Center for Independent Experts (CIE) Independent Peer Review Report of the SEDAR 40 Atlantic Menhaden Review Workshop

Carmen Fernández

January 2015

Executive Summary:

This report provides a review of the Atlantic menhaden 2014 benchmark stock assessment. The report follows the structure requested by the CIE and addresses the 9 ToRs set for this review (shown in the CIE Statement of Work, Appendix 2 to this report). The ToRs cover stock assessment data and methods, biological reference points and stock status in relation to them, research recommendations, and development of ecological reference points. A review meeting took place in Atlantic Beach (North Carolina) during December 9-11 2014, in which the reviewers (three CIE reviewers and the panel chair, listed in Appendix 3 to this report) had a good opportunity to discuss all these aspects with the assessment scientists.

Menhaden along the east coast of the USA has been historically exploited by reduction and bait fisheries, with the reduction fishery accounting for most of the landings (around 80% in the last fifteen years, although its volume with respect to the bait fishery has been decreasing through time). Recreational fisheries also exist, but are very small in relation to the commercial fisheries. The 2014 menhaden benchmark stock assessment contains many changes with respect to the previous assessment, both in terms of the data used in the assessment and the model configuration (although the modeling platform remains the same as before, the Beaufort Assessment Model). The assessment covers the years 1955-2013.

A detailed discussion of my comments is provided later in this report and a complete bullet point list of suggestions and recommendations is presented at the end of the main body of this report. Here I provide a concise summary of only the points I identify as most relevant.

I consider the Atlantic menhaden stock assessment to be consistent with best available science. The work conducted by the Assessment Team is of high quality, methods are sound, and choices well justified and consistent with standard practices. Following from discussions during the review meeting, the reviewers have proposed a new base run that differs in the following aspect from the base run originally proposed by the Assessment Team: the new base run downweights the length composition data of the abundance indices in the assessment model by a factor of 10 (relative to the weights used in the base run originally proposed by the Assessment Team); all other model settings were unchanged. The reasons for proposing the new base run are explained briefly in this Executive Summary and in more detail under ToRs 1 and 2 of the "Summary of findings for each ToR" section of this report.

Concerning the data used in the assessment, I consider the treatment of the commercial fisheries data (reduction and bait fisheries) to be sound and robust.

Fishery-independent abundance indices for the adult (NAD and SAD indices) and juvenile (JAI index) menhaden population were developed for the 2014 stock assessment benchmark; each of these is a composite index that combines several survey indices using the hierarchical method of Conn (2010). Much of the discussion and explorations during the review meeting were in relation to these composite indices. Whereas continuing exploration of these indices during coming years is desirable, my conclusion is that the JAI, NAD and SAD indices are appropriate for use in the menhaden stock assessment. On the other hand, the method used to derive length compositions for NAD and SAD (directly combining the lengths of all observed fish in the component surveys, without applying any kind of weighting) does not seem entirely appropriate, given the heterogeneity of the lengths observed in different component surveys and the far from uniform weights the surveys received in Conn's method. The stock assessment fitted poorly to the length compositions of NAD and SAD and there was concern among the reviewers that these length composition data may lead to biases in assessment results (e.g. in estimated population abundances). After several trials during the review meeting (removing the length composition data completely, downweighting them, or removing them while treating the values of the indices' selectivities-at-age as known) the reviewers proposed that the run that downweights length composition data by a factor of 10 (relative to the base run originally proposed by the Assessment Team) be considered as the new base run. I agree with this recommendation. I also suggest exploring for the next benchmark stock assessment the possibility of obtaining age composition data for NAD and SAD, as well as exploring possibilities for obtaining more representative length compositions.

The Assessment Team has performed a comprehensive and useful uncertainty and sensitivity analysis of the assessment results, including also stock status relative to reference points and future projections. The combination of evaluating uncertainties via a Monte Carlo Bootstrap approach, and conducting sensitivity runs and a retrospective analysis, gives a very complete picture and adds significantly to the understanding and confidence in the results. I agree with the overall Monte Carlo Bootstrap approach for evaluating uncertainty, although I suggest giving some further consideration to the sources of uncertainty included in the simulations and that correlations be taken into account whenever possible.

The work conducted by the Assessment Team (their initially proposed base run, uncertainty analyses and sensitivity runs) indicates that the stock is not overfished (relative to the stock fecundity threshold) and is not undergoing overfishing (relative to the fishing mortality threshold). The stock is also estimated to be currently above the fecundity target and below the fishing mortality target. These conclusions hold with the current reference points (based on 15%SPR for thresholds and 30%SPR for targets) and the revised reference points proposed by the Assessment Team (obtained from the maximum and median F at age 2 during 1960-2012, corresponding to 20%SPR for thresholds and 36%SPR for targets). The same conclusions also hold for the new base run proposed during the review meeting; for this new base run, the revised reference points proposed by the Assessment Team (based on maximum and median F at age 2 during 1960-2012) correspond to 20%SPR for thresholds and 39%SPR for targets.

The Assessment Team proposed to revise the reference points as indicated in the previous paragraph in order to provide a more appropriate measure of sustainability given the historical perception of stock development from the 2014 benchmark assessment (which is quite different from the previous assessment). In the absence of more clearly specified management objectives for this stock, the proposal made by the Assessment Team seems sensible and would be expected to be in line with sustainable exploitation (assuming no major changes happen in the ecosystem relative to the situation generally experienced during 1960-2012). Some comments and suggestions concerning the technical aspects of the reference points calculation (range of ages to consider for reference F in the method, and range of years to consider for fishery and biological parameters) are presented in the body of this report (see discussion of ToR 4e). If the suggestions are followed, the %SPR corresponding to the reference points will likely be somewhat different from 20% and 39%.

The review also had a ToR to provide preliminary feedback on ways forward for the development of Ecological Reference Points (ERP) taking into account menhaden's role as a forage fish. A range of possibilities, of different levels of complexity, was presented by the Assessment Team and discussed during the review meeting. Almost all the approaches could address some aspects of ERPs, and Annex E of the stock assessment document usefully identifies types of objectives that each of the approaches could address. Most of the approaches could address the objective of having a sustainable menhaden fishery in the light of forage pressure, but if management objectives include having enough prey to support predator species at preferred biomass levels (which may differ from current biomass levels) the range of possible approaches is reduced. Additionally, if feedbacks from prey abundance to predator abundance are expected to be important (in addition to impacts of predators on prey) then the development of a multispecies model seems most relevant. This could perhaps be done in parallel to the development of some of the simpler approaches. I would suggest that the multispecies model be kept as simple as possible, although it must obviously account for the main interactions so that the resulting modeled population dynamics are sufficiently realistic to be useful for management purposes. I expect an iterative process of dialogue with managers (likely also involving stakeholders at some stage, although this depends on governance frameworks in the USA, with which I am not familiar) will be needed to find the most appropriate way forward.

Background:

The current assessment of Atlantic menhaden covers the years 1955-2013. Commercial landings are split into four fleets: reduction fishery (two fleets: north and south) and bait fishery (two fleets: north and south). The north fleets take their catches in the North Atlantic and Middle Atlantic NMFS statistical reporting areas for menhaden, and the south fleets in the Chesapeake Bay and South Atlantic reporting areas. Menhaden discards in commercial fisheries are considered to be minimal, and commercial landings are assumed to be equal to removals. The recreational fishery represents <1% of the total stock removals; for stock assessment purposes, the recreational removals were incorporated into the bait fishery.

The reduction fishery accounts for around 80% of the landings, although its volume relative to that of the bait fishery has been decreasing through time. Landings from the reduction fishery were at their highest during the 1950s (with a peak of around 700,000 metric tons, mt) and then decreased strongly during the 1960s (particularly in the northern area, due to geographical contraction of the menhaden stock). The reduction fishery experienced some increase during the 1970s and 1980s, once again expanding into northern areas, with annual landings around 300,000-400,000 mt. During the 1990s, menhaden again became scarce north of Long Island Sound and the reduction fishery landings decreased. The number of reduction processing plants has decreased from over 20 in the 1950s to just one left since 2005 (located in Reedville, Virginia); annual reduction landings since 2005 have been 160,000 mt on average. Total reduction landings statistics are believed to be both accurate and precise. Age compositions from the reduction fishery are estimated at the port/week/area level, in order to account for the annual cycle of menhaden migration, where larger and older fish tend to be distributed further north during the summer. The sampling intensity and estimation procedure are considered to be adequate.

Bait menhaden harvest originates from directed and bycatch fisheries. Reported bait landings have historically been incomplete for various reasons, but data collection has been improving in many areas. Although bait landings records go back to 1955 (and are included in the stock assessment from 1955), the most reliable values are considered to be those since year 1985. Bait landings statistics show overall stability during 1985-2006, with a clear increase since 2007. Part of this increase may be due to better data collection, but also to increased interest in menhaden for bait as a consequence of limitations on catch of Atlantic herring, which was traditionally used for bait in some fisheries. It has been possible to characterize the age compositions of the menhaden bait fishery from 1985 to present, albeit based on limited data and only at the year/area level (hence, with coarser resolution than for the reduction fishery). Sampling in the bait fishery has increased substantially in 2013.

