Report on SEDAR 27:

Gulf of Mexico menhaden

review

November 1 – 4, 2011

St. Petersburg, Florida

P.L. Cordue

Fisheries consultant

New Zealand

For **The Center for Independent Experts**

28 November 2011

Executive summary

During 1-4 November 2011, a SEDAR Review Workshop was convened in St. Petersburg, Florida, to review a draft stock assessment of Gulf of Mexico menhaden. I was a member of a five-person Review Panel, which included three CIE reviewers.

The base model and the main sensitivities were developed using a forward-projecting statistical model as implemented in the Beaufort Assessment Model (BAM). Two supporting models were also used: a Bayesian model configured with *MSY* and F_{MSY} as key parameters and termed a Stock Reduction Analysis (SRA); and the non-equilibrium production model ASPIC.

Data inputs included a well-characterized landings history from 1948–2010 and an extensive catch-at-age matrix from 1964–2010. Two independent time series of juvenile-abundance indices were available from Gulf-state sampling of coastal waters. The base model included a purported adult-abundance time series derived from fishery-independent gillnet sampling in the coastal waters of Louisiana. An alternative adult-abundance index derived from the reduction-fishery CPUE from 1948–2010 was also used in some runs.

The menhaden assessment was not technically defensible and was rejected by the Panel. The gillnet indices used in the base model appear dubious as indices of adult abundance (since they measure density from part of the edge of the adult distribution) and the alternative CPUE indices are clearly problematic over such an extended timeframe in a changing and targeted purse-seine fishery. The base model, implemented in BAM, was poorly structured and data weighting was inappropriate. The supporting SRA and ASPIC models could not be taken to provide reliable estimates of stock status or benchmarks on their own. The SRA model runs used output from the BAM base model and did not use the extensive catch-at-age data. Likewise, the ASPIC model could not use the catch-at-age data and also did not incorporate important biological characteristics of the species.

Remedial actions to provide a reliable quantitative stock assessment are detailed in this report. Work is required on the data inputs to (perhaps) derive more defensible adult abundance indices and good practice needs to be adopted while developing a base model and sensitivities within the BAM framework.

A formal recommendation on stock status was not made given the absence of an accepted quantitative stock assessment. However, a qualitative assessment of the data by the Panel, in conjunction with some "worst case" model runs, did allow the Panel to tentatively suggest that the stock is probably not over-fished and that over-fishing is probably not occurring.

To improve the accuracy and precision of future menhaden stock assessments it is recommended that a reliable fishery-independent adult-abundance index is developed. This cannot be done by surveying only coastal waters and will require a survey covering most of the spatial extent of the adult population.

Background

During 1-4 November 2011, a SEDAR Review Workshop was convened in St. Petersburg, Florida, to review a draft stock assessment for Gulf of Mexico menhaden (SEDAR27-SAR1). The Review Panel operated under the SEDAR review workshop guidelines although the Data and Assessment workshops had been convened under the GSMFC review process.

I was one of three CIE reviewers appointed to the five-person Review Panel. The meeting had a non-CIE Chair and a GSMFC 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 (ToRs) for the review (Appendix 2, annex 2). There was general agreement among the Panel with regard to most findings and recommendations and I expect that my findings will be consistent with those in the joint report. However, I have yet to see a full draft of the joint summary report.

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 document in detail prior to the meeting.

The ToRs for the review were identical to the ToRs for the data and assessment workshops. I suggested to CIE that the ToRs be changed to reflect that we were reviewing rather doing the assessment. One change was made, to ToR 5, in that we were no longer required to actually "perform retrospective analyses" and there was an assurance that more appropriate ToRs would be produced for future reviews.

We were advised by the SEDAR coordinator that yellowtail snapper would no longer be reviewed at the meeting as the assessment was not ready for review.

I participated in a breakfast meeting with the Chair, Panel members, and the SEDAR coordinator on the first morning of the meeting. Initial writing assignments for the joint report were assigned. However, these were later revised given the outcome of the review.

Meeting

The meeting began on schedule at 9.00 am on Tuesday morning and generally followed the agenda for the first two days (Appendix 2, annex 3). However, due to the Panel's eventual rejection of the assessment, the meeting was adjourned early, on Thursday afternoon. The Panel did continue to meet, for discussion of the joint report, through to the scheduled conclusion at lunchtime on Friday.

Tuesday was mainly taken up with presentations on the fishery-dependent and fishery independent data and the BAM model runs. The presentations were excellent and good discussions were held on the merits of the various data sources. Concerns were raised by myself and others with regard to the reliability of the adult abundance indices. The obvious problems with the reduction-fishery CPUE indices were acknowledged by everyone. The views on the reliability of the gillnet indices were more mixed. Some Panel members thought that it could be a reliable index. I indicated that my main concern was that it was sampling on the edge of the adult distribution. There was also the issue, with all of the fishery independent indices, that they were measuring density rather than abundance (i.e., biomass is density multiplied by area; the state-sampling by gillnet, seine, and trawl, were all just measures of density in coastal waters).

I requested that a map be presented showing how much/little the gillnet sampling stations overlapped with the fishery (to provide an indication of whether the sampling was on the edge or in the "middle" of the adult distribution). I also queried the diagnostic plots shown in the report for the residuals of the fishery independent indices (derived using a delta-lognormal model). These were plotted by year and the residuals were clearly not centered around zero, in each year, as they should have been. I requested that we see more diagnostics from the derivation of the indices and that the residual plots in the assessment report be explained.

On Wednesday, new plots of the residuals by year were produced for the fishery-independent indices showing that the residuals were indeed centered about zero. The plots in the document were in error. It was perhaps of little consequence, but the plots in the document had been presented because they had been the basis for excluding the first few years of some of the indices (because of the residual pattern shown). There was further discussion on the reliability of the gillnet indices. In response to the Chair, I made the point that seeing (for the first time) sensible diagnostics for the gillnet time series had little bearing on whether it could be considered as a reliable index of abundance.

