Center for Independent Experts (CIE) independent peer review report for the 2014 benchmark stock assessment of Atlantic Menhaden

Anders Nielsen

January 2015

Executive Summary

The assessment team presented a very complete and comprehensive assessment report prior to this meeting, and were able to answer any questions posed by the panel in a very efficient manner. All details were well documented and made available to the review panel - even the source code of the assessment model. It was clear that this assessment represented a lot of work and a major step forward, compared to the previous benchmark assessment, in terms of available data and modeling approach.

The fishery is well monitored and the data series are long, consistent, reliable, and sufficient to inform the assessment model. The decisions about biological parameters are sound and robust, and consistent with normal practice.

The length compositions from the two new adult indices were not in agreement with the predicted length compositions from the assessment model. The two new adult indices are constructed by adding all length samples across a number of individual surveys (unweighted). This approach may be causing the mismatch, as the different individual surveys are conducted in different areas and at different times. It was pragmatically chosen to down weight the length compositions (by a factor of 10) and use that as the new base run, to limit their influence. Suggestions for more long term solutions were made.

Uncertainties were investigated via a parametric Monte Carlo Bootstrap approach, (where the data and fixed inputs were sampled using relevant uncertainties), a wide range of sensitivity analysis, and retrospective analysis.

Across the uncertainties, sensitivity analysis, old base run, and the new base run (where the results were presented at the meeting) the main conclusion remained: Atlantic Menhaden is not overfished and is not undergoing overfishing.

Finally options for reference points were reviewed. Without an informative stock-recruitment relationship it seems reasonable to base the reference points on quantiles from the historic period. It was further suggested to use a wider range of ages as basis for the fishing mortality reference points.

Looking beyond the single species reference points the modeling options were reviewed, and a minimally complex combination of the current assessment model and multi-species production models was suggested.

Background

This report is prepared for the Center for Independent Experts (CIE). It contains an independent and impartial review of the assessments of Atlantic menhaden. The SEDAR 40 Assessment Review was held in December 9-11, 2014 in Atlantic Beach, NC.

The assessment scientists presenting were: Amy Schueller (Lead analyst), Genny Nesslage, Jason McNamee, and Joe Smith.

Prior to the meeting the review panel were given a link to an ftp site with background documents (appendix 1). This reviewer's statement of work can be found in appendix 2, and a list of review meeting participants in appendix 3.

Description of the reviewer's role

This reviewer has independently read the assessment report, its appendices and all supplementary documents deemed necessary in preparation for the review, traveled and participated actively in the review meeting, identified key issues in the assessment, contributed to the review panel's summary report, and independently authored this review report.

Findings for each term of reference

To ensure that all terms of reference are covered, and that comments are interpreted with reference to the correct terms, the terms are listed with corresponding reviewer comments following.

> Evaluate the data used in the assessment.
>
> Are data decisions made during the DW and AW justified (i.e. sound and robust)?
>
> Are input data series reliable and sufficient to support the assessment approach and findings?
>
> Are data applied properly within the assessment?
>
> Are data uncertainties acknowledged, reported, and within normal or expected levels?

The assessment report provides a thorough and concise description of the different data sources used in the assessment. The assessment team had made a big effort to improve the available data, both in terms of quality checking and in making new fisheries independent surveys available. In terms of available data this assessment represents a big improvement.

The two important commercial fisheries are the reduction fleet and the bait fleet. In the beginning of the data series the bait fleet was only a small part of the fishery, but in recent years it has increased to about 23%. Both of the fleets are split into a northern and a southern part to allow for different selectivity in the two areas. Age compositions are available for all four fleets.

The fishery from two reduction fleets (north and south) appears well monitored. It is covered by logbooks, the age sampling is consistent and sufficient to split the catch into age groups, and the so-called "Hopper" device used to measure the

landings has been thoroughly checked for accuracy. The assessment team stated that the age reading from scales is not considered problematic for Atlantic menhaden, but that no good comparisons to age reading from otoliths have been conducted.

The samples used for age and length measurements are taken from the top of the tank. Some fishing trips will take most of their catch far from the landing site, but then "top off" the catch closer to the landing site. This practice can potentially bias the results from the samples, but the logbooks are used to compensate for this effect by assigning samples to correct areas.

The two bait fleets (north and south) are monitored less. There are concerns that species are lumped in reporting in some areas and that some are fished for personal bait or private sales (and hence not reported). Not all gears catching bait were included in the bait landings prior to 1985, so the data quality has likely improved since 1985. The age sampling from the bait fishery is more limited (compared to the reduction fishery), but it has increased over time as the bait fishery has become a larger part of the total catches. The recreational catches are estimated to be less than one percent of the total catches, so overall of little importance. The bait catches are slightly inflated to somewhat account for the recreational catches.

Construction of indices of abundance has been a major undertaking by the assessment team. The Atlantic menhaden assessment used to rely on only one fisheries dependent index based on Potomac River pond net fishery and one fishery independent index based on state seine surveys. As part of this assessment the assessment team have scanned and evaluated 14 fishery dependent and 45 fishery independent data sets to identify which were suitable to be included. The selection process was based on clearly stated sensible criteria. The selected time series were standardized w.r.t. variables affecting catchability. The resulting separate indices were then combined, using Conn's method (Conn 2010), into three composite indices: a Young Of Year index (YOY), a Northern ADult index (NAD), and a Southern ADult index (SAD). The review panel used considerable time investigating the methods of standardization and combining (see the panels summary report). Using information from all these "new data" is overall a big improvement.