Three composite fishery-independent menhaden abundance indices (each derived from several separate survey indices, combined following the hierarchical method proposed in Conn, 2010) were developed for the 2014 stock assessment benchmark: one juvenile index (JAI) and two adult indices (SAD for the south and NAD for the north). The assessment also uses length composition data for SAD and NAD.

The age composition data from both reduction and bait fisheries indicate that fish caught in the north are generally older than in the south. Likewise, the NAD index sees larger fish than the SAD index.

The stock is assessed using the so-called Beaufort Assessment Model (BAM). It is a forward-projection statistical catch-at-age model with an annual time step and covers ages 0-6+. The data used in the assessment are the landings and age compositions of the four fleets, the three abundance indices (JAI, NAD and SAD), and the length compositions of the NAD and SAD indices. Double logistic functional forms are assumed for the selectivity-at-age of the four commercial fleets and the SAD index, whereas logistic selectivity-at-age is assumed for the NAD index (the parameters of the corresponding functional forms are estimated within the assessment model); the JAI index is assumed to represent exclusively age-0 menhaden. Natural mortality, mean length, weight, proportion mature and fecundity (number of eggs) at age are inputs to the stock assessment (all these parameters are treated as known in the assessment). Most of these biological parameters were first derived as a function of length; subsequently, a weighted average with weights proportional to cohort-specific mean length-at-age values was calculated for each age, resulting in annually-varying biological parameters for each age (with the exception of natural mortality, which is assumed to be age-dependent but constant over time).

Observation equations (likelihoods) are assumed to be log-Normal for catch and abundance indices, and robust Multinomial for age and length composition data. BAM is written in AD Model Builder, and the software allows for maximization of the objective function and for calculation of the Hessian matrix. For menhaden, however, uncertainty in assessment results was not calculated from the Hessian matrix, but from a Monte Carlo bootstrap procedure that also incorporated uncertainty in M and maturity. A thorough sensitivity analysis to various model configurations was additionally conducted, as well as a retrospective analysis.

The assessment estimates an exceptionally high year class in 1958. After this, recruitment has generally been higher from the mid-1970s to the mid 1980s than in other time periods. Recruitment in the last 20 years has mostly been below the time series average. Despite this, in recent years stock biomass has been high and fecundity is close to the highest observed levels; the proportion of older individuals in the population is estimated to have increased, in line with the lower fishing mortalities estimated for the last decade.

Biological reference points (the existing ones and the ones proposed in the 2014 stock assessment benchmark) are based on %SPR targets and thresholds. Relative to the biological threshold reference points, the stock is not overfished nor is it undergoing overfishing. Menhaden is a forage fish, with predators such as striped bass, bluefish and weakfish. The ASMFC is working on the development of ecological reference points to better take into account species interactions and the role of menhaden in the ecosystem as a forage fish.

Detailed discussion of several aspects of the input data, assessment model, stock status and reference points is provided under the “Summary of findings for each ToR” section of this report.

Description of review activities and reviewer’s role:

The review was organized around a meeting held at Atlantic Beach, North Carolina, during December 9-11, 2014. The documents indicated in the Bibliography section of this report were provided to the reviewers two weeks in advance of the meeting and constituted the central material for the review. Nine ToRs were given for the review process, which are shown in Appendix 2 to this report.

The review meeting followed closely the planned agenda of presentations, developing as follows:

Tuesday, December 9, 2014
Most of the day was spent on presentations given by the Assessment Team. Each presentation was followed by questions and discussion. The following topics were presented:

Regulatory History; Life History; Commercial Reduction Fishery; Commercial Bait Fishery; Indices of Abundance; Assessment Model and Results; Reference Points and Stock Status; Projection Methodology.

At the end of the day, the reviewers made several requests to the Assessment Team, regarding extra assessment model runs and additional explorations to be prepared for consideration on the following day.

Wednesday, December 10, 2014
The morning started with a summary presentation by the review panel and follow up discussion. The Assessment Team then presented results from the requests made on the previous day and this led to considerable additional comment and discussion.

In addition, there were two more scheduled presentations: Research and Modeling Recommendations; Ecological Reference Points Methods.

Discussion and additional model runs were conducted during the afternoon, which led to proposing a new base run for the assessment of the Atlantic menhaden stock. The public meeting was formally closed at the end of the day.

Thursday, December 11, 2014
The reviewers met to organize production of the report and to start working on it.

All three CIE reviewers and the panel chair fully participated in all aspects of the review. The procedure I followed to provide this independent review report was to read carefully in advance the documents provided for the review, then to exchange views and clarify questions with the assessment scientists and the other reviewers during the meeting and, finally, to review some of the documents once again (benefiting from the insights gained during the meeting) and go through some additional literature as a follow up to some of the discussions held during the meeting.

The meeting was very well organized and ran efficiently. All review materials had been provided by the assessment scientists two weeks in advance (even a bit earlier than that), which I found very useful as it gave me time to read carefully through them in advance of the meeting. The materials were well structured and clear, which again facilitated my task as a reviewer. The assessment scientists helpfully clarified questions during the review meeting and, in my opinion, there was a good exchange of knowledge and experience in both directions. I was impressed by the amount of work that had taken place within the SEDAR process before the review meeting, including a data workshop and an assessment workshop. This was reflected in the good quality of the work available for the review.

Summary of findings for each ToR:

This section presents the main points that arose during the review, according to my own perspective and understanding of the issues discussed. A few additional thoughts from following up after the meeting on some aspects of the work presented and discussed there, are also included. This section is organized following the ToRs.

ToR 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?**

The data sources used in the stock assessment were already explained in the "Background" section of this report. During the review meeting, the Assessment Team provided presentations about all of them.

The reviewers had some questions about the commercial fisheries data, which were mainly for clarification purposes. Reduction fisheries appear to be well sampled and the procedures used to raise the age samples to age compositions of the fishery seem appropriate. Previous concerns about a potential "topping off" bias due to the last purse-seine set of the trip (the set that may be sampled from the trip) being sometimes taken in a more southern location than most other sets in the trip appear to have been resolved by the use of Captains Daily Fishing Reports (logbooks) to allocate the weekly landings to fishing areas. On the other hand, the Assessment Team explained that reporting of bait landings has historically been incomplete, although the situation has been improving over time. A power analysis was conducted in 2012 to statistically determine the sampling level needed to adequately represent the age composition in the bait fishery; following from that analysis, sampling in the bait fishery has increased substantially in 2013. This is a good development and I endorse the Assessment Team's recommendation to continue the current level of sampling in the bait fishery.

Considerable discussion took place during the review meeting regarding the fishery-independent composite abundance indices (JAI, NAD and SAD) and the length compositions of NAD and SAD. These indices are an important development since the previous assessment, and there was a lot of interest from the reviewers to understand as well as possible how they had been constructed. The composite indices JAI, NAD and SAD are a combination of several separate indices. In brief terms, the following process has been followed to arrive at each of the composite indices:

- An initial selection among all possible fishery-independent surveys was made using a standard set of criteria (absence of hyperstability; sufficiently long time series; spatial extent; takes zero values into account; consistency through time; identification to species level; standardization model converges; information to determine if the index represents juvenile or adult fish)
- The indices were standardized to account for factors that might affect catchability of menhaden (as the surveys were not designed for menhaden). The standardization followed a set of common guidelines provided to all analysts doing this work.
- The selected standardized indices (obtained from the previous steps) were then included in Conn's hierarchical model (Conn 2010) to calculate a composite index.

This procedure was used to produce a coast wide juvenile index (JAI, made up of 16 separate surveys which use a variety of gears) and two adult indices (SAD for the south and NAD for the north). Nine separate trawl surveys were available for adult menhaden. Based on a Principal Components Analysis of their length compositions, the nine surveys were split in two sets: two surveys catching smaller adult fish were used to create the SAD index, whereas the other seven surveys catching larger adult fish were used to create the NAD index.

The reviewers requested additional information about the standardization process, aiming in particular to ensure that the variables used to standardize the indices were solely representative of catchability (i.e. not confounded with abundance). In response to this, the Assessment Team prepared an additional presentation concerning the standardization procedure. The presentation made clear that the standardization guidelines given to the analysts emphasized that the variables used for standardization should be those that might affect catchability but not abundance; therefore, provided analysts were able to implement this guideline successfully, the standardization should have worked well. The presentation also displayed graphs of the indices before and after standardization. Considering the overall set of survey indices, the differences between the indices before and after standardization were not major, although they were substantial for some of the surveys. The reviewers concluded that the overall approach followed to select the surveys and to standardize the indices is appropriate.