A map of gillnet stations was presented and it illustrated that there was very little overlap between the station locations and the fishery. It was pointed out that the fishery was forced offshore to some extent by regulation and that the coastal stations may well still be in "prime" menhaden habitat. I reiterated that the stations, even if in good menhaden habitat, were clearly just on the edge of the adult distribution. I pointed out that a reliable index had to cover a large proportion of the spatial extent of the population's distribution. It was suggested that "mixing rates" were the determinant of whether sampling at the edge was adequate to index abundance. I suggested that it was more to do with how the fish distributed themselves spatially. For sampling at the edge to be adequate required that a doubling of the biomass doubled the density of the fish in the costal waters. This might be expected to work in a swimming pool, but seemed less likely in the open ocean where the population could expand and contract spatially.

The Panel requested some exploratory runs of the BAM model to try to determine why there was such a strong residual pattern in the catch-at-age data. While the assessment author did these runs, presentations were made on the SRA and ASPIC models. The SRA

model had used inputs from the BAM base model which was problematic given concerns the Panel had about that run. The main runs had also been done using the less-preferred state-wide gillnet indices, because the SRA model had had great difficulty fitting the Louisiana-only gillnet indices (no doubt because of the big spikes in 2008 and 2009).

The ASPIC runs had used the state-wide seine and trawl juvenile indices, (offset by one year, to make them look like adult indices) in various combinations with the reduction-CPUE and Louisiana gillnet indices. An annual 1% adjustment had been made to the reduction CPUE to allow for improvements in fishing efficiency from 1948 to 2010. Near the end of the day, I suggested that some "worst" case ASPIC runs be done showing the effect of assuming a 2% annual increase in efficiency. At this stage we had seen that we were making little progress with the BAM exploratory runs, so I thought it might be useful to have the additional ASPIC runs to see if they revealed any possible concerns with regard to stock status.

Before adjourning for the day, the Panel, in consultation with the assessment author, formulated some more exploratory BAM runs to be performed overnight.

On Thursday morning the additional BAM and ASPIC runs were presented. The BAM runs all had a problem with the initial biomass; the model having found its best fit at a very low biomass in 1948 with subsequent recruitment being very strong. The Panel had been aware that there was a penalty to encourage the initial biomass to be equal to virgin biomass. However, we had not realized that the penalty would have very little effect (in the exploratory runs). In the BAM base model the low initial biomass had been eliminated by down-weighting early age data – an entirely inappropriate response. Since the model was starting in 1948 and there were no data to inform with regard to initial biomass the specification should have required that initial biomass be equal to virgin biomass. A run with this specification was requested after the "failed" exploratory runs had been partly presented. The results showed a sensible looking biomass trajectory (compared to the other exploratory runs) and caused some consternation amongst some Panel members. They were concerned that an apparently simple change had dramatically affected the model results. They wondered what other switches could be changed and how that might affect results.

The additional ASPIC runs showed that it was possible that the stock could be at or below B_{MSY} and that over-fishing might be occurring – with the qualification that these were "worst" case scenarios.

There was much discussion about how the meeting should proceed. The Panel agreed that we were rejecting the assessment and that we therefore could not recommend a stock status. However, we did agree that we could provide some qualitative advice on stock status using the catch history in conjunction with the catch-at-age data, and the "worst" case ASPIC runs.

After the Panel agreed to reject the assessment, I was assigned the task of producing initial drafts of "Qualitative advice on stock status", "Remedial actions necessary to

provide a reliable stock assessment", and "Research recommendations". In my report I have made only minor modifications to the initial drafts of these sections (so, for these sections, my report and the joint report are very similar).

Post-meeting

On my return to New Zealand I made some agreed changes to my draft sections for the joint report and emailed the revised sections to the Panel members for comment. A single comment was received and I made a minor revision to the sections, which were sent to the Chair for his use in compiling the joint report.

Two of the other Panel members circulated their draft sections for comment. I reviewed the sections and suggested some changes. At the time of completing this report I have not seen the final draft section from one of the Panel members or the final version of the joint summary report.

Summary of findings

The menhaden assessment was not technically defensible. The gillnet indices used in the base model appear dubious as indices of adult abundance and the alternative CPUE indices are clearly problematic over such an extended timeframe in a changing fishery. The base model, implemented in BAM, was poorly structured and the data weighting was inappropriate. The supporting SRA and ASPIC models could not be taken to provide reliable estimates of stock status or benchmarks on their own. The SRA model runs used output from the BAM base model and did not use the extensive catch-at-age data. Likewise, the ASPIC model could not use the catch-at-age data, or model important biological characteristics of the species.

Each of the TOR are specifically considered below.

1. *Evaluate precision and accuracy of fishery-dependent and fishery-independent data used in the assessment.*

   A single stock was assumed for the Gulf of Mexico menhaden. Data included landings from the commercial reduction fishery, an extensive catch-at-age matrix, two juvenile abundance indices (beach seine and trawl from Gulf-state sampling), and possible adult abundance indices from the reduction-fishery CPUE and fishery-independent gillnet sampling from Louisiana. In addition, growth, weight, and maturity at length/age were estimated using fishery dependent and independent information, while age-specific natural mortality was estimated from tagging data and an inverse relationship between size-at-age and natural mortality.