The length compositions for each of the adult indices (SAD and NAD) were constructed by adding all length samples across surveys (unweighted). This approach may be problematic, as the different individual surveys are conducted in different areas and at different times, which may have different length distributions. Adding these length distributions may result in a combined length distribution which is not representative of the stock anywhere.

A couple of additional issues were raised during the review meeting w.r.t. scanning of relevant indices to include: 1) Simple correlation calculation is sensitive to a single extreme point, so when correlation is part of the selection procedure it is advisable to check for the influence of single points. 2) In some cases the standardization model did not converge, which was part of the team’s selection

procedure. This is a somewhat artificial criterion, as model convergence is influenced by arbitrary initial values, and details of the implemented software. Only one index was excluded because of this criterion, and it was a borderline case for other reasons, but this criterion could also influence which explanatory variables are selected for the standardization. Overall the effects are judged to be at most minor. The criterion is not judged to cause a systematic bias.

For natural mortality it was chosen to input age specific, but time invariant numbers based on a Lorenzen curve scaled to mortality estimate from tagging. Strong consideration was given to using time varying M estimated from a MSVPA run, but the MSVPA run had other limitations and only spanned the later part of the data period (from 1982).

Overall the data decisions leading up to this assessment are justified, and the data series are reliable and sufficient to support the assessment approach and findings. The only slightly problematic part is the combined length compositions.

More details about the model approach will follow later (TOR 2 and 3), but in summary: the data are applied properly within the assessment, and the uncertainties acknowledged, reported, and within normal or expected levels.

> Evaluate the methods used to assess the stock, taking into account available data.
>
> Are methods scientifically sound and robust?
>
> Are assessment models configured properly and used consistent with standard practices?
>
> Are the methods appropriate for the available data?
>
> If multiple models or model configurations were considered, evaluate the explanation of any differences in results and justification of a base model.

The stock assessment model used for Atlantic Menhaden is the Beaufort Assessment Model (BAM). BAM is a statistical catch at age model, which is a model type used for most statistical fish stock assessments worldwide. Other commonly applied statistical catch at age models are SCAA and SS3. BAM has previously been used in SEDAR assessments (e.g., Spanish mackerel, Gulf Menhaden, and red grouper). The version of BAM was set up to match the data requirements of Atlantic Menhaden. The assessment team clearly demonstrated that they were comfortable modifying both configuration and source code, and hence were not treating it as a `black box'. BAM for Atlantic Menhaden is thoroughly documented both in mathematical terms and by openly sharing its source code (appendix C), which allows for review at the most detailed level if desired. All of the above strengthens confidence in the model being scientifically sound, robust, and appropriate for the available data.

The predicted removals from the four fleets closely match the observed (fig. 7.1.1-4), which is expected, as the model assumes a fixed low CV. The predicted age compositions capture all main features in the observed age compositions (fig. 7.1.5-8). A minor indication of a shift in selectivity around the year 2003 is seen for the

northern bait fleet (fig. 7.1.11). Predicted abundance indices for the three combined surveys (YOY, NAD, and SAD) are in agreement with the observed (7.1.13-15). The model is not able to predict the length compositions of the two adult combined surveys (7.1.16-17). The length compositions are mainly used to inform the age specific selectivity used to predict the abundance indices from the two adult surveys, this age specific selectivity may be approximated even when the detailed length compositions are not matched. The different components of the likelihood were weighted according to Francis (2011).

The overall conclusion was that BAM was configured properly and used consistent with standard practices.

The problematic mismatch between model predicted and observed length compositions was further investigated by the review panel and assessment team. The concern was that the misfit was biasing other estimated quantities from the model (e.g. stock sizes). A requested sensitivity run where the CV around the growth function was set to half of its estimated value showed that stock sizes were influenced. To minimize the problem it was chosen to down weight the length composition by a factor of 10. A configuration where the length compositions were completely removed, but where the estimated selectivity (from the length data) were input as fixed was also done, which gave similar results as down weighting by a factor of 10. Down weighting by a factor of 20 was also attempted, but resulted in failed convergence.

Down weighting by a factor of 10 was pragmatically chosen as the new base run.

It is recommended that future versions of the model do not use the length composition data, or find better ways to predict them. One possibility is to use the individual surveys (instead of the combined). Using the individual surveys would allow individual selection curves, which would better reflect that these surveys are coming from different areas and times.

Using the surveys individually within the assessment model, instead of first using an external model to combine the selected surveys and then using the combined survey in the assessment model, is one example of so-called integrated modeling. The advantage of integrated modeling is that nothing gets thrown away by the link between the two models. Consider the survey example: The model standardizing and combining the surveys delivers an index and associated standard deviations, which is then used in the assessment model. From this point, the data are reduced to that index, and all model diagnostics of the assessment model is using only that index. If the original data e.g. gave correlated indices, then that information is lost. If the original data were used within the model, then the correlations would not be ignored, and would be propagated to the estimated quantities of interest. Also the model diagnostics would be allowed to be evaluated against real observations. Another important example in this assessment is the growth model, which is done external to the model. It could be considered to move growth to a more integrated approach in future generations of this assessment.

The model uses dome-shaped selectivity for all fleets, except for the NAD fleet. Dome-shaped selectivity pattern can be caused by many mechanisms (e.g. spatial distribution). It is seen that the predicted catch-at-age compositions match the observed catch-at-age compositions. There are several possible reasons to expect dome-shaped selectivity: 1) The larger fish could be further from the shore than where the main part of the fishery is being conducted. 2) The reduction fleet uses spotter planes to locate fish schools, so they may bypass smaller schools and choose to fish on larger schools. If larger fish tend to swim in smaller schools, and smaller fish in larger schools, then this practice till lead to a dome-shaped selectivity pattern. 3) Remaining bias from the "topping off" practice. If age samples are taken from the from the top of the tank (caught close to shore), but the majority of the catch is caught further away, that could lead to a bias, but should already be corrected for.