The reviewers also wanted to understand the weights that Conn's method gave to the different surveys and how they differed, for example, from a weighted average with weights proportional to the areal extent of each survey. The Assessment Team explained that the index- and year- specific sampling variances used as inputs in Conn's method, had been taken from the coefficients of variation estimated in the standardization process. Conn's method assumes an extra component for the variance of each individual index; this component represents the variance of annual random deviates of the index' catchability and is estimated together with the derivation of the composite index. In order to gain an understanding of the weighting Conn's method gave to each individual index, the inverse of the total variance (sampling variance + extra variance component) for each individual index was extracted from Conn's output and averaged over the years for which that index is available. For the NAD composite index there was a large difference between the weights of the different individual indices (about a factor 10). The differences were not quite as large for the JAI index (about a factor 4), and were minor for the two individual indices that make up the SAD index. The rather large differences found for the weights in the NAD index created some concern among the reviewers and a quick attempt was made during the review meeting to calculate an areal-weighted index as an alternative to NAD (for comparison). This exercise is, however, not straightforward, as it requires making assumptions about the relative catchabilities of different surveys and additional complications arise because not all surveys used to create NAD span the same years. The quick attempt at calculating an areal-weighted index during the review meeting resulted in an index with unusual features that suggested the calculation method used was not appropriate; this was not further pursued.

An areal-weighted index had been calculated for juvenile menhaden by the Assessment Team and used in the stock assessment document for a sensitivity run (this is the sensitivity run that replaces the original JAI with the "areal extent JAI"). This "areal extent JAI" has more pronounced spikes than the original JAI (obtained from Conn's method); see Figure 5.3.16 of the stock assessment document. The sensitivity run using the "areal extent JAI" produces fairly similar results to the sensitivity run without any juvenile index at all. It was unclear to start with if this was due to the "areal extent JAI" being more coherent with other data sources (such as SAD and NAD) than the JAI index, or whether it was due to the "areal extent JAI" having little weight in the assessment model fit. The Assessment Team showed the coefficients of variation that had been assumed for the "areal extent JAI" annual indices, which were seen to be considerably larger than those of the original JAI; this should at least partly explain why the "areal extent JAI" sensitivity run produced fairly similar results to the sensitivity run that did not use any juvenile abundance index.

After discussing all the above matters in the review meeting, my conclusion is that, whereas continuing exploration of these indices during the coming years is desirable, the approach and methods followed by the Assessment Team to produce the JAI, SAD and NAD composite indices are appropriate and the results can be used in the menhaden stock assessment.

There was also considerable discussion during the review meeting regarding the length composition data of the SAD and NAD composite indices. The discussion was partly motivated by the poor fit of the model-predicted length compositions to the observed ones (Figures 7.1.16 and 7.1.17 of the stock assessment document). The Assessment Team explained that the length compositions of NAD and SAD had been obtained by directly combining the lengths of all observed fish in all surveys, without applying any type of weighting. They also showed a graph of the observed length compositions of each individual survey used to construct the NAD index, which indicated considerable heterogeneity between the lengths of different component surveys. Given this heterogeneity and the fact that Conn's method calculates far from uniform weights for the component surveys, re-examining in the future how the observed lengths from the different surveys can best be combined to represent the composite index seems very relevant. The method used at present is probably quite noisy and can also lead to a mismatch between the actual composite index value (obtained from Conn's procedure) and its allocated length compositions. The Assessment Team explained that their main aim with the index length compositions had been to get an idea of the index selectivities-at-age rather than to achieve very good fits to the observed length compositions. However, for the reasons explained above and the potential influence these length compositions can have on other assessment results (e.g. recruitment or population age structure estimates), the reviewers decided to explore alternative model configurations. These explorations led to proposing a new base run that gives less weight to the observed length compositions in the assessment model; I concur with this recommendation and provide further discussion of this under ToR 2.

Summarizing the comments above, the specific questions in this ToR can be answered as follows:

a. **Are data decisions made during the DW and AW justified (i.e. sound and robust)?**
   For the commercial fisheries data, they appear to be sound and robust. For the composite abundance indices, there is some concern about the length compositions of NAD and SAD, and I recommend future work to try to improve this.

b. **Are input data series reliable and sufficient to support the assessment approach and findings?**
   Taken as a whole, the data are reliable and sufficient to support the assessment approach and findings.

c. **Are data applied properly within the assessment?**
   They are applied properly on the whole. However, because of the problems identified concerning the length compositions of the NAD and SAD indices, the reviewers recommended that the length composition data be downweighted in the current assessment; I concur with this recommendation.

d. **Are data uncertainties acknowledged, reported, and within normal or expected levels?**
   Data uncertainties are clearly acknowledged and reported, and within normal or expected levels.

ToR 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.**

The Beaufort Assessment Model (BAM) is used to assess the Atlantic menhaden stock. BAM is a forward-projection statistical catch-at-age model, written in AD Model Builder. The assessment covers years 1955-2013 and ages 0-6+.

The data used in the stock assessment are: landings of four fleets (reduction fishery, north and south, and bait fishery, north and south), since 1955, age compositions of the two reduction fleets (since 1955) and the two bait fleets (since 1985); three composite abundance indices (JAI for age 0, since 1959, and the adult indices NAD, since 1980, and SAD, since 1990), length compositions of NAD (since 1988) and SAD (since 1990). These data were discussed under ToR 2.

The following biological parameters are treated as fixed (known) inputs in the assessment:

- Natural mortality: assumed to be age-dependent but constant over time. M-at-age is calculated from Lorenzen (1996) but is scaled to tagging estimates of natural mortality for ages 4-6+.
- Mean length-at-age: a Von-Bertalanffy model is fit on a cohort-basis (to account for expected density dependence in growth). Because the length-age data available come from the fishery and do no represent well the larger and smaller individuals in the population, the bias correction method of Schueller *et al.* (2014) was implemented in the estimation of mean length-at-age in the population (not in the landings).
- The annually-varying mean length-at-age is combined with (time-invariant) weight-at-length, maturity-at-length and fecundity-at-length to calculate annually-varying weight-at-age, maturity-at-age and fecundity-at-age values which are input to the stock assessment.

The assessment model has an annual time-step. Annual recruitment is modeled using random deviates around a Beverton-Holt stock-recruitment functional form with steepness fixed at 0.99, whereas the variance of the log-recruitment annual deviates is fixed at 0.6; the R_0 parameter of the Beverton-Holt curve is estimated within the stock assessment, as are the annual recruitment deviates. In practical terms, this is similar to estimating recruitment as free annual parameters.

All selectivities are specified as functions of age, with the following functional forms: double-logistic (hence dome-shaped) for the four fleets and the SAD index, and logistic (hence, asymptotic) for the NAD index. This reflects the fact that the NAD component surveys see larger fish than the SAD component surveys and the fishing fleets. The JAI index is assumed to represent exclusively age 0.

Allowing the selectivities of all four fishing fleets to be dome-shaped does of course have substantial impact on the assessment results. The stock assessment report presents several arguments to justify why allowing all fishing fleets to have dome-shaped selectivity is appropriate, namely, comparison of fishery lengths with fishery-independent and observer data indicates the presence of larger fish in the population than taken by the fishery, and the shape and CV of the fishery length-at-age data also suggest that larger fish are missing from those data; this is all supported by several graphs in the stock assessment report. The stock assessment report suggests possible reasons for the dome-shaped selectivity: menhaden undertake annual migrations, with larger individuals migrating farther north as the summer proceeds, whereas the main fishery is centered more to the south; purse-seines may target preferentially larger schools, typically made up of younger individuals; older fish may be located further offshore than the usual fishing locations. Given the information presented in the assessment report, I agree with the Assessment Team that allowing dome-shaped selectivity for the four fishing fleets is appropriate.

Assuming the selectivity (at age) of SAD to be dome-shaped and that of NAD to be asymptotic also seems appropriate. In order to fit to the length compositions of these indices, a transformation from model-predicted numbers-at-age selected by the index to numbers-at-length must be performed. This transformation assumes a Gaussian distribution for length-at-age with mean equal to the mean length-at-age estimated (outside the assessment) by the cohort-based Von-Bertanlanffy model (with bias correction); only the CV is estimated within the stock assessment model.

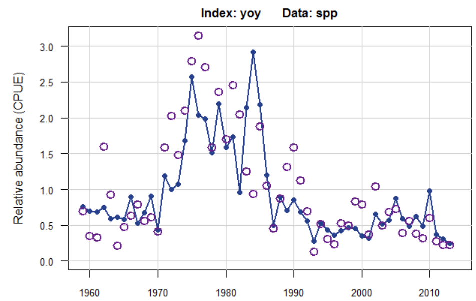
Selectivities of the two reduction fleets were modelled using three time blocks (with time-invariant selectivities within each block). A single time block was used for the selectivities of the bait fleets and the NAD and SAD indices. There is a single catchability parameter for each of NAD and SAD, whereas two catchability blocks are assumed for the JAI index (one until 1986, when only one survey is available for JAI, and another one since 1987, when many more component surveys are used in the JAI index).

