviewer contact information to the COTR who will in turn forward this to the Project Contact.
October 6, 2008:	The CIE reviewers will receive the pre-meeting documents from the Project Contact in preparation for the SEDAR 17 panel review meeting.
October 20-24, 2008:	The CIE reviewers shall participate during the SEDAR 17 panel review meeting, and conduct an independent peer review in accordance with the ToR.
October 24, 2008:	The CIE reviewers shall assist Chair in the development of the first draft of review panel consensus report(s) at the conclusion of the review workshop.
November 7, 2008:	Review panel members submit final review panel consensus report(s) contributions to workshop Chair.
November 14, 2008:	Workshop Chair submits final review panel consensus report(s) and SEDAR summary reports to SEDAR Coordinator.
November 14, 2008:	CIE reviewers shall submit their independent peer review reports to CIE.
December 1, 2008:	SEDAR Coordinator submits final review panel consensus report(s) and SEDAR stock assessment report(s) to CIE.
December 1, 2008:	CIE submits individual CIE reviewer reports to the COTR.
December 5, 2008:	COTR notifies CIE regarding individual reviewer report acceptance.
December 8, 2008:	CIE provides final individual CIE reviewer reports to COTR.
December 15, 2008:	COTR provides final CIE reviewer reports to SEFSC (Acting) Director and SEDAR Coordinator.
December 19, 2008:	SEDAR submits individual CIE reviewer reports to the SEDAR Steering Committee and Councils.

Key Personnel:

Contracting Officer's Technical Representative (COTR):

William Michaels
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

Stephen K. Brown
NMFS Office of Science and Technology
1315 East West Hwy, SSMC3, F/ST4, Silver Spring, MD 20910
Stephen.K.Brown@noaa.gov Phone: 301-713-2363 ext 133

<u>Contractor Contacts</u>:

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

SEDAR Project Contact (or Emergency):

Dale Theiling, 4055 Faber Place Drive, Suite 201, North Charleston, SC 29405
Dale.Theiling@safmc.net Phone: 843-571-4366.

Request for Changes:

Requests for changes shall be submitted to the Contracting Officer at least 15 working days prior to making any permanent substitutions. The Contracting Officer will notify the Contractor within 10 working days after receipt of all required information of the decision on substitutions. The contract will be modified to reflect any approved changes. The Terms of Reference (ToR) and list of pre-review documents herein may be updated without contract modification as long as the role and ability of the CIE reviewers to complete the SoW deliverable in accordance with the ToR are not adversely impacted.

DRAFT AGENDA

SEDAR 17 REVIEW WORKSHOPS

South Atlantic Vermilion Snapper
South Atlantic Spanish Mackerel

October 20 - 24, 2008
Hampton Inn and Suites, Savannah, GA

Dr. TBN, Chair

Monday, October 20, 2008

1:00 p.m. **Convene**

1:00 – 1:30 **Introductions and Opening Remarks** **Mr.**
Dale Theiling
- Agenda review, TOR review, and Task assignments **Chair**

1:30 – 3:30 **Vermilion Snapper Presentation** **Dr.**

Kyle Shertzer

3:30 – 3:45 **Break**

3:45 – 6:00 **Vermilion Snapper Discussion** **Chair**
- Data, Methods and Results evaluation
- Identify additional analyses, sensitivities, and corrections

Tuesday, October 21, 2008

8:00 a.m. – 12:00 p.m. **Vermilion Snapper Discussion** **Chair**
- Review additional analyses and sensitivities
- Initial recommendations and comments

12:00 p.m. – 2:00 p.m. **Lunch Break**

2:00 p.m. – 4:00 p.m. **Spanish Mackerel Assessment Presentation** **Dr.**

Paul Conn

4:00 p.m. – 4:15 p.m. **Break**

4:15 p.m. – 6:15 p.m. **Spanish Mackerel Discussion** **Chair**
- Data, Methods and Results evaluation
- Identify additional analyses, sensitivities, and corrections