The model diagnostic plots can be improved. Minor modifications to the residual plots could make it easier to evaluate if the assessment model is a satisfactory description of the observed data, e.g: 1) When a log-normal distribution is assumed it is useful to show residuals and observed versus predicted values on a logarithmic scale. 2) Showing standardized residuals is helpful for evaluating the assumed variance structure.

The assessment model uses a single CV for length-at-age. Given the problems of matching the length composition data it may be useful to investigate if that model is flexible enough.

The assessment report states the model stock synthesis was also configured for the stock, but the model was not presented at the review meeting. The assessment team had determined that seasonality had to be included in stock synthesis first.

> Consider how uncertainties in the assessment, and their potential consequences, are addressed.
>
> Comment on the degree to which methods used to evaluate uncertainty reflect and capture the significant sources of uncertainty in the population, data sources, and assessment methods.
>
> Are the implications of uncertainty on technical conclusions are clearly stated?

The assessment team have put a lot of effort into investigating the uncertainties. Minimum common practice would have been to supply uncertainties derived from the inverse Hessian matrix of the objective function at its minimum. This is a standard output from most model fitting software, but it would not have been valid here for two reasons. First of all many quantities of importance (e.g. natural mortality) are entered as known constants, even if knowledge about them is uncertain. Secondly assigning arbitrary weights to likelihood components and deviance variances also affects the Hessian derived uncertainties.

Instead the assessment team used a parametric Monte Carlo Bootstrap (MCB) method, where the data and fixed inputs (including natural mortality) were sampled

using relevant uncertainties. For each of 1000 complete sampled sets of input the model was re-estimated, which results in a simulated distribution of all estimated quantities. This approach correctly propagates the uncertainty through the nonlinear model equations to the quantities of interest.

Two minor concerns about the implementation details of the approach were raised. 1) The fixed assigned weights of the likelihood components likelihood were kept fixed and not part of the sampling. 2) all quantities were sampled independently. For instance the two parameters of a logistic function were each simulated uniformly from their 95% interval. If these model parameters are correlated, then sampling them independently will result in unlikely pairs (and hence unlikely logistic curves). It was suggested to use the joint distribution where available.

In addition to the MCB method the assessment team had prepared a wide range of sensitivity runs. These included: leaving out entire data sources, including ageing uncertainties, changing index calculation method, different assumptions about natural mortality, different weighting of likelihood components, and different time varying assumptions. The results were seen to be robust to most alternatives, and to react as expected to others (fig. 7.4.1.1-77). The results were most sensitive to changes in assumed natural mortalities and omitting the NAD index.

Finally a retrospective analysis was presented, where the last 1, 2, 3, or 4 years of data were left out to demonstrate that the final year's estimates are not severely biased. For the estimates of fishing mortality and biomass no systematic retrospective bias is seen. For the recruitment estimates a negative bias is seen, but it is small and only supported by four points.

Overall the method used to evaluate uncertainty reflect and capture the significant sources of uncertainty in the population, data sources, and assessment methods, and the implications of uncertainty on technical conclusions are clearly stated (fig. 8.3.2.1-12).

> Evaluate the assessment findings with respect to the following:
> Are estimates of biomass, abundance, and exploitation rate reliable and consistent with input data and population biological characteristics? Are they useful to support inferences on stock status?

The conclusion is that the estimates of biomass, abundance and exploitation rate are reliable and consistent with input data and biological characteristics. They are also useful to support inferences on stock status. The only real exception is the inconsistency between the model predicted length compositions and the observed combined length compositions (described above). The review panel investigated the consequences of that and recommended a modified base run that down weighted this problematic part of the data to minimize its influence.

The reliability of the estimates are investigated via sensitivity runs and via the Monte Carlo Bootstrap. The biomass estimates are (not surprisingly) sensitive to the assumed natural mortality. The level changes, but the major trends over time

remains the same, even when the time varying natural mortalities from the multi-species (MSVPA) are used. The biomass is however also sensitive to leaving out the NAD index, which is potentially more problematic, as the NAD index uses the selectivity, which is informed by the inconsistent length compositions. It should however be noted that only the much coarser selection-at-age is used, which may still be a fair approximation. Recruitment and fecundity are similarly sensitive to natural mortality and the NAD index. In years prior to 1960 the estimates of fishing mortality are unrealistically high (F(age 3)>8.0 in one year), so in this initial phase all estimates from the model are likely less reliable.

> Is the stock overfished relative to biomass or abundance threshold reference points? Where is the stock relative to biomass or abundance management targets? What information supports this conclusion?

The stock is not overfished, as it is above its abundance threshold and target reference points. This is supported by the base run in the assessment report and the corresponding Monte Carlo bootstrap analysis (e.g. fig. 8.3.2.5-6) and a wide range of sensitivity analysis (Fig. 7.4.1.64-7.4.1.77). The base run was modified during the review meeting (see TOR 2), but results presented during the meeting did not modify this conclusion.

> Is the stock undergoing overfishing relative to fishing mortality threshold reference points? Where is the stock relative to fishing mortality management targets? What information supports this conclusion?