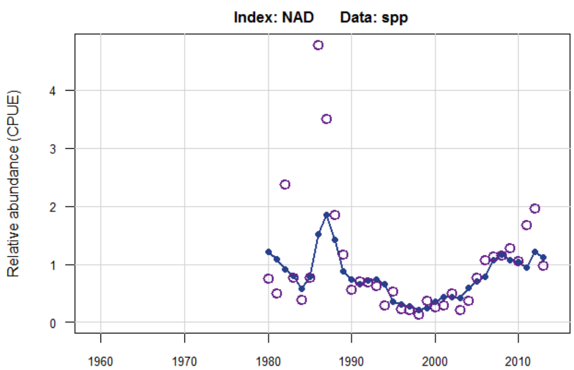
The likelihood functions are log-Normal for the landings of each of the four fleets and for the three abundance indices, and robust multinomial for the age compositions of the fleets and the length compositions of the NAD and SAD indices. The different likelihood components were weighted broadly following the suggestions of Francis (2011). The CVs of the log-Normal distributions and the effective sample sizes of the multinomial distributions initially reflected the amount of uncertainty in the different data sets and appear to have been chosen appropriately. All multinomial effective sample sizes and the CVs of the three indices were then iteratively modified until the standard deviation of their normalized residuals (SDNRs) reached close to 1. As this procedure did not produce a good fit to the indices, the Assessment Team decided to increase the weights of the indices. After trialing several possibilities, they concluded that giving weights to the indices that resulted in their SDNRs being approximately equal to 2 produced a good compromise that led to improved fits to the indices without strongly diminishing the fits to other data components; this was proposed as the base run by the Assessment Team (see stock assessment report for more detail).

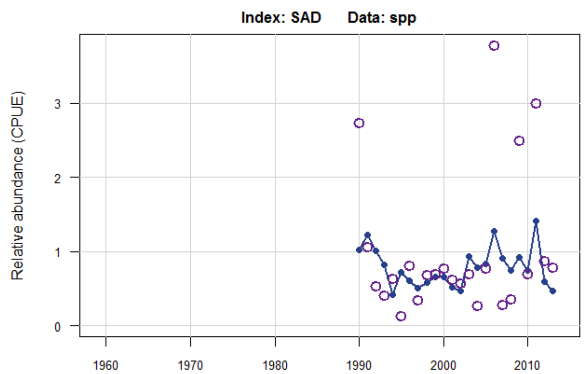
The stock assessment document presents the assessment results in detail (Sections 7.1-7.3). The model fits very closely to the landings of each of the four fleets, which is unsurprising given the low CVs assigned to these data. Fits to the JAI, SAD and NAD indices are also quite acceptable. The residuals of age composition data from the fisheries do not show strong patterns that may cause concern, although the residuals for the bait fishery in the north suggest there may perhaps have been some shift in selectivity around 2003-2005. However, concern arose during the review meeting because of the poor fit to the length composition data of NAD and SAD (Figures 7.1.16 and 7.1.17 of stock assessment document). Although the Assessment Team explained that their intention with these length composition data was only to be able to estimate the indices' selectivity-at-age, and not particularly to get a very good fit to the length compositions, there was concern among the reviewers that the poor fits may introduce bias in the assessment results.

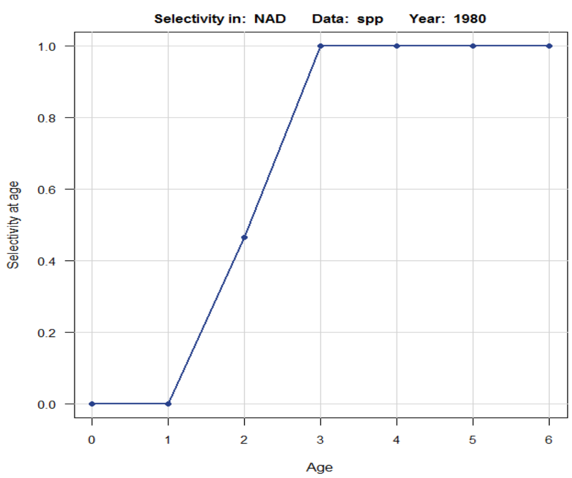
The reviewers noted that the model-predicted length compositions appear to be too wide in comparison with the observed compositions (suggesting that the estimated CV of the age to length transformation might be too large) and, in the case of SAD, model-predicted length compositions are also somewhat displaced towards larger length values. The reviewers requested a sensitivity run where the CV of length-at-age was fixed at half the value estimated in the base run. As expected, this led to narrower predicted length compositions but to no real improvement in the length compositions fit; however, the results indicated that the fit to the length compositions affected other aspects of the assessment results.

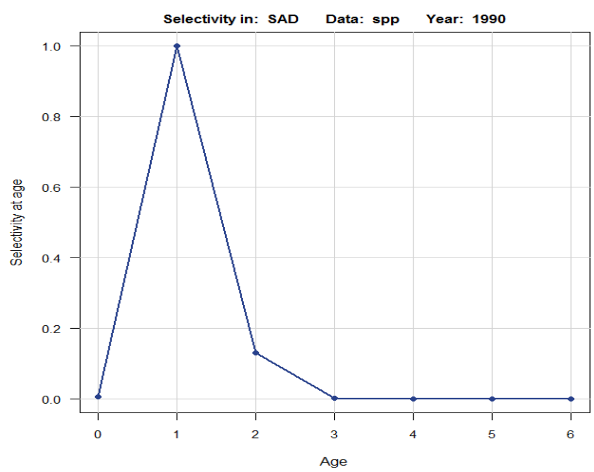
The Assessment Team explained that the length compositions of the composite indices come from a mix of surveys and that no weighting was applied to derive these length compositions; they also noted that the component surveys take place at various times in the year, which adds heterogeneity to the length compositions of the composite indices and will likely also increase interannual variability in observed lengths. Given this, and the impossibility to do anything that might potentially improve the length composition data during the review meeting, the reviewers suggested downweighting the length compositions in the assessment, or even removing these data completely. As a pragmatic way forward, downweighting the length compositions by a factor 10 was attempted. Graphs of observed and fitted values and residual plots indicated only minor changes to the fits of the commercial data (compared with the fits from the base run proposed by the Assessment Team); only minor changes were seen in the estimates of selectivity-at-age of the four commercial fleets. The fits to the three indices were also similar to before (see graphs below for JAI (labelled YOY in the graph below), NAD and SAD). However, the fits to the length compositions of the NAD and SAD indices were a bit more different from before (as would be expected, given that these data were downweighted in this run) and the estimated selectivities-at-age of these indices were also different, particularly for NAD at age 2 (now estimated to be only half selected versus fully selected in the base run proposed by the Assessment Team). The combination of these changes led to some changes in the estimates of recruitment, F and stock biomass and fecundity. As the graphs below show, F in this run is estimated to be below the threshold $F_{15\%}$ and the target $F_{30\%}$, and total stock fecundity is estimated to be above the threshold $FEC_{15\%}$ and the target $FEC_{30\%}$ (as was the case in the base run).