Wednesday, October 22, 2008

8:00 a.m. – 12:00 p.m. **Spanish Mackerel Discussion** **Chair**
- Review additional analyses and sensitivities
- Initial recommendations and comments

12:00 p.m. – 2:00 p.m. **Lunch Break**

2:00 p.m. – 4:00 p.m. **Vermilion Snapper and Spanish Mackerel** **Chair/**

Discussion as needed **Stock Leaders**

4:00 p.m. – 4:15 p.m. **Break**

4:15 p.m. – 6:15 p.m. **Vermilion Snapper and Spanish Mackerel** **Chair/**
Discussion as needed **Stock Leaders**

<u>Thursday, October 23, 2008</u>

8:00 a.m. – 12:00 p.m. **Review Workshop Consensus Summary Chair/Stock**
- Review draft Consensus Report sections
Leaders

12:00 p.m. – 2:00 p.m. **Lunch Break**

2:00 p.m. – 5:00 p.m. **Review Workshop Advisory Report Chair/Stock**
Review draft Summary Reports
Leaders

<u>Friday, October 24, 2008</u>

8:00 a.m. – 12:00 p.m. **Final Review of Panel Documents** **Chair**
- Final review of Consensus Reports and Summary Reports

12:00 p.m. **ADJOURN**

The timing of particular events is tentative, and the Chair may modify this schedule during the workshop as needed to complete stated tasks. However, to accommodate travel planning the workshop will start as scheduled and will conclude no later than the stated time.

SEDAR is a public process, and the public is welcome to attend SEDAR workshops. Although no formal public comment period is scheduled, the workshop Chair will allow opportunity during the meeting for the public in attendance to comment on discussion items.

Annex I. SEDAR Review Panel Consensus Summary Report Contents

I. Terms of Reference

List each Term of Reference and provide a summary of Panel discussions and recommendations regarding the particular item. Include a clear statement indicating whether or not the criteria in the Term of Reference are satisfied.

II. Further Analyses and Evaluations

Summary and findings of review panel analytical requests not previously addressed in TOR discussion above.

III. Additional Comments

Summary of any additional discussions not captured in the Terms of Reference statements.

IV. Recommendations for Future Workshops

Panelists are encouraged to provide general suggestions to improve the SEDAR process.

V. Reviewer Statements

Each individual reviewer should provide a statement attesting whether or not the contents of the Consensus Report provide an accurate and complete summary of their views on the issues covered in the review. Reviewers may also make any additional individual comments or suggestions desired.

ANNEX II: Contents of CIE Independent Peer Review Report

1. The reviewer report shall be prefaced with an executive summary of findings and recommendations.

2. The main body of the reviewer report shall consist of a background, description of the individual reviewer's role in the review activities, a summary of findings, and summary of conclusions and recommendations in accordance with the ToR.

a. Reviewers should described in their own words the review activities completed during the meeting, including providing a detailed summary of findings, conclusions, and recommendations.

b. Reviewers should discuss their views on each ToR even if these were consistent with those of the panel 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 SEDAR process including suggestions for improvements of both process and products.

e. While it is expected that reviewers would not simply repeat the contents of the summary report, the report is to represent a stand-alone document that could be used by others who may not have read the summary report to be able to understand the proceedings and findings of the meeting.

3. The reviewer report shall include as separate appendices a copy of the CIE Statement of Work and a bibliography that includes all materials provided for review.