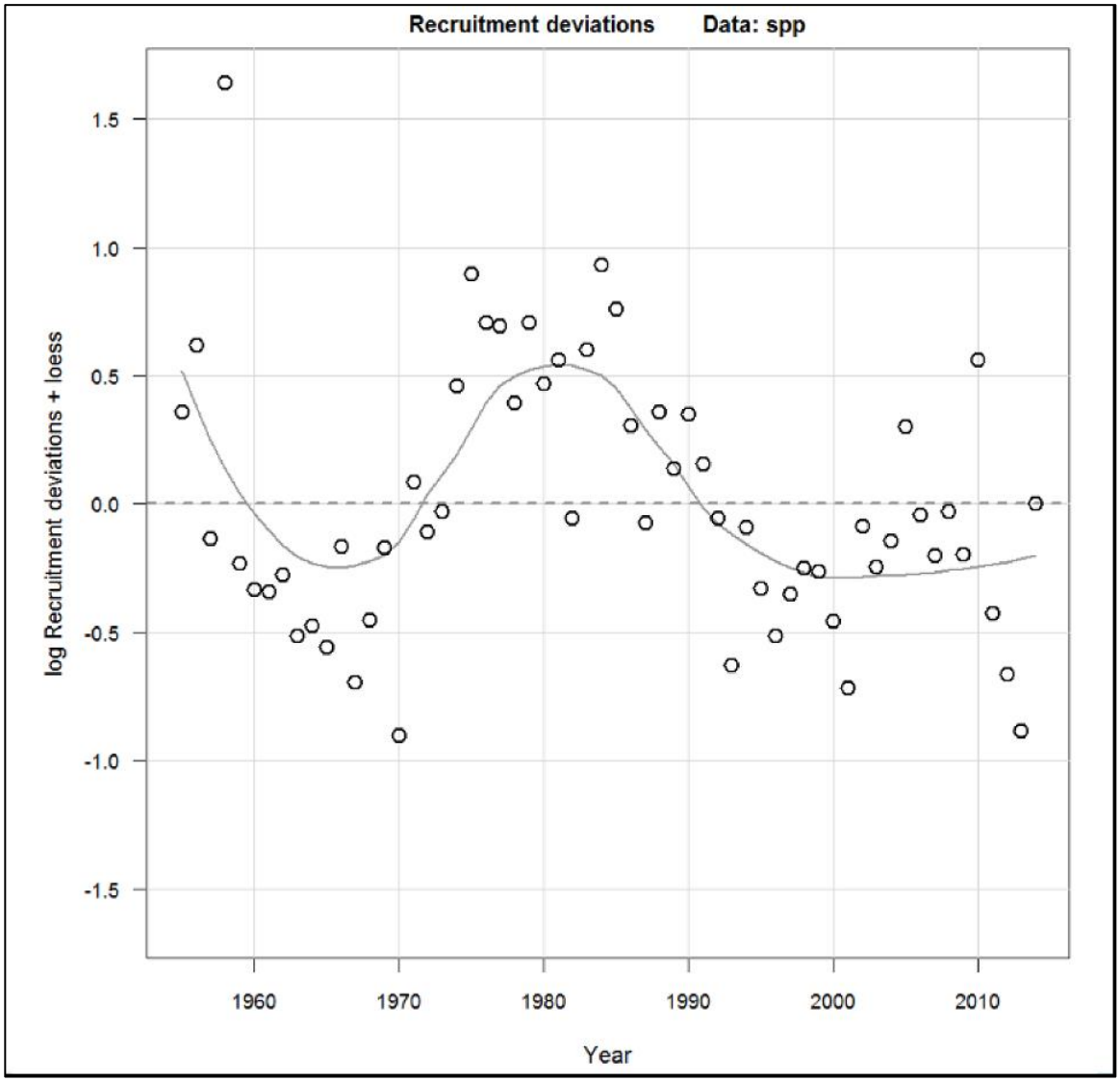
The stock is not undergoing overfishing, as the level of fishing is estimated to be below the fishing mortality threshold and target reference points. This is supported by the base run in the assessment report and the corresponding Monte Carlo bootstrap analysis (e.g. fig. 8.3.2.3-4) and a wide range of sensitivity analysis (Fig. 7.4.1.50-63). The base run was modified during the review meeting (see TOR 2), but results presented during the meeting did not modify this conclusion.

> Is there an informative stock recruitment relationship? Is the stock recruitment curve reliable and useful for evaluation of productivity and future stock conditions?

There is no informative stock-recruitment relationship for this stock. The assessment model assumes a Beverton-Holt curve, but the steepness is fixed at 0.99. Penalized deviances from the curve are allowed. This effectively gives a model with a constant median recruitment at all stock sizes (within the range observed).

An interesting pattern is seen in figure 7.2.3.6, which shows the (log) recruitment deviances plotted against time. The deviances are not independent (as the model currently assume), but shows clear periods of high or low recruitment. This is an indication that something, which is currently not part of the model, is influencing the recruitment. This could for instance be an environmental indicator or the abundance of some other stock. Identifying this relationship could be helpful for evaluation of productivity and future stock conditions. It is also possible that a more

informative stock-recruitment relationship could be seen after correcting for the influencing factor(s). Identifying such an influencing factor is not easy, and care must be taken to ensure that an added explanatory variable is improving the prediction power of the model and not merely over fitting the historic period (see comment on prediction power under TOR 8).



Log recruitment deviations against time

> Are the quantitative estimates of the threshold reference points reliable for this stock? If not, are there other indicators that may be used to inform managers about stock trends and conditions?

The technical committee recommends to "adopt SPR reference points based on the maximum F value experienced at age-2 during the 1960-2012 time period as the threshold and the median F value experienced at age-2 during the 1960-2012 time period as the target, along with the associated FEC values".