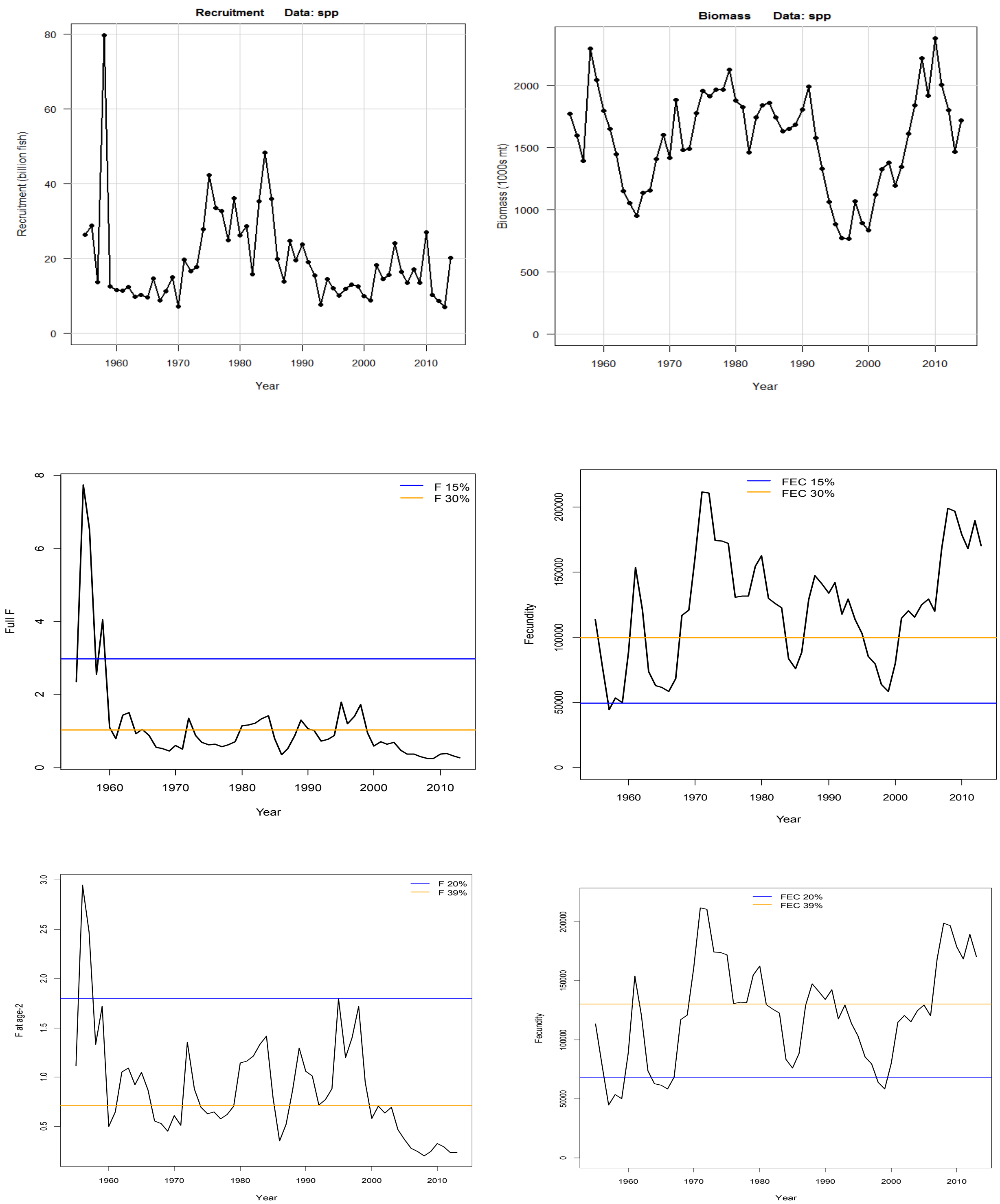












Results from run that downweights length composition data of NAD and SAD by a factor 10
(in all other respects this run is configured exactly like the base run the Assessment Team had proposed; the right-most point in the recruitment graph, corresponding to 2014, should be ignored; the bottom left panel shows F(age 2) whereas the panel above it shows full (i.e. apical) F)

Another run was attempted completely removing the length composition data but the run did not converge (not surprisingly, as one would expect some information about the index selectivities is needed in order to get model estimates). A further run where length compositions were downweighted by a factor 20 was also attempted, but this run did not converge either. Finally, a run was attempted removing the length composition data and fixing their selectivities-at-age at the values estimated in the base run; this run converged and, again, produced abundance estimates which were different from those of the base run.

After considerable discussion about the best way forward, the reviewers proposed the run that downweights length compositions by a factor 10 (relative to the base run proposed by the Assessment Team) as the new base run. I agree with this recommendation.

a. **Are methods scientifically sound and robust?**
   The methods used by the Assessment Team to conduct the menhaden assessment are sound and robust. The assessment model BAM is an appropriate choice and makes appropriate use of the available data. The only concern is in relation to the length composition data of the NAD and SAD composite indices. As explained above, this was discussed at length during the review meeting and a run where these length composition data are downweighted by a factor 10 (relative to the weights in the base run the Assessment Team had proposed) has been proposed as the new base run.

b. **Are assessment models configured properly and used consistent with standard practices?**
   Yes, the BAM assessment assessment model has been configured properly and used consistently with standard practices; however, see the discussion above about the proposed new base run with downweighted length composition data.

c. **Are the methods appropriate for the available data?**
   Yes, they are.

d. **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 document has a clear and defensible explanation of how the base run proposed by the Assessment Team was arrived at. The review meeting agreed with the choices made, except for the concerns about the length composition data, which led the reviewers to request further explorations during the review meeting and, finally, to propose a new base run with downweighted length composition data.

ToR 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 clearly stated?**

The stock assessment document clearly shows that the Assessment Team put a lot of effort in evaluating uncertainty in assessment results and in stock status relative to reference points.

The Assessment Team considered that directly using the Hessian matrix calculated from the base run model fit would underestimate uncertainty, as all biological parameters are treated as fixed known inputs in the stock assessment (whereas in reality there is uncertainty around them). Therefore, they evaluated uncertainty using Monte Carlo Bootstrap (MCB), resampling the data (1000 samples) according to their assumed statistical distributions and uncertainty levels, and also incorporating uncertainty in some biological parameters. I consider this to be a good approach; if implemented appropriately, it should lead to a more realistic characterization of true uncertainty than simply using the Hessian. In my opinion, the difficulty lies in incorporating a realistic amount of uncertainty in the biological parameters (not too little nor too much). Uncertainty was incorporated in M (which makes sense, as uncertainty exists about M and it usually has considerable impact on assessment outputs) and also in maturity (by randomly and uniformly sampling from the 95% confidence intervals of the slope and intercept parameters of the logistic regression for maturity-at-length). Whereas I consider that incorporating uncertainty in biological parameters via MCB is a good approach, I also think that it will produce more realistic outcomes if correlations between parameter estimates are taken into account (e.g. between the intercept and slope parameter estimates of the logistic regression for maturity-at-length); otherwise, the approach may produce a potentially high proportion of samples with unrealistic maturity-at-length configurations, in which case uncertainty will not be realistically characterized. I also realize now that I forgot to ask (but should have asked) why uncertainty was incorporated in maturity but not in other biological parameters (such as weight or fecundity; also note that length-at-age is a common variable in the calculation of maturity, fecundity and weight, and uncertainty in length-at-age would thus introduce both uncertainty in and correlation between biological parameters); this is my mistake, for I forgot to ask this question during the review meeting. In summary, I agree with the overall MCB approach, but suggest giving some further thought in the future to the sources of uncertainty included in the MCB and taking into account correlations between the estimates of different variables whenever possible.

The MCB approach used did not take into account uncertainty in the weights given to each data source in the model fit (i.e. likelihood component weights); instead, fixed weights equal to those used in the base run were used. Of course, if some uncertainty around these weights was incorporated in the MCB procedure this would also result in wider uncertainty intervals for the assessment output. However, getting a good understanding of what may be a realistic uncertainty level around these weights (for input to the MCB procedure) is probably not straightforward. The Assessment Team explored the impact of alternative weighting options via sensitivity runs, which I found useful.

A very complete exploration of the robustness of assessment results was conducted by the Assessment Team via sensitivity runs. Several runs examined the sensitivity of model outputs to inclusion or exclusion of the JAI, NAD and/or SAD indices, and to replacement of JAI by an alternative areal-weighted index. Other runs considered sensitivities to biological parameters (M, maturity, growth) or ageing uncertainty. Yet other runs examined the impact of model configurations: e.g. different weighting assigned to the indices, no selectivity time blocks for the reduction fleets, logistic selectivities for all four fleets, single catchability over the entire time series for the JAI index, etc. A very comprehensive set of graphical outputs is presented in the stock assessment document (Figures 7.4.1.1-7.4.1.77). No particular surprises emerged from these outputs, but it is very good that such a thorough examination was undertaken. As expected, larger values of M led to higher estimated recruitment and stock biomass. Excluding the NAD index led to much higher estimated stock biomass and stock fecundity (very likely related to the fact that NAD is the only data set with asymptotic selectivity in this assessment, so it will be influential in scaling the population abundance estimates). Removing the SAD index seems to have the opposite effect, i.e. lower estimates of stock biomass and fecundity are obtained. Assuming logistic selectivity for all four fleets also leads, unsurprisingly, to lower estimates of stock biomass and fecundity. The fits to the JAI, NAD and SAD indices were fairly similar for all sensitivity runs considered. All results were within the range of sensitivities that I consider as normal in stock assessments.

Importantly, uncertainty (via the MCB procedure) and sensitivity (through the set of sensitivity runs) of stock status relative to reference points (F relative to the threshold and target F, and Fecundity relative to the threshold and target Fecundity) was also examined (stock assessment document: Figures 7.4.1.50-7.4.1.77 for sensitivity runs and Figures 8.3.2.3-8.3.2.12 for MCB uncertainty). Stock status relative to reference points typically drives management actions, so evaluating its uncertainty and sensitivity is highly relevant. The reference points were recalculated for each MCB sample or sensitivity run, so that they could be examined consistently with assessment outputs. This seems like the correct approach to me.

A retrospective analysis was also presented, with terminal year varying from 2013 (base run) to 2009 (Figures 7.4.2.1-7.4.2.18). It can be seen that, during the period covered by the retrospective analysis, recruitment estimates for the most recent assessment years tend to be revised upwards when more years of data are incorporated in the assessment. This suggests that the SAD and NAD indices, and possibly also the age composition data of the commercial catch, may be giving a more positive signal than the recruitment index. At the same time, there is a minor retrospective pattern in the catchabilities of the SAD and NAD indices, which appear to be revised slightly upwards when more years of data are added to the assessment, and in the estimate of R_0, for which the opposite occurs; in line with this, slightly downward historic revisions of estimated stock 1+ biomass and fecundity are observed when more years of data are incorporated in the assessment. Despite these observations, I do not feel the retrospective analysis gives any cause for serious concern regarding the reliability of assessment outputs.