   Landings data from the commercial purse-seine reduction fishery are well characterized from 1948 to the present. The bait and recreational fisheries are insignificant in comparison to the reduction fishery (the bait fishery peaked at 2.5% of total landings in the mid 1990s; in recent years the bait and recreational

fisheries combined are less than 0.2% of total landings). The annual discards from the shrimp fishery were estimated from 1987 to 2010 for use in sensitivity runs. The estimates were small in terms of tonnage and in terms of the numbers of juveniles (assumed age 0) compared to the removals from the reduction fishery. The estimates were derived from menhaden CPUE in the SEAMAP surveys and shrimp-fishery effort. Since the SEAMAP surveys are mainly outside the range of menhaden (SEAMAP data were not used to provide menhaden indices for this reason) the estimates are probably biased low.

Data from biological sampling for length and age for the reduction fishery were available from 1964–2010. Samples are obtained from the "last catch of the day" using a dip-net on shore. Ages are determined using scales and the same single reader has been used since the early 1970s. The catch-at-age matrix is derived from weekly age-length keys at each port, which is possible because of the high sampling frequency relative to the number of age classes in the fishery.

Two data sources were used to estimate potential adult relative abundance indices: gillnet sampling of menhaden by the Gulf of Mexico states (1975–2010); and a simple total-catch over total-effort index (CPUE) from the reduction fishery (1948–2010).

The gillnet indices were calculated using all states and also using only Louisiana. The latter was preferred by the Assessment Workshop because Louisiana is at the center of the fishery and it was seen as problematic to combine the data across states (Louisiana used a "strike" gillnet while the other states set the gillnets for at least 1 hour). A delta-lognormal model was used to standardize the indices; catch rates/presence-absence at each site being explained by various fixed-effects including year, state, season, salinity, and temperature.

The reduction-fishery CPUE indices used an effort measure of vessel-ton-week, where at least one trip by a vessel caused its tonnage to be included in the effort total for the week. The annual index was the total landings divided by the total vessel-ton-weeks. The use of vessel-tons is an ad hoc allowance for changes in fleet compositions (larger vessels can catch more fish each week). An annual 1% increase in catching efficiency was also assumed to account for improvements in technology over the period. The assessment team were well-aware of the potential problems with this time series because of the nature of a targeted purse-seine fishery (i.e., such indices are probably hyper-stable). Detailed logbook data (Captains Daily Fishing Reports, CDFRs) are available from 1983 to 2010. These data had not been used to produce a standardized CPUE time series, presumably because of the same concerns.

Data from state beach-seine sampling (1974–2010) and trawl sampling (1967–2010) were used to construct two potential juvenile abundance time series. Data across all states were combined and a delta-lognormal model was used for standardization in each case.

Specific questions specified in TOR 1. are addressed below.

a. *Discuss data strengths and weaknesses (e.g. temporal and spatial scale, gear selectivities, aging accuracy, sampling intensity).*

Strengths of the fishery-dependent and fishery-independent data:

- The reduction fishery landings are well estimated since 1948 and they represent close to 100% of the landings.
- The biological sampling of the catch for age and length has been conducted in a consistent manner since 1964.
- Daily logbooks (CDFRs) are available since 1983, which detail purse-seine set locations and estimated catch.
- Scales have been used for ageing since the 1960s, and have been read by the same person since the early 1970s. Ageing error has been estimated from scale-scale comparisons and scale-otolith comparisons.
- The two potential juvenile abundance indices are positively correlated and show a very similar pattern of recruitment for the period in which they overlap.

Data weaknesses:

- The by-catch of age-0 menhaden in the shrimp fishery is not well estimated. It is probably not important given the assumed high natural mortality at that age, but the potential under-estimation of the numbers discarded could be substantial.
- Sampling just from the last catch of the day (or trip) is of some concern because of potential bias. For example, it may be that as boats fill up their holds that they fish in deeper water where they tend to catch larger fish. Such effects could be investigated to some extent using the existing CDFR data.
- The early catch-at-age data (pre 1971) contain very few old fish and this coincides with a period when multiple readers were used. The absence of old fish may simply be due to ageing error.
- There is no defensible adult abundance index available for the stock:
    - the reduction CPUE time series has only been standardized in an ad hoc fashion; it covers periods of enormous change in the fishery including its initial development and a substantial reduction in the number of factories and vessels; the CDFR data should be analyzed, but even after proper standardization, using only data from periods of relative stability, there will still be the concern that the indices are hyper-stable because it is a target purse-seine fishery supported by spotter planes;
    - the gillnet indices are derived from samples in coastal waters – they are sampling on the edge of the menhaden distribution (the fishery is necessarily further offshore due to regulation, but there is no doubt that the bulk of the population is found outside the coastal regions sampled by the states); combining the data across states is highly problematic

because of the differences in gear and given there may be different trends in each state and the required weighting across states will not be reflected in the sampling effort; the Louisiana-only indices are from using a gillnet as a "strike net" with the operator herding fish into the net by circling it two or three times – it seems that catch rates could be highly dependent on the individual operators each year (and this effect was not considered in the standardization).

- The construction of the juvenile indices from the trawl and seine data also suffer from the problem of relative weighting across states if the data show different trends within state over time; or if there have been shifts in the juvenile distribution. The delta-lognormal modeling uses weights based on the number of stations and this may be very different from the weightings that are actually required (given different trends within states and shifts of distribution). Properly stratified random surveys, which allow density to be scaled up to biomass by area, are needed to produce fully defensible abundance indices.
- The 2010 data points in each time series may have been compromised by changes in fish distribution due to the BP oil spill. Certainly, the fishery was radically different in 2010 and any CPUE analysis, aimed at producing abundance indices, should omit the 2010 data.

b. Report metrics of precision for data inputs and use them to inform the model as appropriate.

This is an instruction to the Assessment Team.
The use of the number of tows sampled, capped at 200, was appropriate as initial effective sample sizes for the catch-at-age data. The use of the CVs estimated by the delta-lognormal modeling of state-sampling seine, trawl, and gillnet was not appropriate. These time series are noisy indices at best and an appropriate level of process error must be "added on" when used in the assessment modeling.

c. Describe and justify index standardization methods.