Appendix 2: Bibliography of supplied material

Document #	Title	Authors
Documents Prepared for the Data Workshop		
SEDAR17-DW01	South Atlantic Vermilion Snapper Management Information Worksheet	J. McGovern (SERO) R. DeVictor (SAFMC)
SEDAR17-DW02	South Atlantic Spanish Mackerel Management Information Worksheet	J. McGovern (SERO) R. DeVictor (SAFMC)
SEDAR17-DW03	South Atlantic Vermilion Snapper Assessment History	D. Vaughan (SEFSC)
SEDAR17-DW04	South Atlantic Spanish Mackerel Assessment History	D. Vaughan (SEFSC)
SEDAR17-DW05	South Atlantic Vermilion Snapper Commercial Chapter	D. Vaughan (SEFSC)
SEDAR17-DW06	South Atlantic Spanish Mackerel Commercial Chapter	D. Vaughan (SEFSC)
SEDAR17-DW07	A review of Spanish mackerel (*Scomberomorus maculatus)* age data, 1987-2007, Atlantic collections only, from the Panama City Laboratory, SEFSC, NOAA Fisheries Service	C. Palmer, D. DeVries, C. Fioramonti and L. Lombardi-Carlson (SEFSC)
SEDAR17-DW08	Vermilion Snapper Length Frequencies and Condition of Released Fish from At-Sea Headboat Observer Surveys in the South Atlantic, 2004 to 2007	B. Sauls, C. Wilson, D. Mumford, and K. Brennan (SEFSC)
SEDAR17-DW09	Development of Conversion Factors for Different Trap Types used by MARMAP since 1978.	P. Harris (MARMAP)
SEDAR17-DW10	Discards of Spanish Mackerel and Vermilion Snapper Calculated for Commercial Vessels with Federal Fishing Permits in the US South Atlantic	K. McCarthy (SEFSC)
SEDAR17-DW11	Standardized catch rates of vermilion snapper from the headboat sector: Sensitivity analysis of the 10-fish-per-angler bag limit	Sustainable Fisheries Branch (SEFSC)

SEDAR17-DW12	Estimation of Spanish mackerel and vermilion snapper bycatch in the shrimp trawl fishery in the South Atlantic (SA)	K. Andrews (SEFSC)
Documents Prepared for the Assessment Workshop		
SEDAR17-AW01	SEDAR 17 South Atlantic Vermilion Snapper Stock Assessment Model	To be prepared by SEDAR 17
SEDAR17-AW02	SEDAR 17 South Atlantic Spanish Mackerel Stock Assessment Model	To be prepared by SEDAR 17
SEDAR17-AW03	Development of an aging error matrix for the vermilion snapper catch-at-age stock assessment model	E. Williams (SEFSC)
SEDAR17-AW04	Catch curve analysis of age composition data for Spanish mackerel	E. Williams (SEFSC)
SEDAR17-AW05	Catch curve analysis of age composition data for vermilion snapper	E. Williams (SEFSC)
SEDAR17-AW06	Methods for combining multiple indices into one, with application to south Atlantic (U.S.) Spanish mackerel	P. Conn (SEFSC)
SEDAR17-AW07	Extrapolation of Spanish mackerel bycatch by commercial shrimp trawl fisheries	P. Conn (SEFSC)
SEDAR17-AW08	A Bayesian approach to stochastic stock reduction analysis, with application to south Atlantic Spanish mackerel	P. Conn (SEFSC)
SEDAR17-AW09	Preliminary Surplus–production Model Results of Vermilion Snapper off the Southeastern United States	R. Cheshire (SEFSC)
SEDAR17-AW10	Preliminary Surplus–production Model Results of Spanish Mackerel off the Southeastern United States	R. Cheshire (SEFSC)
SEDAR17-AW11	AD Model Builder code to implement catch-age	K. Shertzer (SEFSC)