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.

As I explained through my comments above, the Assessment Team has performed a very comprehensive uncertainty and sensitivity analysis, which I find very useful. The combination of evaluating uncertainties via the MCB approach, conducting sensitivity runs and a retrospective analysis, gives a very complete picture and adds significantly to the confidence in assessment results.

As I noted above, I agree with the overall MCB approach, but suggest that in the future some further thought be given to the sources of uncertainty included in the MCB simulations and to accounting for correlations between the estimates of different variables whenever possible.

b. Are the implications of uncertainty on technical conclusions clearly stated?

The implications are clearly stated, both in terms of assessment estimates and the robustness of stock status relative to reference points.

The methodology for short term projections (presented in Annex D of the stock assessment document) carries the uncertainty estimated via MCB consistently into the forecast period (each of the 1000 bootstrap samples gives rise to 1 future population trajectory; the 1000 trajectories are then summarized using e.g. 5, 50 and 95 percentiles for each future year). In addition to showing the forecast distributions of recruitment, F and total stock fecundity, I think it could also be relevant to show F and fecundity relative to the reference points (this would allow a quantification of the probability of staying within targets or exceeding thresholds). Of course, it should be remembered that these are model projections and, hence, results are conditional on model specifications and the sources of uncertainty incorporated in the MCB procedure.

ToR 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?

I consider this to be the case. I have already discussed the appropriateness of the stock assessment model and its configuration choices, and explained why the reviewers proposed a slightly different base run which reduces the weight of the length composition data of the NAD and SAD indices by a factor of 10 (relative to the base run originally proposed by the Assessment Team). Graphs with results of the proposed new base run, including also historic time series of F and fecundity estimates relative to reference points, are shown in this report under ToR 2. This change in configuration from the Assessment Team's base run did not change the conclusion on the state of the stock.

Although there was no time during the review meeting to conduct uncertainty or sensitivity analyses from the base run proposed by the reviewers, I expect the essential features of the analyses conducted by the Assessment Team for their originally proposed base run will most likely hold for the newly proposed base run. Based on this perception, and the comments I provided under ToR 3, I consider the estimates of biomass, abundance, and exploitation rate to be reliable and consistent with input data and population biological characteristics, and 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?

From the base run originally proposed by the Assessment Team, the sensitivity runs and the uncertainty evaluations (via MCB), it is concluded that the stock is not overfished relative to the stock fecundity threshold ($FEC_{15\%}$) or the revised threshold proposed by the Assessment Team ($FEC_{20\%}$). The stock is also estimated to be above (with more than 50% probability) the original target reference point ($FEC_{30\%}$) and the revised target proposed by the Assessment Team ($FEC_{36\%}$). These conclusions are supported by Figures 8.3.2.5-8.3.2.8, 8.3.2.11, 8.3.2.12 and 8.2.2 of the stock assessment document.

Such an extensive analysis was not replicated for the new base run proposed by the reviewers. However, the fecundity time series graphs displayed under ToR 2 of this report show that fecundity estimates from this run are above $FEC_{20\%}$ during almost the entire time series and above $FEC_{39\%}$ during the last seven years. Therefore, it is again concluded that the stock is not overfished relative to the fecundity threshold and the stock is estimated to be currently above the fecundity target; this holds for the current reference points ($FEC_{15\%}$ and $FEC_{30\%}$) and the revised reference points proposed by the Assessment Team ($FEC_{20\%}$ and $FEC_{39\%}$).

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?

From the base run originally proposed by the Assessment Team, the sensitivity runs and the uncertainty evaluations (via MCB), it is concluded that the stock is not undergoing overfishing relative to the fishing mortality threshold ($F_{15\%}$) or the revised threshold proposed by the Assessment Team ($F_{20\%}$). The stock is also estimated to be below (with more than 50% probability) the original target reference point ($F_{30\%}$) and the revised target proposed by the Assessment Team ($F_{36\%}$). These conclusions are supported by Figures 8.3.1.1, 8.3.2.3, 8.3.2.4, 8.3.2.7-8.3.2.10 and 8.2.1 of the stock assessment document.

Such an extensive analysis was not replicated for the new base run proposed by the reviewers. However, the fishing mortality time series graphs displayed under ToR 2 of this report show that estimates of full F from this run are below $F_{15\%}$ during almost the entire time series and below $F_{30\%}$ continuously since the late 1990s; estimates of F at age 2 are below $F_{39\%}$ continuously since year 2000. Therefore, it is again concluded that the stock is not undergoing overfishing relative to the fishing mortality threshold and fishing mortality is estimated to be currently below the target; this holds for the current reference points ($F_{15\%}$ and $F_{30\%}$) and the revised reference points proposed by the Assessment Team ($F_{20\%}$ and $F_{39\%}$).

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

The stock assessment document notes that an attempt was made to fit a Beverton-Holt stock-recruitment relationship within the assessment, but that the steepness parameter always ended up on a bound near 1. Annual recruitment values are estimated in the assessment from log-normal distributions centered at a Beverton-Holt curve with steepness parameter fixed at 0.99 and assuming a large coefficient of variation. In practical terms, this is similar to estimating a free recruitment parameter for each year and seems like an appropriate approach. The assessment estimates an exceptionally high year class in 1958. After this, there was a period of higher recruitment from the mid-1970s to the mid-1980s, whereas recruitment has generally been lower in most other years. Recruitment in the last 20 years has been mostly below the time series average (see estimated recruitment time series graph displayed under ToR 2; the rightmost value in the graph, 2014, should be ignored). Within the range of stock abundance values experienced during the assessment period (1955-2013) there seems to be no clear sign of recruitment increases or decreases depending on stock fecundities: recruitment was high in almost every year from the mid-1970s to the mid-1980s, even though total stock fecundity experienced a broad range of values during this period (being high at the start of the period and low at the end, this being a period of increasing fishing mortality); recruitment has been lower since the late 1980s, even though stock fecundity has been at many different levels during that time, including the particularly high stock fecundity in the seven most recent years. On the other hand, the highest recruitment during the assessment years corresponds to 1958, and this was one of the years with lowest estimated stock fecundity. Given the above, estimating an informative stock-recruitment relationship does not seem possible based on the current data.

In this situation, if applying a stock-recruitment relationship is needed (for example, for Management Strategy Evaluations or for medium-term projections), I suggest using a hockey-stick functional form with breakpoint at the lowest observed total stock fecundity, or at a value a bit higher than the lowest observed fecundity for extra protection. Using a hockey-stick stock-recruitment function of this form in such analyses will provide some protection against overoptimistic recruitment assumptions at low stock abundances (such as assuming the same mean recruitment at all stock fecundity levels would likely be, since we know that recruitment will decrease if stock fecundity becomes sufficiently low).

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?

The current estimates of fishing mortality and fecundity reference points are based on 15%SPR for thresholds and 30%SPR for targets (where SPR is interpreted as fecundity-per-recruit for this stock). The stock assessment report states that these choices of %SPR values were based on the last stock assessment, when the exploitation level was thought to be around $F_{8\%}$. The 2014 benchmark stock assessment provides a very different perspective from the previous assessment, and F in 2013 is estimated to correspond to a much higher %SPR (of the order of 65%-70%). Therefore, the stock assessment document proposes that new reference points be set, to be more in line with the current understanding of how the stock has been developing during the period covered by the assessment, and with the aim of providing a measure of the sustainability of the stock and fisheries. To this end, the Assessment Team proposes that the threshold and target reference points be respectively based on the maximum and median F at age 2 estimated to have occurred during 1960-2012. The assessment document notes that these historic values of F correspond in %SPR terms to $F_{20\%}$ for the threshold and $F_{36\%}$ for the target; correspondingly, $FEC_{20\%}$ and $FEC_{36\%}$ were proposed as the stock fecundity threshold and target, respectively. These calculations were based on the results of the base run originally proposed by the Assessment Team. For the new base run proposed by the reviewers, the maximum and median F at age 2 estimated to have occurred during 1960-2012 correspond to 20% and 39% SPR, respectively; hence, the revised reference points proposed by the Assessment Team correspond to 20% and 39% SPR, for thresholds and targets, respectively.

In the absence of more clearly specified management objectives for this stock, the Assessment Team's proposal for reference points (based on historical exploitation levels, aiming to be in line with sustainability) seems sensible. Fishing mortalities around the target proposed by the Assessment Team would be expected to keep total stock fecundity and recruitment within the observed historic range with high probability (assuming no major changes happen in the ecosystem relative to the situation generally experienced during 1960-2012); since the stock has historically been able to sustain this level of fishing pressure, one would, in principle, expect that the same will likely continue to hold in future years. The fishing mortality proposed as target is likely to maintain a level of stock abundance that will give managers some flexibility in terms of possible harvest strategies; for example, some flexibility for deviating from targets (while always avoiding exceeding thresholds) in some years is likely possible, but this will depend on the chosen harvest strategies, which should preferably be evaluated according to a Management Strategy Evaluation analysis. Considerations of menhaden as a forage fish, and potential development of ecological reference points, will be addressed under ToR 8. Using values of M (i.e. values that realistically represent the amount of predation menhaden undergoes) in the assessment and reference points calculation is a starting point in this direction (although this in itself may not be sufficient if substantial increases in the biomass of predators are desired); M values would normally be expected to change over time, mostly linked to changes in predators' abundances.