Another instruction to the Assessment Team. The standardization methods were generally appropriate for the fishery independent time series although the documentation of diagnostics and estimated effects was inadequate. Also, state-year interactions should have been investigated for the indices that used data from multiple states. If the trends across time within each state are similar, then the data-weighting issue goes away.

d. Justify weighting or elimination of available data sources.

The choice of the gillnet time series as an adult abundance index in the base model is dubious but understandable. Because there is no defensible adult abundance index, it is a choice of the lesser evil (the gillnet and the ad hoc CPUE time series are both very unlikely to index abundance) or admitting that there isn't

a defensible adult index. Weighting of data sets was done in an ad hoc manner – see bullet points for "problems with the base model" below.

2. *Evaluate models used to estimate population parameters (e.g., F, biomass, abundance) and biological reference points.*

The Beaufort Assessment Model (BAM) was used to provide the main assessment runs. This is a statistical forward-projection model with separable selectivities using the Baranov catch equation. Catch histories, catch-at-age, juvenile and adult abundance indices were all fitted in the model assuming a single fishery. A complex parameterization was used for the selectivities with multiple time blocks and annual estimation of selection at age 1 during the most recent time block. Lognormal likelihoods were assumed for the catch histories and indices, with multinomial likelihoods for the catch-at-age data.

Natural mortality was assumed to be age dependent following an inverse relationship between size and natural mortality which was scaled to a tagging experiment estimate at age 2. A Beverton-Holt stock recruitment relationship was assumed and stock steepness was estimated. Initial biomass was allowed to differ from virgin biomass but a penalty was used to encourage equality.

The base model has a number of strengths:

- well tested software, population dynamics equations, and likelihoods;
- reasonable certainty in the catch history over an extended period;
- extensive catch-at-age data from the main fishery sampled in a consistent manner over many years;
- defensible recruitment indices (sampled from the main juvenile habitat and well correlated with each other).

However, there are many problems with the base model:

- the gillnet time series is not defensible as an adult abundance index (it measures density on one part of the edge of the adult distribution);
- there are strong residual patterns for the reduction fishery catch-at-age which were not able to be explained (use of a domed selectivity should have removed the pattern but it did not);
- data weightings were arrived at in a very ad hoc fashion (early age data were down-weighted to "fix" a problem with initial biomass deviating markedly from virgin biomass and causing unrealistic estimates of fishing mortality; subsequent iterative re-weighting was then over-ridden manually by adjusting likelihood multipliers);
- stock steepness is estimated even though the lack of contrast across $\%B_0$ indicates that there can be very little information in the data to support such an estimate;
- initial biomass is estimated in the absence of any data which could provide information to support the estimate (in the Panel's exploratory runs it was

shown that the penalty encouraging initial biomass to equal virgin biomass allowed huge deviations – they should simply be specified to be equal).

Specific questions specified in TOR 2. are addressed below.

a. Did the model have difficulty finding a stable solution?

This was not explicitly investigated by the Panel. Bounds were often hit in the Panel's exploratory runs, but this was because initial biomass was being estimated as very small and current biomass as being very large.

b. Were sensitivity analyses for starting parameter values, priors, etc. and other model diagnostics performed?

An extensive set of sensitivity runs were performed for the base model including higher and lower *M*, alternative weights on data sets, and alternative selectivities. None of the runs were close to what I would have considered an acceptable specification.

c. Have the model strengths and limitations been clearly and thoroughly explained?

These were discussed to some extent in the assessment document and were also considered by the Panel (see above).

d. Have the models been used in other peer reviewed assessments? If not, has new model code been verified with simulated data?

BAM has been used in several other peer-reviewed assessments. BAM is not the issue; the problems are with how it was used.

e. Compare and discuss differences among alternative models.

BAM, SRA, and ASPIC were all used in runs to assess stock status. They all gave similar results with the spawning biomass/egg-production estimated to be above B_{MSY} and F to be currently less than F_{MSY}. However, all of the models have technical deficiencies that render the assessment results unreliable. BAM has been discussed above. The SRA runs used outputs from the BAM base run and did not use the catch-at-age data. The results are highly dependent on the assumed prior distributions and the BAM model outputs, which are dubious. SRA should be developed to use the catch-at-age data so that it can provide an alternative stand-alone assessment for menhaden. ASPIC, as a production model (using either Schaffer or Fox), is not sufficiently complex to model important biological characteristics of menhaden or the fishery. It cannot use the catch-at-age data and only had available, what are, dubious adult abundance indices (i.e., the offset juvenile indices, and the dubious CPUE and gillnet indices).

3. *State and evaluate assumptions made for all models and explain the likely effects of assumption violations on model outputs.*

These were discussed under TOR 2. However, each point in the checklist is addressed below.

a. *Calculation of M.*

The assessment author investigated this using sensitivity runs, which gave predictable results.

b. *Choice of selectivity patterns.*

A very complex parameterization was used, presumably to try to reduce the residual patterns in the catch-at-age data. The whole model formulation, including the selectivity parameterization, needs to be revisited.

c. *Error in the catch-at-age matrix.*

There may be some bias in the catch-at-age data due to samples always being taken from the top of the fish hold at the end of a trip (assumed to be from the last catch of the day). There may also be ageing error or drift in age readings across time (despite using the same reader since about 1970). Model outputs may be biased if there are such biases in the catch-at-age data. Some of my recommendations concern investigating the potential magnitude and direction of bias in the catch-at-age data.

d. *Choice of a plus group for age-structured species.*

A plus group was used at an appropriate age.

e. *Constant ecosystem (abiotic and trophic) conditions.*

Ecosystem conditions are unlikely to have been constant over the period in which the stock was modeled. There are attempts in the model to deal with changing conditions in terms of year-specific mean weight-at-age.

f. *Choice of stock-recruitment function.*

A Beverton-Holt stock recruit relationship was assumed, which is the usual choice. A sensitivity run was done using a Ricker relationship.

g. *Choice of reference points (e.g. equilibrium assumptions).*

The use of explicit MSY based reference points is inappropriate given that a limited range of spawning stock sizes (in terms of $\%B_0$) appears to have been

observed. This makes the estimation of F_{MSY} and B_{MSY} very uncertain. It would be better, at this stage, to use a proxy (e.g., choose a target level of F after an evaluation of equilibrium conditions based on a range of plausible biological parameters – including stock-recruit relationships).