Without an informative stock-recruitment relationship it seems reasonable to base the reference points on quantiles from the historic period. As the realized selectivity may be changing over time (e.g. because the bait fishery has become a larger fraction of the total fishery) it would be preferable to base the fishing mortality definition of a wider range of ages (e.g. from age 2 to age 4) instead of only age 2.

Looking beyond the single species setting the reference points should include considerations on Atlantic menhaden's role as a forage fish for other important species (see TOR 8).

> If a minority report has been filed, review minority opinion and any associated analyses. If possible, make recommendation on current or future use of alternative assessment approach presented in minority report.

No minority report was filed.

> Review the Technical Committee's recommendations on research, data collection, and assessment methodology and make any additional recommendations or prioritizations, if warranted.

This reviewer generally agrees with technical committee's recommendations, but will offer some comments here. Additional recommendations will be offered under TOR 7.

The short term recommendation to "continue current level of sampling from bait fisheries ..." is important, as the bait fishery has become relatively more important in recent years.

All actions aimed at validating and improving ageing and fecundity data used in the model is important, as these measures directly affect the result of the assessment and the reference points.

Placing observers on board to collect at-sea samples will be a good way to investigate if the practice of "topping off" the catch close to the landing site is in any way biasing the collected data.

The suggestions to investigate relationship between fish size and school size, and fish size and distance from shore are important to understand and explain the dome-shaped selectivity pattern observed.

The technical committee's highest recommendations w.r.t. data collection is to develop a fishery-independent adult abundance at age. This reviewer fully supports this effort. Having an age specific abundance at age index will remove the need use the length composition data to infer selectivity at age (via an assumed growth function), which has been a challenging part of this assessment.

To this reviewer conducting a Management Strategy Evaluation may be interesting, but at present difficult to do in a meaningful way, because of the lacking informative stock recruitment relationship for this stock. It is mentioned under TOR 4d, but also

a recommendation, that one hope to improve the stock recruitment relationship is to investigate if any environmental covariates are influencing recruitment.

The assessment team was comfortable modifying and adapting their current assessment platform (BAM), so it seems they would be able to extend it in any way needed. Moving to a different platform would seem like a low priority to this reviewer. Except that it is often useful to run two models to verify implementations, and study implications of different assumptions.

Developing seasonally fully spatially-explicit models with movement rates would be interesting, but is also a major task. It is therefore correctly listed in the "long term" category. It is important to make sure that data are available to support such a model before investing a lot of man power developing it. Also simpler alternatives are often available. For instance focusing on the spatial differences in key biological parameters may be enough.

> Provide guidance on key improvements in data or modeling approaches which should be considered when scheduling the next assessment.

There is some overlap between TOR 6 and TOR 7. Here focus will be on suggestions not mentioned by the Technical Committee.

Addressing the problematic fit to the length compositions from the SAD and the NAD indices would be this reviewer's highest priority, and of those the NAD is the most influential. This could be addressed in several ways. 1) Each combined adult abundance index is combined using Conn's method (Conn 2010) on a set of carefully selected individual indices, but the combined length compositions are constructed simply by adding all length samples across surveys (unweighted). This approach may be too simple and resulting in a length distribution which is a mix of length distributions, and hence not representative of the stock anywhere. It could be investigated if an approach is available (or could be developed) which would use the data to predict the length distribution at a relevant place and time of year (e.g. the center of the survey area and the mean time of the survey). 2) It could be considered to simply avoid using the length composition data, and simply using an assumed selectivity at age to predict the adult abundance index. 3) The growth parameters and CV of length-at-age are assumed to be constant. It could be considered to allow these model parameters to be time varying quantities. The usual way to do this in the BAM model is by allowing penalized deviances, but it could also be considered to use a time-series formulation (using the built-in tools for random effects in AD Model Builder). 4) Finally it could be considered to include individual surveys independently into the model, which would allow individual selection curves, which would better reflect that these surveys are coming from different areas and times.

The assessment team stated that the age reading from scales is not considered problematic for Atlantic menhaden, but that no good comparison to age reading from otoliths had been conducted. Age readings are very important for this assessment, as the length compositions have proven difficult, so any additional validation is useful.

The catches are dominated by relative few ages (Fig. 7.2.1.1-10), so an alternative to using parametric shape-constrained selectivity-at-age functions, would be to simply use a parameter for each age class. This would give a more flexible selectivity with few extra model parameters.

When propagating uncertainties via the parametric Monte Carlo bootstrap it would be more realistic to sample parameters from their joint distribution (accounting for correlation). In cases where that is difficult, because some parts of the modeling is done external to the main assessment model, it could be considered to use an integrated approach instead. Further it could be considered to include the likelihood weighting constants in the parametric Monte Carlo bootstrap procedure.

For reference points it would be preferable to base the fishing mortality definition on a wider range of ages (e.g. from age 2 to age 4) instead of only age 2.

In preparation for the long term objective of moving to ecosystem-based management, it could be considered, to extend the current single-species age-based assessment with a multi-species production component.

> Provide feedback on the proposed ecological reference points that account for Atlantic menhaden's role as a forage fish. Evaluate the appropriateness and feasibility of the proposed approach. Provide alternative suggestions, if necessary. Note: this TOR is aimed at obtaining preliminary feedback on a proposed reference point development approach that would inform future ecosystem-based management plans. Further technical development and peer review would be required before these reference points would be used in management.

The discussion at the review meeting was focused on selecting the right level of complexity for models which could be used as basis for ecological reference points that account for Atlantic menhaden's role as a forage fish. The assessment team presented the panel with a comprehensive review of sensible options. These ranged from using simple ecosystem indicators (e.g. predator/prey ratios), over multi-species production models, to full featured multi-species catch-at-age models and Ecosim.