Next I provide some technical comments regarding the reference points calculations.

Because selectivity-at-age on the stock as a whole (considering the combined removals of all fleets) is changing from year to year, I think the average F over a range of highly selected ages (e.g. ages 2-4) would be more representative of fishing pressure on the stock in any year, and more comparable among different years, than focusing on F at a single age. So I suggest exploring this. Note that applying the method for reference points calculation proposed by the Assessment Team but based on average F(ages 2-4) instead of F(age 2) will likely lead to %SPR values for the reference points that are somewhat different from 20% and 39% (which were obtained applying the method on F(age 2)).

The Assessment Team recommended calculation of reference points based on the average biological parameters of the period 1955-2013 (maturity, fecundity, weight, M; although M has been assumed to be time-invariant in the current assessment) and the average fishery selectivity of the last three years. I agree with the use of recent fishery selectivity for reference point calculation, as the recent situation in the fishery is more likely to be relevant for the next few years (this is assuming that stock status in relation to reference points may be used to direct, or evaluate the likely implications of, management actions, in which case I think it makes sense to calculate the reference points in such a way that they are applicable for management in the next few years). In principle, the use of a longer time series average for biological parameters also seems reasonable to me, although one would need to carefully examine if there have been clear changes in productivity or in other biological parameters that may make the use of a more recent average more relevant for the next few years than a longer term average. The assessment results indicate that recruitment has been mostly below the time series average in the last two decades and growth at present seems to be different from how it was during the 1970s and 1980s (growth is believed to be density-dependent for this stock, suggesting that linkages might be expected between recruitment strength and biological parameters). Although in principle using the longer time series for biological parameters seems sensible, it may also be useful to check how the recent productivity/biological features would impact on the reference points.

ToR 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.

No minority report has been filed.

ToR 6. Review the Technical Committee's recommendations on research, data collection, and assessment methodology and make any additional recommendations or prioritizations, if warranted.

The stock assessment document lists recommendations in Section 9, so I will not repeat them here. These recommendations were presented during the review meeting. The reviewers agreed with the recommendations overall and provided some additional comments, of which I highlight the ones I consider most salient here.

In terms of data collection, there was agreement that developing a coast-wide fishery-independent index of adult abundance at age should be a top priority. Along similar lines (but likely simpler), the reviewers felt that obtaining age compositions of the NAD and SAD composite indices would be very helpful in the estimation of selectivity-at-age for these indices. I also suggest trying to improve the length composition data for the NAD and SAD indices (especially if age composition data can not be obtained for them).

In terms of assessment methodology, there was also agreement that conducting a Management Strategy Evaluation to evaluate harvest strategies and possibilities for reference points should also have high priority. Estimating growth inside, rather than outside (as done as present), the stock assessment model was considered worth investigating (if length composition data continue to be used in future assessments), although it was also noted that this may lead to a substantial increase in model complexity. If the Assessment Team decides to pursue this work, the BAM model platform could be extended to incorporate this feature (which has the advantage of the Assessment Team already having great expertise on BAM); alternatively, a platform that already allows for this feature (such as SS3) could be used. Finally, some reservations were expressed about the recommendation to develop a seasonal spatially-explicit model incorporating movement, regarding the difficulties this approach entails (e.g. in terms of data needs and overall scientific effort) versus the added value it may offer for informing menhaden management. Since the Assessment Team presented this as a long-term recommendation, I suggest that this is again reconsidered after some years.

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

I note here main points which I think would be useful to investigate for the next assessment. Note that I think of them as points to investigate for the next benchmark assessment, not in an update assessment.

- Explore the possibility of obtaining representative age composition data for the NAD and SAD composite indices.
- Explore more appropriate ways of deriving length compositions for the NAD and SAD indices.
- If the tasks in the previous two bullet points are not possible, then explore model configurations that do not require the use of (age or length) composition data for the NAD and SAD indices. Evaluate the robustness of assessment results to alternative model configurations (such as assumptions about the indices' selectivities-at-age) that may be considered plausible.

- Assuming that deriving representative age or length compositions for NAD and SAD is possible, find ways of improving the fit of the model to their age or length compositions.
- The latter bullet point may benefit from the estimation of (time-varying) growth within the assessment model (assuming length compositions remain in the model). However, this could substantially increase model complexity. A relatively simple alternative may be to allow some flexibility (e.g. by using a constrained prior distribution centered at the values estimated outside the assessment) in the growth parameters used in the fit to the length composition data.
- Another possibility for improving the model fit could be to model the fleet and index selectivities using age-specific parameters (while assuming the same selectivity for a group of older ages) instead of pre-selecting functional forms (logistic or double-logistic in the current assessment).
- Along similar lines, fleets' selectivities could be allowed to vary from year to year in a relatively constrained manner (e.g. with some kind of constrained random walk) instead of using time blocks.
- I also suggest exploring the use of more realistic estimates of M. Given menhaden's role as a forage species, using a time-varying M (responding mainly to predation changes) would seem very appropriate (although this is of course conditional on being able to obtain realistic estimates of such a time-varying M).
- Related to the previous point, I also suggest continuing exploring the development of a multispecies model that can take the main predator-prey interactions into account (see discussion under ToR 8).
- Finally, I suggest conducting an in-depth evaluation of reference points using MSE.

ToR 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.***

Annex E of the stock assessment report describes methodological options for developing ecological reference points (ERP) for menhaden taking into account its role as a forage fish. It also presents ecosystem monitoring and modeling approaches that could be used to support ecosystem-based management more broadly. The approaches are presented in order of increasing complexity and I discuss them in the same order here. The approaches can be grouped into: ecosystem indicators, nutrition reference points, production models, single-species age-structured models, and multispecies models (see Table 1 of Annex E).

One of the difficulties often encountered with this type of task is the lack of specific management objectives, apart from some generally stated goals. Any evaluation involving multiple species or aspects of the ecosystem inevitably has to deal with identification of trade-offs and prioritization among different combinations of management objectives and outcomes. It is not a simple task for anyone, but in my opinion it is relevant to tackle this issue for menhaden, given its role as a forage fish as well as providing for important directed fisheries. Table 2 of Annex E usefully identifies types of objectives on which each of the approaches could inform. The identified potential types of objectives are: low disease prevalence, adequate nutrition levels, enough prey to support predator species at preferred biomass levels, sustainable menhaden fishery in light of forage pressure, high menhaden abundance at younger ages, broad menhaden age-structure. The methods examined in Annex E tackle exclusively biological aspects and do not directly inform about economic and social objectives that likely also exist.

Ecosystem indicators: as Annex E states, these do not directly provide quantitative reference points for the menhaden stock, but give information on several aspects of the ecosystem and may help to guide management decisions in some cases. Annex E considers environmental indicators (e.g. plankton abundance, chlorophyll concentration, SST and climate indicators linked to productivity in various parts of the ecosystem), indicators of forage fish abundance (for a suite of relevant forage species), and a range of simpler to more complex prey:predator ratio indicators. No clear methodologies for setting reference points for such indicators appear to be presently available, so reference levels will be to a certain degree subjective and based on expert judgment of what indicator levels may be considered to reflect 'desirable' ecosystem conditions; past experience may be used to inform interim reference levels.

Nutritional indicators are meant to represent nutritional status of predators and reference points should characterize 'desirable' and 'undesirable' nutritional status. Past experience and scientific judgment may be used to inform interim reference levels.

Production models: Annex E presents two ideas, both based on modelling the annual dynamics of menhaden biomass based on a Schaefer model. The first idea is a Steele-Henderson model, where explicit account is taken of fishery removals and predation by a set of pre-specified main predators; the annual biomasses of each predator are model inputs and predation is assumed to follow a Type III functional response. Menhaden fishing mortality (F) and predation mortality (M_2) arising from each predator are directly obtained from model outputs. Maximum useable production reference points for menhaden are calculated as $K/2$ for biomass and $r/2$ for fishing + predation mortality. The second idea is a Schaefer model with time-varying intrinsic growth rate (r). This model explicitly accounts for fishery removals, but not for predation removals. The time-varying r is assumed adequately to account for all substantial changes in predation pressure on menhaden without having to explicitly specify the predator field or the exact mechanism behind the changes experienced by the menhaden stock. MSY reference points for menhaden could be calculated ($K/2$ for biomass and the time-varying $r/2$ for fishing mortality).