4. *Evaluate uncertainty of model estimates and biological or empirical reference points.*

The assessment was rejected because of the large potential for bias in the assessment results given the technical deficiencies of the methods used. The use of bootstrapping in the BAM models is not ideal for the characterization of uncertainty. It would be better to use a Bayesian approach such as in SRA.

Specific questions specified in TOR 5 are addressed below.

a. *Choice of weighting likelihood components.*

An ad hoc and inappropriate weighting method was used in the BAM runs. Early catch-at-age data were down-weighted to try to fix a structural problem in the model specification. Iterative re-weighting was then applied but then was over-ridden by the use of multipliers of likelihood components. The *final* CVs and effective sample sizes from the iterative re-weighting were not documented.

5. *Review the findings from the retrospective analyses, assess magnitude and direction of retrospective patterns detected, and discuss implications of any observed retrospective pattern for uncertainty in population parameters (e.g., F, SSB), reference points, and/or management measures*.

A retrospective analysis was performed by the assessment author for the base model. There were no retrospective patterns of any consequence, nor were there likely to be. It is a common misconception that retrospective analysis can, of itself, provide useful information on estimator uncertainty or bias. Retrospective patterns, if they are particularly severe, can be an indication of problems with the model that would warrant further investigation. However, retrospective patterns are common for estimators and of themselves do not indicate a problem. It is easy to construct a very poor estimator that will have no retrospective pattern. It is also true that many good estimators will, purely by chance, often show a retrospective pattern.

6. *Recommend stock status as related to reference points.*

The assessment was rejected and therefore no recommendation is made with regard to stock status. However, I do offer qualitative advice on stock status that is consistent with the Panel's advice in the joint report.

The landings peaked in 1984 with a catch of almost 1,000,000 t and in the 1980s there were six consecutive years with landings of over 800,000 t. If these removals had been associated with high fishing mortality they would have caused a contraction in the age structure of the landings during the period of high catches and in subsequent years. There is no strong evidence of this in the catch-at-age data. For example, the proportion of fish 3 years and older in the landings shows little trend from 1980 through to 2010.

The average annual landings from 2000 to 2010 were approximately 480,000 t. This is a substantial reduction in landings from the peak period and suggests that the current stock status is probably "not overfished" and "not overfishing".

The "worst case" ASPIC runs (using an annual 2% increase in efficiency since 1948 for the reduction CPUE indices) suggest that over-fishing is not currently occurring (3 out of the 4 runs have F_{2010} less than F_{MSY}). The runs do allow the possibility that the stock may be over-fished (B less than $0.5B_{MSY}$), but 3 out of the 4 runs have B_{2010} approximately equal to B_{MSY} and the point estimate from the 4th run does not breach the biomass threshold. Given that these are "worst case" scenarios, the runs suggest that the most likely stock status is "not overfished" and "not overfishing".

7. *Develop detailed short and long-term prioritized lists of recommendations for future research, data collection, and assessment methodology. Highlight improvements to be made by next benchmark review.*

Remedial actions necessary to provide a reliable stock assessment

No remedial actions are required with regard to SRA and ASPIC, as such, in that neither can currently provide a reliable assessment. The SRA model needs further development to use catch-at-age data so that selectivities can be internally estimated and that there are reliable data for estimating recruitment patterns. ASPIC is of limited utility for menhaden as it cannot use the catch-at-age data and is not provided with information on the biological characteristics of the species (e.g., cannot accurately estimate F_{MSY}). Also, a reliable ASPIC assessment requires reliable adult abundance indices, which are not currently available.

It should be possible to provide a reliable, although potentially very uncertain, stock assessment using the BAM package. However, preparatory work is required with regard to the current indices and the catch-at-age data; and the BAM runs need to be correctly implemented.

With regard to data inputs:

More work needs to go into producing defensible abundance indices from existing data.

- The CDFR data should be used to construct a standardized CPUE time series (with catch as the response variable and various effort variables included as potential explanatory variables). The resulting time series would be an improvement on the existing reduction CPUE time series but would still need to be interpreted as a potentially hyper-stable index. Some of the early years in the data may need to be omitted if they are from a period when the fishery was changing dramatically.
- The fishery independent data needs to be more fully analyzed, starting with a descriptive analysis of the temporal and spatial distribution of catch rates at stations. After that is done, it may be possible to determine better methods to use the data to construct abundance indices (e.g., some post-stratification and scaling by area may be possible). Efforts should be concentrated on the juvenile indices as the gillnet data can never be corrected for sampling just the edge of the adult distribution.
- Full documentation of proposed indices needs to be provided including standardization diagnostics and estimated effects. For indices derived by combining across states, state-year interactions should be investigated to see if there may be different trends between states.
- Exclude 2010 data from all analyses if their inclusion is problematic because of probable changes in fish distribution or fishing practice due to the BP oil spill.

The reduction fishery catch-at-age data should be checked for possible ageing problems:

- There are very few fish 3 years or older in the catch-at-age matrix from 1966 to 1970 inclusive. This corresponds to a period when multiple readers were used. Scales from the period 1964–1970 should be re-read by the current reader.
- A sample of scales throughout the whole time period should be re-read by the current reader to check for drift in the age readings.