There is always a desire to develop and use the most complex model in an effort to account for as many details as possible of the real system. There are however serious practical downsides to these highly complex models. The complex models take a long time to develop and configure correctly, and often run times are too long to do comprehensive simulation studies. Such complex models are hard to debug, and it is difficult to verify that the model actually worked as intended.

Aside from the practical obstacles the model choice should be guided by the model's ability to describe the observations. Here a key criteria should be prediction power. Adding more model parameters and covariates to a model will always improve the fit to the observations, but adding too many parameters will lead to over-fitting. As illustrated below using a too simple model will introduce a bias in the model

predictions, using a too detailed model will start to adapt the model parameters to the observation noise (over-fitting), which will diminish the prediction power. Choosing the right level of model complexity will make the model better at predicting future observations.

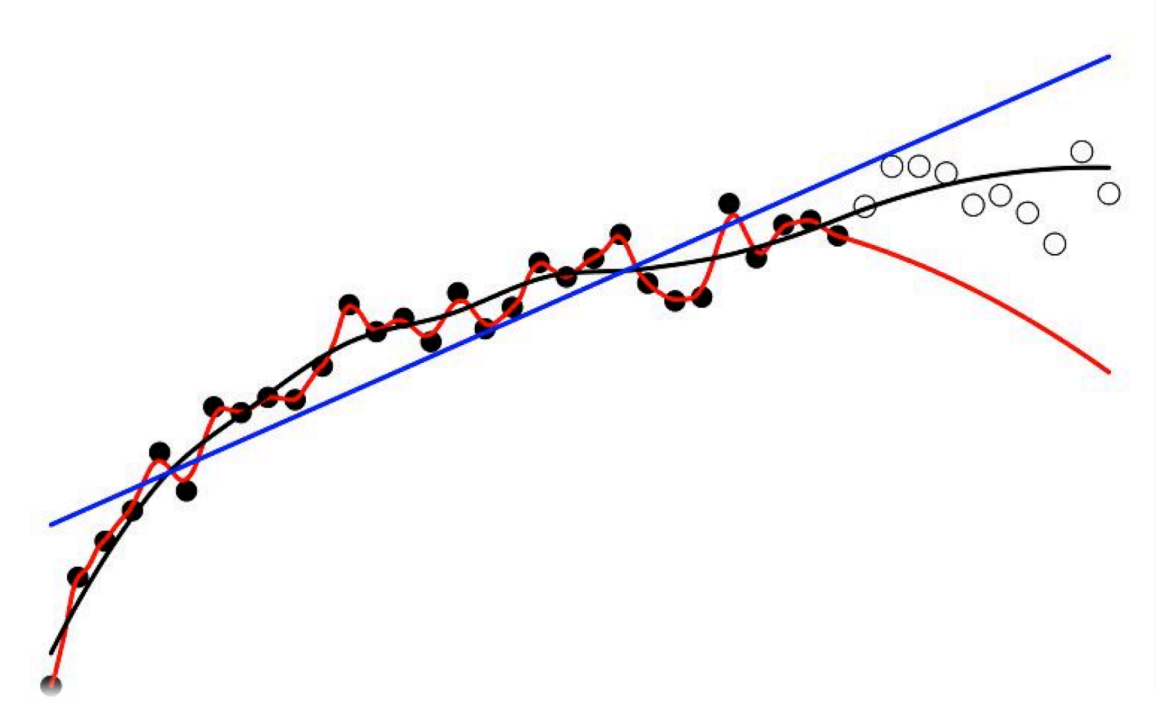


Illustration of prediction ability of models with different level of complexity

In practice the different candidate models should be compared by 1) Splitting the data set in two parts. 2) Optimizing the model only w.r.t. the first part of the data set. 3) Predicting the second part of the data set. 4) Comparing results across models. Using the prediction power as part of the model selection criteria will help avoid over fitting.

Another thing that should guide the model choice is the objectives of the resulting reference points. It was clear from the assessment team's presentation that these objectives were not clearly stated. If the objective is ensure Atlantic menhaden's role as a forage fish (as indicated by this TOR), then the model to be used for this should be one which can answer questions like: If the stock size of Atlantic menhaden is reduced by this amount, what influence is it likely to have on other interesting stocks (e.g., bluefish, striped bass, and weakfish)? The simplest level of models able to answer such questions are the multi-species production models.

Concern was however raised about how one would arrive at one joint conclusion if for instance the detailed age-specific single-species assessment model for Atlantic menhaden gave one result, but the more coarse multi-species production model gave a substantially different result. As a way forward this reviewer suggested that the assessment team would build a combined model. The existing and working

single-species assessment model could be extended by linking it to a multi-species production model, which is simple in comparison. The stock size of menhaden would be the common part (entering both models). Then inference about fishing mortality, stock sizes, and recruitment would only have to be drawn from one model. A combined model like that would still be a relatively simple, but it would be a logical first step to answer the important questions about menhaden's influence on other stocks.

> Prepare a peer review panel advisory report summarizing the panel's evaluation of the stock assessment and addressing each peer review term of reference. Develop a list of tasks to be completed following the workshop. Complete and submit the report within 4 weeks of workshop conclusion.

This task was completed jointly with the other panel members and the chair. This reviewer especially helped shape the report for terms of reference 2 and 3.

Appendix 1: List of documents

Atlantic menhaden review workshop document list.