	assessment model of vermilion snapper	
SEDAR17-AW12	AD Model Builder code to implement catch-age assessment model of Spanish mackerel	P. Conn (SEFSC)
SEDAR17-AW13	ASCII file populated by results of VS base catch-age model	K. Shertzer (SEFSC)
Documents Prepared for the Review Workshop		
SEDAR17-RW01	SEDAR 17 South Atlantic Vermilion Snapper Document for Peer Review	To be prepared by SEDAR 17
SEDAR17-RW02	SEDAR 17 South Atlantic Spanish Mackerel Document for Peer Review	To be prepared by SEDAR 17
Final Assessment Reports		
SEDAR17-AR01	Assessment of the Vermilion Snapper Stock in the US South Atlantic	To be prepared by SEDAR 17
SEDAR17-AR02	Assessment of the Spanish Mackerel Stock in the US South Atlantic	To be prepared by SEDAR 17
Reference Documents		
SEDAR17-RD01	South Atlantic Vermilion Snapper Stock Assessment Report, SEDAR 2, 2003	SEDAR 2
SEDAR17-RD02	Update of the SEDAR 2 South Atlantic Vermilion	SEDAR
SEDAR17-RD03	Fishery Management Plan for Spanish Mackerel, Atlantic States Marine Fisheries Commission, 1990	L. P. Mercer L. R. Phalen J. R. Maiolo
SEDAR17-RD04	Mitochondrial and nuclear DNA analysis of population subdivision among young-of-the-year Spanish mackerel (*Scomberomorus maculatus*) from the western Atlantic and Gulf of Mexico	V. P. Buonaccorsi E. Starkey J. E. Graves
SEDAR17-RD05	George Fishes MD TAFS 28 1-49	W. A. George
SEDAR17-RD06	Excerpt – Goode 1878 stats 7-1-99	Goode
SEDAR17-RD07	Excerpt – Henshall Comparative Excellence TAF 13 1-115	Henshall

SEDAR17-RD08	Stock Assessment Analyses on Spanish and King Mackerel Stocks, April 2003	Sustainable Fisheries Div, SEFSC
SEDAR17-RD09	Hooking Mortality of Reef Fishes in the Snapper-Grouper Commercial Fishery of the Southeastern United States	D.V. Guccione Jr.
SEDAR17-RD10	Effects of cryptic mortality and the hidden costs of using length limits in fishery management Lewis G Coggins Jr	L. G. Coggins Jr. and others
SEDAR17-RD11	Discard composition and release fate in the snapper and grouper commercial hook-and-line fishery in North Carolina, USA	P. J. Rudershausen and J. A. Buckel
SEDAR17-RD12	A multispecies approach to subsetting logbook data for purposes of estimating CPUE	A. Stephens and A. MacCall
SEDAR17-RD13	The 1960 Salt-Water Angling Survey, USFWS Circular 153	J. R. Clark
SEDAR17-RD14	The 1965 Salt-Water Angling Survey, USFWS Resource Publication 67	D. G. Deuel and J. R. Clark
SEDAR17-RD15	1970 Salt-Water Angling Survey, NMFS Current Fisheries Statistics Number 6200	D. G. Deuel
SEDAR17-RD16	User's Guide: Delta-GLM function for the R Language /environment (Version 1.7.2, revised 07-06-2006)	E. J. Dick (SWFSC/NMFS)
SEDAR17-RD17	Reproductive biology of Spanish mackerel, *Scomberomorus maculatus*, in the lower Chesapeake Bay. M.A. Thesis, Virginia Institute of Marine Science. (Selective pages)	C. L. Cooksey
SEDAR17-RD18	The summer flounder chronicles: Science, politics, and litigation, 1975–2000	M. Terceiro
SEDAR17-RD19	Use of Angler Diaries to Examine Biases Associated with 12-Month Recall on Mail	N. Connelly and T. Brown

	Questionnaires	
SEDAR17-RD20	Comparing 1994 Angler Catch and Harvest Rates from On-Site and Mail Surveys on Selected Maine Lakes	B. Roach
SEDAR17-RD21	Response Errors in Canadian Waterfowl Surveys	A. Sen
SEDAR17-RD22	Exaggeration of Walleye Catches by Alberta Anglers	M. Sullivan
SEDAR17-RD23	Effects of Recall Bias and Non-response Bias on Self-Report Estimates of Angling Participation	M. A. Tarrant and M. J. Manfredo
SEDAR17-RD24	Influence of Survey Method on Estimates of Statewide Fishing Activity	T. Thompson
SEDAR 17-RD25	Final Amendment 6 to the Fishery Management Plan for the Shrimp Fishery of the South Atlantic Region	SAFMC, 2004