For the BAM modeling:

- Either start the model in 1948 with initial biomass equal to virgin biomass with equilibrium age structure and/or start when the age data begin and estimate initial age structure.
- Only estimate the recruitment deviations during the period when there are data available to support the estimates.
- Do not estimate steepness unless the biomass trajectory covers a broad range of biomass in terms of percent B_0.
- Aim for a parsimonious model in terms of the number of parameters; only introduce additional parameters if they can be more than justified in terms of improved model fit (use AIC or an even more stringent rule of thumb).
- An age-invariant M is probably adequate for this model given the small number of age classes in the catch. It may be possible to estimate M given the extensive catch-at-age data. However, work with a fixed M first until the basic model structure is sorted out. Then do a likelihood profile for M to see if there is any suggestion of information on M.

- Complex selectivity parameterizations are undesirable. Start with a single time invariant selectivity and introduce additional complexity only if necessary (e.g., blocking if supported by known/suspected changes in the fishery).
- The estimated CVs for indices derived from CPUE analysis and/or linear modeling are almost always far too low for stock assessment purposes. One rule of thumb is to set them at a starting value of 20% in each year.
- Continue to use number of tows for effective sample sizes of composition data (with a cap of 100 or 200).
- Always check the SDNRs to make sure they are approximately equal to 1 and, if not, use iterative re-weighting. After re-weighting check that there is still an adequate fit to "reliable" abundance indices (if not, up-weight the abundance indices). Fully document the final CVs and effective sample sizes.
- Alternatively, or as a sensitivity test, follow the recommendation of Francis (2011) for weighting of composition data; this may give even lower effective sample sizes (than the capped number of tows).
- Do not use multipliers of the likelihood functions except as a quick way to test the sensitivity of results to emphasizing alternative data sets.
- Fully document the results of sensitivity runs including tables/figures with the estimates of all parameters for easy comparison across runs.
- Key outputs that are useful for understanding how a model is behaving are the biomass/egg-production trajectory in absolute terms and as a percentage of virgin biomass/egg-production. Effects can be very hard to understand when only B/B_{MSY} ratios are plotted. The trajectory of the absolute value of F is also useful.
- Include a run that does not use any of the supposed adult abundance indices. This should be the base run until a defensible adult-abundance time series can be developed.
- As a reality check it may be worthwhile doing an un-tuned VPA using a range of terminal Fs (with low and high Ms in addition to the best estimate).

Research Recommendations

Short-term (prioritized):

Adult abundance index: Review methods that could be used to provide a reliable fishery-independent adult-abundance time series. A pilot survey should be implemented as soon as possible. Development of a long-term time series is needed to increase the certainty of menhaden stock assessments.

Analysis of CDFR data: These data may contain an abundance signal on a weekly and/or an annual basis. In the long-term the data should be fully analyzed in this regard. In the short-term, a standardized CPUE time series should be developed from the data for use in stock assessment.

Further analysis of fishery independent state indices: These data need to be fully analyzed with regard to determining the best methods to use the data to provide potential juvenile (and perhaps adult) abundance indices.

Ageing: The consistency of the age readings throughout the whole time series should be checked. The current reader has read scales since 1969 and there may be some drift in her readings. Also, other readers participated up to the early 1970s and there is evidence of relative bias in the readings up to 1970 that should be investigated.

Further development of the SRA: The incorporation of catch-at-age data into the SRA approach is encouraged as this would allow the method to provide an alternative stand-alone stock assessment for menhaden.

Long-term (prioritized):

Adult abundance survey: The existing state sampling of coastal waters is not adequate for providing a defensible adult abundance index. In the absence of such an index, stock assessment of menhaden will continue to be problematic. The development of a fishery-independent adult-abundance index should be given a very high priority. A review of possible methods is the first step (see Short-term recommendations). Aerial surveying using visual estimation and/or LIDAR should be considered among the options.

Biological data: All biological parameters pertinent to the stock assessment should be updated. Subsequently, they should be monitored every few years.

Catch sampling: The potential bias associated with sampling only the last catch of the day should be investigated. It is important to know if there could be a bias and whether it is towards larger/older fish or smaller/younger fish.

Critique of the NMFS review process

The ToRs for this review were originally identical to the ToRs for the Assessment and Data Workshops. I expressed concern about the ToRs prior to the review. There was one change made to them before the meeting in that we were not required to "perform retrospective analyses". Obviously, ToRs for a review need to be written differently than those for a data or assessment workshop. I understand that new generic ToRs for SEDAR reviews are currently being written and I support that initiative.

The basic structure of data-workshop, assessment-workshop, and then review is sound. The nature of the review being essentially acceptance or rejection of an existing assessment with limited room to change the assessment is also sound. It is important that the assessment is not modified by a review panel to the extent that it becomes a new assessment and has therefore not undergone review. The inclusion of experts with knowledge of the local jurisdiction in the review panel is also highly desirable.

In some ways this review was an endorsement of the process in that a sub-standard assessment was reviewed and rejected. As a Panel, we have been able to provide recommendations on how a reliable assessment can produced. However, most of the

suggestions with regard to the modeling are simply for use of best/good practice. It is a concern that good practice was not followed at the assessment workshop. This was not a failure of the (inexperienced) analyst. The process should have provided them with sufficient support and technical review during the data and assessment workshops to ensure that good practice was followed.

I think that in some cases it would be useful for independent/outside technical experts to participate in the data and/or assessment workshops to provide additional peer review during the development of the assessments.

Conclusions and Recommendations

The menhaden assessment was not technically defensible and stock status recommendations using the results from the base model and supporting models were not able to be made. However, a qualitative assessment of the data by the Panel, in conjunction with some "worst case" model runs, did allow the Panel to tentatively suggest that the stock is probably not over-fished and that over-fishing is probably not occurring.