Document #	Title
Assessment Report	
SEDAR40 – 1.1	Atlantic Menhaden Benchmark Stock Assessment (main body of report)
SEDAR40 – 1.2	Assessment Report Tables
SEDAR40 – 1.3	Assessment Report Figures
SEDAR40 – 1.4	Appendix A. 2014 MSVPA update report & appendices
SEDAR40 – 1.5	Appendix B. Atlantic menhaden tagging report
SEDAR40 – 1.6	Appendix C. Atlantic menhaden Beaufort Assessment Model (BAM) equations and code
SEDAR40 – 1.7	Appendix D. Projections methodology and example assuming constant landings
SEDAR40 – 1.8	Appendix E. Ecological Reference Points (ERP) report
Supplementary Materials	
SEDAR40 – 2	Fishery Dependent Indices
SEDAR40 - 3	Powerplant Impingement
SEDAR40 – 4	Fishery Independent Survey Standardization
SEDAR40 – 5	Fishery Independent Index Standardization Guidelines
SEDAR40 – 6	Atlantic Menhaden Stock Assessment Update – 2012
SEDAR40 – 7	Atlantic Menhaden Benchmark Stock Assessment and Review Panel Reports – 2010
SEDAR40 – 8	Hierarchical analysis of multiple noisy abundance indices. (P. Conn 2010)
SEDAR40 – 9	A proposed, tested, and applied adjustment to account for bias in growth parameter estimates due to selectivity (Schueller et al. 2013)

Appendix 2: Statement of work for Anders Nielsen

SEDAR 40 ASMFC Atlantic menhaden Review Workshop

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 40 will be a CIE assessment review conducted for ASMFC Atlantic menhaden. 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 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 40 are within the jurisdiction of the Atlantic States Marine Fisheries Commission and the states of Florida, Georgia, South Carolina, North Carolina, Virginia, Maryland, Delaware, Pennsylvania, New Jersey, New York, Connecticut, Rhode Island, Massachusetts, New Hampshire, and Maine.

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 in total a combination of expertise in stock assessment, statistics, fisheries science, and marine biology sufficient to complete the primary task of providing peer-review advice in compliance with the workshop Terms of Reference. It would be preferable for CIE reviewers to have expertise in forage fish

population dynamics and ecology, age-based assessment modeling, multi-species/ecosystem modeling and ecological reference points, and/or management strategy evaluations/decisional frameworks.

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 **Atlantic Beach, North Carolina** during **December 9-11, 2014**.

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/

http://deemedexports.noaa.gov/compliance_access_control_procedures/noaa-foreign-national-registration-system.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 during the panel review meeting at the **Atlantic Beach, North Carolina** during **December 9-11, 2014.**
3) Conduct an independent peer review in accordance with the ToR (**Annex 2**) in Atlantic Beach, North Carolina during **December 9-11, 2014.**

4) **No later than January 2, 2015**, each CIE reviewer shall submit an independent peer review report addressed to the "Center for Independent Experts," and sent to Dr. Manoj Shivlani, CIE Lead Coordinator, via email to shivlanim@bellsouth.net, and Dr. David Sampson, CIE Regional Coordinator, via email to 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**.

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

November 3, 2014	CIE sends reviewer contact information to the COTR, who then sends this to the NMFS Project Contact
November 24, 2014	NMFS Project Contact sends the CIE Reviewers the pre-review documents
December 9-11, 2014	Each reviewer participates and conducts an independent peer review during the panel review meeting
January 2, 2015	CIE reviewers submit draft CIE independent peer review reports to the CIE Lead Coordinator and CIE Regional Coordinator
January 16, 2015	CIE submits CIE independent peer review reports to the COTR
January 21, 2015	The COTR distributes the final CIE reports to the NMFS Project Contact and regional Center Director

Modifications to the Statement of Work: This 'Time and Materials' task order may require an update or modification due to possible changes to the terms of reference or schedule of milestones resulting from the fishery management decision process of the NOAA Leadership, Fishery Management Council, and Council's SSC advisory committee. A request to modify this SoW must be approved by the Contracting Officer at least 15 working days prior to making any permanent changes. The Contracting Officer will notify the COTR within 10 working days after receipt of all required information of the decision on changes. 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) The CIE report shall completed with the format and content in accordance with **Annex 1**,
(2) The 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:

Allen Shimada
NMFS Office of Science and Technology
1315 East West Hwy, SSMC3, F/ST4, Silver Spring, MD 20910
Allen.Shimada@noaa.gov Phone: 301-427-8174

William Michaels
NMFS Office of Science and Technology
1315 East West Hwy, SSMC3, F/ST4, Silver Spring, MD 20910
William.Michaels@noaa.gov Phone: 301-427-8155

Manoj Shivlani, CIE Lead Coordinator
Northern Taiga Ventures, Inc.
10600 SW 131st Court, Miami, FL 33186
shivlanim@bellsouth.net Phone: 305-968-7136

Key Personnel:

NMFS Project Contact:

Julia Byrd
SEDAR Coordinator
4055 Faber Place Drive, Suite 201
North Charleston, SC 29405
(843) 571-4366
julia.byrd@safmc.net

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: Tentative Terms of Reference for the Peer Review

SEDAR 40 ASMFC Atlantic menhaden Review Workshop

1. Evaluate the data used in the assessment.
   a. Are data decisions made during the DW and AW justified (i.e. sound and robust)?
   b. Are input data series reliable and sufficient to support the assessment approach and
   findings?
   c. Are data applied properly within the assessment?
   d. Are data uncertainties acknowledged, reported, and within normal or expected levels?

2. Evaluate the methods used to assess the stock, taking into account available data.
   a. Are methods scientifically sound and robust?
   b. Are assessment models configured properly and used consistent with standard practices?
   c. Are the methods appropriate for the available data?
   d. If multiple models or model configurations were considered, evaluate the explanation
   of any differences in results and justification of a base model.