Remedial actions to provide a reliable quantitative stock assessment are detailed in this report. Work is required on the data inputs to (perhaps) derive more defensible adult abundance indices. And, good practice needs to be adopted while developing a base model and sensitivities within the BAM framework.

The main recommendation, to improve the accuracy and precision of future menhaden stock assessments, is to develop a reliable fishery-independent adult-abundance index (see detailed recommendations under ToR 7). This cannot be done by surveying only coastal waters and will require a survey covering most of the spatial extent of the adult population.

References

Francis, R.I.C.C. 2011. Data weighting in statistical fisheries stock assessment models. *Can. J. Fish. Aquat. Sci. 68*: 1124–1138.

Appendix 1: Bibliography of supplied material

Document #	Title	Authors
Documents Prepared for the Data Workshop		
SEDAR27-DW-01	History of Assessments of the Menhaden Stock along the U.S. Gulf of Mexico Coast	Douglas S. Vaughan
SEDAR27-DW-02	Age, Growth and Reproduction of Gulf Menhaden	Douglas S. Vaughan, Joseph W. Smith and Amy M. Schueller
SEDAR27-DW-03	Life History-Based Estimates of Natural Mortality for Gulf Menhaden	Amy M. Schueller
SEDAR27-DW-04	History of the Gulf Menhaden Fishery and Reconstruction of Historical Commercial Landings	Joseph W. Smith and Douglas S. Vaughan
SEDAR27-DW-05	Harvest, Effort, and Catch-at-Age for Gulf Menhaden	Joseph W. Smith and Douglas S. Vaughan
SEDAR27-DW-06	Management Unit Definition for the Gulf Menhaden Stock in the U.S. Gulf of Mexico	Steve VanderKooy
SEDAR27-DW-07	Habitat Description for the Gulf Menhaden Stock in the U.S. Gulf of Mexico	Steve VanderKooy
SEDAR27-DW-08	Regulatory History for the Gulf Menhaden Stock in the U.S. Gulf of Mexico	Steve VanderKooy
SEDAR27-DW-09	Report on the distribution and abundance of menhaden (*Brevoortia spp*.) larvae captured in ichthyoplankton samples during fishery-independent resource surveys in the Gulf of Mexico	Joanne Lyczkowski-Shultz and David S. Hanisko
Documents Prepared for the Assessment Workshop		
SEDAR27-AW-01	Surplus production models of gulf menhaden, *Brevoortia patronus*	Michael H. Prager and Douglas S. Vaughan
Documents Prepared for the Review Workshop		
SEDAR27-RW-01	The Beaufort Assessment Model (BAM) with application to gulf menhaden: mathematical description, implementation details,	NOAA Beaufort Laboratory

	and computer code	
Final Stock Assessment Reports		
SEDAR27-SAR1	Gulf of Mexico Menhaden	
Reference Documents		
SEDAR27-RD01	Fishery Independent Sampling: Alabama	
SEDAR27-RD02	Fishery Independent Sampling: Mississippi	
SEDAR27-RD03	Fishery Independent Sampling: Florida	
SEDAR27-RD04	Fishery Independent Sampling: Texas	
SEDAR27-RD05	Fishery Independent Sampling: SEAMAP Trawl	
SEDAR27-RD06	Fishery Independent Sampling: Louisiana	
SEDAR27-RD07	Sampling Statistics in the Atlantic Menhaden Fishery	Alex Chester
SEDAR27-RD08	Gulf menhaden (*Brevoortia patronus*) in the U.S. Gulf of Mexico: Fishery characteristics and biological reference points for management	Douglas S. Vaughan, Kyle W. Shertzer, Joseph W. Smith
SEDAR27-RD09	Red snapper: Iterative re-weighting of data components in the Beaufort Assessment Model (SEDAR 24-RW-03)	NOAA Beaufort Laboratory

Appendix 2: Statement of Work for Patrick Cordue

External Independent Peer Review by the Center for Independent Experts

SEDAR 27 Gulf of Mexico Menhaden and Southeast Yellowtail Snapper Review

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.org.

Project Description: SEDAR 27 will be an assessment review for conducted for Gulf of Mexico Menhaden and Southeast Yellowtail Snapper. 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 stocks assessed through SEDAR 27 are within the jurisdiction of the Gulf of Mexico and South Atlantic Fisheries Management Councils and the states of Texas, Louisiana, Mississippi, Alabama, Florida, Georgia, South Carolina, and North Carolina. 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: Three 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 stock assessment, statistics, fisheries science, and marine biology sufficient to complete the primary task of reviewing the technical details of the methods used for the assessment. Expertise with data poor assessment methods would be preferable. 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 scheduled in Saint Petersburg, Florida during November 1-4, 2011.

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.

Contract Deliverables - Independent CIE Peer Review Reports: 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.

Other Tasks – Contribution to Summary Report: 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 in the panel review meeting in Saint Petersburg, Florida during November 1-4, 2011.
3) In Saint Petersburg, Florida during November 1-4, 2011 as specified herein, conduct an independent peer review in accordance with the ToRs (**Annex 2**).
4) No later than November 18, 2011, each CIE reviewer shall submit an independent peer review report addressed to the "Center for Independent Experts," and sent to Manoj Shivlani, CIE Lead Coordinator, via email to shivlanim@bellsouth.net, and CIE Regional Coordinator, via email to Dr. David Sampson david.sampson@oregonstate.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.

September 27, 2011	CIE sends reviewer contact information to the COTR, who then sends this to the NMFS Project Contact
October 18, 2011	NMFS Project Contact sends the CIE Reviewers the pre-review documents
November 1-4, 2011	Each reviewer participates and conducts an independent peer review during the panel review meeting
November 18, 2011	CIE reviewers submit draft CIE independent peer review reports to the CIE Lead Coordinator and CIE Regional Coordinator
December 2, 2011	CIE submits CIE independent peer review reports to the COTR
December 9, 2012	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.

Support Personnel:

William Michaels, Program Manager, 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

Roger W. Peretti, Executive Vice President
Northern Taiga Ventures, Inc. (NTVI)
22375 Broderick Drive, Suite 215, Sterling, VA 20166
RPerretti@ntvifederal.com Phone: 571-223-7717

Key Personnel:

Julie A Neer, SEDAR Coordinator
4055 Faber Place Drive, Suite 201
North Charleston, SC 29405
julie.neer@safmc.net Phone: 843-571-4366

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

SEDAR 27 Gulf of Mexico Menhaden and Southeast Yellowtail Snapper Review

1. Evaluate precision and accuracy of fishery-dependent and fishery-independent data used in the assessment:
    a. Discuss data strengths and weaknesses (e.g. temporal and spatial scale, gear selectivities, aging accuracy, sampling intensity).
    b. Report metrics of precision for data inputs and use them to inform the model as appropriate.
    c. Describe and justify index standardization methods.
    d. Justify weighting or elimination of available data sources.
2. Evaluate models used to estimate population parameters (e.g., F, biomass, abundance) and biological reference points.
    a. Did the model have difficulty finding a stable solution?
    b. Were sensitivity analyses for starting parameter values, priors, etc. and other model diagnostics performed?
    c. Have the model strengths and limitations been clearly and thoroughly explained?
    d. Have the models been used in other peer reviewed assessments? If not, has new model code been verified with simulated data?
    e. Compare and discuss differences among alternative models.
3. State and evaluate assumptions made for all models and explain the likely effects of assumption violations on model outputs, including:
    a. Calculation of M.
    b. Choice of selectivity patterns.
    c. Error in the catch-at-age matrix.
    d. Choice of a plus group for age-structured species.
    e. Constant or variable ecosystem (e.g., abiotic) conditions.
    f. Choice of stock-recruitment function.
    g. Choice of reference points (e.g. equilibrium assumptions).
4. Evaluate uncertainty of model estimates and biological or empirical reference points.
    a. Choice of weighting likelihood components.
5. Perform retrospective analyses, assess magnitude and direction of retrospective patterns detected, and discuss implications of any observed retrospective pattern for uncertainty in population parameters (e.g., F, SSB), reference points, and/or management measures.
6. Recommend stock status as related to reference points.
7. Develop detailed short and long-term prioritized lists of recommendations for future research, data collection, and assessment methodology. Highlight improvements to be made by next benchmark review.

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.

Annex 3: Agenda

SEDAR 27 Gulf of Mexico Menhaden and Southeast Yellowtail Snapper Review

Florida Fish and Wildlife Research Institute
Saint Petersburg, Florida
November 1-4, 2011

Tuesday

9:00 a.m.	**Convene**	
9:00 – 9:30 a.m.	**Introductions and Opening Remarks** **Coordinator** *- Agenda Review, TOR, Task Assignments*	
9:30 a.m. – 11:30 a.m.	**Assessment Presentation**	**TBD**
11:30 a.m. – 1:30 p.m.	**Lunch Break**	
1:30 p.m. – 3:30 p.m.	**Continue Assessment Presentations**	**TBD**
3:30 p.m. – 4:00 p.m.	**Break**	
4:00 p.m. – 6:00 p.m.	**Panel Discussion** **-** *Assessment Data & Methods* *- Identify additional analyses, sensitivities, corrections*	**Chair**

***Tuesday Goals*:** Initial presentations completed, sensitivities and modifications identified.

Wednesday

8:30 a.m. – 11:30 a.m.	**Panel Discussion** *- Review additional analyses, sensitivities* *- Consensus recommendations and comments*	**Chair**
11:30 a.m. – 1:30 p.m.	**Lunch Break**	
1:30 p.m. – 3:30 p.m.	**Panel Discussion**	**TBD**
3:30 p.m. – 4:00 p.m.	**Break**	
4:00 p.m. – 6:00 p.m.	**Panel Discussion**	**Chair**

Wednesday Goals: Final sensitivities identified, Preferred models selected, Projection approaches approved, Summary report drafts begun

Thursday

8:30 a.m. – 11:30 a.m.	**Panel Discussion** *- Final sensitivities reviewed.* *- Projections reviewed.*	**Chair**
11:30 a.m. – 1:30 p.m.	**Lunch Break**	
1:30 p.m. – 3:30 p.m.	**Panel Discussion or Work Session**	**Chair**
3:30 p.m. - 4:00 p.m.	**Break**	
4:00 p.m. - 6:00 p.m.	**Panel Work Session** *- Review Consensus Reports*	**Chair**

Thursday Goals: Complete assessment work and discussions. Final results available. Draft Summary Report reviewed.

Friday

8:30 a.m. – 12:00 p.m.	**Panel Work Session**	**Chair**
12:00 p.m.	**ADJOURN**	

Appendix 3: Review workshop participant list

Workshop Panel

Luiz Barbieri, Chair ..... FWRI
John Wheeler ..... CIE Reviewer
Patrick Cordue ..... CIE Reviewer
Sven Kupschus ..... CIE Reviewer
Will Patterson ..... GSMFC-appointed Reviewer

Analytic Representation

Amy Schueller ..... NMFS SEFSC Beaufort
Bezhad Mahmoudi ..... FWRI
Mike Prager ..... Prager Consulting

Rapporteur

Wade Cooper ..... FWRI

Observers

Doug Vaughan ..... GSMFC observer
Ron Lukens ..... Omega Protein
Lew Coggins ..... NMFS SEFSC Beaufort

Staff

Julie Neer ..... SEDAR
Rachael Silvas ..... SEDAR
Steve VanderKooy ..... GSMFC