3. Consider how uncertainties in the assessment, and their potential consequences, are addressed.
   a. Comment on the degree to which methods used to evaluate uncertainty reflect and
   capture the significant sources of uncertainty in the population, data sources, and assessment methods.
   b. Are the implications of uncertainty on technical conclusions are clearly stated?

4. Evaluate the assessment findings with respect to the following:
   a. Are estimates of biomass, abundance, and exploitation rate reliable and consistent with
   input data and population biological characteristics? Are they useful to support inferences
   on stock status?
   b. Is the stock overfished relative to biomass or abundance threshold reference points?
   Where is the stock relative to biomass or abundance management targets? What information supports this conclusion?
   c. Is the stock undergoing overfishing relative to fishing mortality threshold reference
   points? Where is the stock relative to fishing mortality management targets? What information supports this conclusion?

d. Is there an informative stock recruitment relationship? Is the stock recruitment curve
reliable and useful for evaluation of productivity and future stock conditions?
e. Are the quantitative estimates of the threshold reference points reliable for this stock?
If not, are there other indicators that may be used to inform managers about stock trends
and conditions?

5. If a minority report has been filed, review minority opinion and any associated analyses. If possible, make recommendation on current or future use of alternative assessment approach presented in minority report.
6. Review the Technical Committee's recommendations on research, data collection, and assessment methodology and make any additional recommendations or prioritizations, if warranted.

7. Provide guidance on key improvements in data or modeling approaches which should be considered when scheduling the next assessment.

8. Provide feedback on the proposed ecological reference points that account for Atlantic menhaden's role as a forage fish. Evaluate the appropriateness and feasibility of the proposed approach. Provide alternative suggestions, if necessary. *Note: this TOR is aimed at obtaining preliminary feedback on a proposed reference point development approach that would inform future ecosystem-based management plans. Further technical development and peer review would be required before these reference points would be used in management.*

9. Prepare a peer review panel advisory report summarizing the panel's evaluation of the stock assessment and addressing each peer review term of reference. Develop a list of tasks to be completed following the workshop. Complete and submit the report within 4 weeks of workshop conclusion.

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 40 ASMFC Atlantic menhaden Review Workshop

Atlantic Beach, North Carolina, December 9-11, 2014.

Tuesday

9:00 a.m.	**Convene**	
9:00 a.m. – 9:30 a.m	**Introductions and Opening Remarks Coordinator**	
	- Agenda Review, TOR, Task Assignments	
9:30 a.m. – 12:00 a.m.	**Assessment Presentation**	**TBD**
12:00 a.m. – 1:30 p.m.	**Lunch Break**	
1:30 p.m. – 3:30 p.m.	**Continue Presentations / Panel Discussion**	**Chair**
	- *Assessment Data & Methods*	
	- Identify additional analyses, sensitivities, corrections	
3:30 – 4:00	**Break**	
4:00 – 6:00	**Continue Discussion**	**Chair**

***Tuesday Goals*:** Initial presentations completed, sensitivity and base model discussion begun

Wednesday

8:30 a.m. – 12:00 a.m.	**Panel Discussion**	**Chair**
	- *Continue deliberations*	
	- Review additional analyses	
12:00 a.m. – 1:30 p.m.	**Lunch Break**	
1:30 p.m. – 3:30 p.m.	**Panel Discussion**	**Chair**
	- Continue deliberations	
	- Review additional analyses	

3:30 p.m. – 4:00 p.m.	**Break**	
4:00 p.m. – 6:00 p.m.	**Panel Discussion/Panel Work Session**	**Chair**
	- Recommendations and comments	

***Wednesday Goals*:** sensitivities and modifications identified, preferred models selected, projection approaches approved, Report drafts begun

Thursday

8:30 a.m. – 10:30 a.m.	**Panel Discussion**	**Chair**
	- Final sensitivities reviewed.	
	- Projections reviewed.	**Chair**
10:30 a.m. – 11:00 a.m.	**Break**	
11:00 a.m. – 1:00 p.m.	**Panel Discussion or Work Session**	**Chair**
	- Review Reports	
1:00 p.m.	**ADJOURN**	

Thursday Goals: Complete assessment work and discussions, final results available. Draft Reports reviewed.

Appendix 3: List of participants.

Review Workshop Panelists

Mike Jones	Review Panel Chair	ASMFC Appointee
Carmen Fernandez	Reviewer	CIE
Anders Nielsen	Reviewer	CIE
John Simmonds	Reviewer	CIE

Analytical Representatives

Amy Schueller	Lead analyst	NMFS Beaufort
Genny Nesslage	Assessment Team	ASMFC
Jason McNamee	Assessment Team	ASMFC TC
Joe Smith	Assessment Team	NMFS Beaufort

Observers

Bob Beale	Executive Director	ASMFC
Louis Daniel	Chairman	ASMFC / NCDMF
Erik Williams	SEFSC	NMFS Beaufort

Council and Comission Staff

Julia Byrd	SEDAR Coordinator	SEDAR
Julie O'Dell	Admin.	SEDAR/SAFMC
Mike Waine	Menhaden Plan Coordinator	ASMFC
Shanna Madsen	Multispecies Coordinator	ASMFC
Pat Campfield	Science Program Director	ASMFC

Review Workshop Attendees

Jud Crawford, Pew Trusts
Eric Fitzpatrick, NOAA
Aaron Kornbluth, Pew Trusts
Ron Lukens, Omega Protein
Mike Prager, Prager Consulting
Kyle Shertzer, NOAA
Doug Vaughan