Both ideas are interesting and can help progress towards the objective of accounting for menhaden's role as a forage fish. Both models can address the objective of a sustainable menhaden fishery in the light of forage pressure. However, if the management objective includes supporting predator species at preferred biomass levels, it seems to me that the Steele-Henderson model would be more useful; this is because it specifically uses predator biomasses as model inputs and would hence allow to inform management policies considering different balances in the prioritization between fishing mortality and predation mortality at alternative levels of predator biomasses. Of course, for such an approach to be useful the predator field and functional responses must be adequately characterized in the model, as otherwise results could be misleading. For this reason, Annex E suggests that both production models are taken forward in tandem, so that the second one can to some extent act as a cross-check of the results obtained from the first one. This seems sensible to me. One thing I would add, though, is that none of these models considers how predator abundance or biomass may change in response to changes in prey biomass; only true multispecies models can incorporate this aspect. If these feedbacks are expected to be important, then I suggest that an appropriate multispecies model should be developed (possibly in parallel to these simpler approaches).

Age-structured single-species approaches: The first approach presented is based directly on the current BAM assessment and considers the possibility of using more conservative reference points, following the principles advocated for forage fish in recent work such as Pikitch *et al.* (2012) and others. A second approach proposes trying to estimate a time-varying M (using an index of predator consumption) within a single-species statistical catch-at-age model (e.g. BAM or SS3). Another approach is a hybrid that uses multispecies MSVPA-X to estimate predation mortalities on menhaden assuming that menhaden predators (striped bass, weakfish and bluefish) were kept at agreed threshold levels and no fishing mortality was applied to menhaden. The resulting natural mortalities for menhaden are then incorporated in projections within the BAM menhaden assessment model. The values of %SPR reference points for menhaden could be recomputed based on the BAM projections with the new natural mortalities; the implication of fishing at various intensity levels could also be examined.

I agree with Annex E that the first of these approaches is best interpreted in a general sense and that it is difficult to know if it would be applicable to menhaden in this region. I would expect that development of reference points specifically addressing the ecosystem menhaden inhabits and management objectives for menhaden would be more useful. The other two approaches can, to some extent, be considered as age-based counterparts of the two approaches proposed based on production models. Some ideas related to linking predation mortality values to predator biomasses were discussed by the ICES multispecies assessment group WGSAM in their 2011 meeting (ICES, 2011; see Chapter 8) and may perhaps be useful in this context. The hybrid approach links some MSVPA-X results with BAM projections and reference point calculation, and accounts for predation mortality on menhaden when predators are at pre-specified biomass levels and assuming F=0 for menhaden. Several relevant caveats about this approach were identified in Annex E and I agree with them (one of these caveats is that direct comparability of the BAM assessment results, based on time-invariant M, with projections based on M derived from MSVPA-X may not be entirely appropriate and should be further examined; another aspect that occurs to me is that the M_2 caused by predators at the pre-specified predator biomass levels may change depending on the menhaden biomass level, and I do not think this is considered in this approach). As was the case with the two approaches based on production models, these approaches do not address either the potential effects that changes in menhaden abundance could have on the abundance of its predators. If these feedbacks are expected to be important, then I suggest that an appropriate multispecies model be used for the development of ERPs.

Multi-species and ecosystem models: Annex E mentions the development of a multispecies statistical catch-at-age model involving two prey species (Atlantic menhaden and scup) and three predators (striped bass, bluefish and weakfish). Finally, Annex E also considers the development of a fuller ecosystem model, based on EwE. I am in no doubt that both approaches could be useful and I would not want to discourage anyone from developing a particular approach. However, my feeling is that a "simple but sufficiently complex" multispecies model would be the most useful option for management purposes. Such a model should be able to account for the main interactions, so as to achieve a level of realism that makes it useful in management, but without making it any more complex than necessary to achieve this. This suggests to me pursuing the development of a multispecies statistical catch-at-age model. The specific detail of how this is done is, of course, strongly case-dependent, but I would imagine that in addition to the predation mechanisms of predators on forage fish, one may potentially need to consider predation of forage fish on predators (e.g. predation on eggs?), possible cannibalism in some species, density-dependence in growth, increased mortality and/or reduced growth for predators under food shortages, etc. Achieving a useful balance between simplicity and realism is obviously not easy, but that is what I would ultimately be aiming to do. An appropriately constructed multispecies model should help to get an understanding of the main trade-offs that may be encountered between yield and abundance of a suite of alternative species, and this should be useful to inform subsequent development of reference points in line with management priorities. It may help to be aware of the work by the ICES working group on multispecies assessment methods (WGSAM), which meets annually. This working group can be found online at http://www.ices.dk/community/groups/Pages/WGSAM.aspx

The selection of specific reference points requires agreement on the management goals and on how trade-offs will be evaluated if contradictory goals are identified. Whereas much of the ground technical work to support the development of ERPs can probably still proceed without formal agreement on a specific set of management objectives, I expect an iterative process of dialogue with managers will be needed (likely also involving stakeholders at some stage, although this will also depend on governance frameworks in the USA, with which I am not familiar). In relation to this dialogue process, the reviewers considered that defining with managers metrics that could be used to measure performance in relation to the managers' objectives would be helpful. It is suggested that scientists could present the potential for different modeling approaches, bringing out the pros and cons, and in discussion with managers (and possibly stakeholders) decide on the most appropriate way forward. Work on Management Strategy Evaluation could then proceed, using an appropriate multispecies model to determine the nature of trade-offs among potentially conflicting objectives as different management strategies (harvest policies) are examined.

ToR 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.

This has been done together with the panel's chair and the other two CIE reviewers. The final panel report will be submitted shortly.

Conclusions and recommendations (in accordance with each ToR):

My conclusions, suggestions and recommendations were incorporated in the detailed discussions provided above for each of the ToRs. In this section I highlight main aspects in bullet point form.

ToR 1: Evaluate the data used in the assessment

- Concerning the commercial fisheries data, the questions asked during the review meeting were mostly for clarification. The treatment of these data for the assessment is sound and robust. The recently improved sampling level in the bait fishery should continue.

- Whereas continuing exploration of the fisheries-independent indices during the coming years is desirable, the approach and methods followed by the Assessment Team to produce the JAI, NAD and SAD composite indices are appropriate and the results can be used for the menhaden stock assessment.

- Concerning the length compositions of the NAD and SAD indices, I recommend further future exploration of ways to combine the observed lengths from the different component surveys so as to better represent the composite indices.

- Related to the previous bullet point, for the stock assessment using the currently available length composition data, it is recommended that the length composition data be downweighted in the assessment model.

ToR 2: Evaluate the methods used to assess the stock, taking into account available data

- The methods used by the Assessment Team to conduct the menhaden assessment are sound and robust. The assessment model BAM is an appropriate choice and makes appropriate use of the available data. The model has been configured properly and used consistently with standard practices. The stock assessment document has a clear and defensible explanation of how the base run proposed by the Assessment Team was arrived at. The only concern with that base run is in relation to the length composition data of the NAD and SAD composite indices. This was explored and discussed at length during the review meeting and a run where these length composition data are downweighted by a factor 10 is proposed as the new base run.

ToR 3: Consider how uncertainties in the assessment, and their potential consequences, are addressed

- The Assessment Team has performed a very comprehensive uncertainty and sensitivity analysis, which I find very useful. The combination of evaluating uncertainties via the MCB approach, conducting sensitivity runs and a retrospective analysis, gives a very complete picture and adds significantly to the confidence in assessment results.

- I agree with the overall MCB approach for evaluating assessment uncertainty, but suggest that in the future some further thought be given to the sources of uncertainty included in the MCB simulations and, particularly, to accounting for the correlations between the estimates of different variables whenever possible.

- The implications of uncertainty on technical conclusions are clearly stated, covering assessment estimates, stock status relative to reference points and short-term projections.

- For short-term projections, in addition to showing the forecast distributions of recruitment, F and total stock fecundity, I think it could also be useful to show F and fecundity relative to the reference points. This would allow a quantification of the probability of staying within targets or exceeding thresholds during the projection period (of course, as these are model projections, results are conditional on model specifications and the sources of uncertainty incorporated in the MCB procedure).

ToR 4: Evaluate the assessment findings with respect to the following:

ToR 4a: 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?

- I consider this to be the case. I have already discussed the appropriateness of the stock assessment model and its configuration choices, and explained why the reviewers proposed a slightly different base run which reduces the weight of the length composition data of the NAD and SAD indices by a factor of 10 (relative to the base run originally proposed by the Assessment Team). This change in configuration from the Assessment Team's base run did not change the conclusion on the state of the stock.

- Although there was no time during the review meeting to conduct uncertainty or sensitivity analyses from the base run proposed by the reviewers, I expect the essential features of the very comprehensive analyses conducted by the Assessment Team for their originally proposed base run will most likely hold for the new proposed base run.

ToR 4b: 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?

- From the base run originally proposed by the Assessment Team, the sensitivity runs and the uncertainty evaluations (via MCB), it is concluded that the stock is not overfished relative to the stock fecundity threshold ($FEC_{15\%}$) or the revised threshold proposed by the Assessment Team ($FEC_{20\%}$).

The stock is also estimated to be above (with more than 50% probability) the original target reference point ($FEC_{30\%}$) and the revised target proposed by the Assessment Team ($FEC_{36\%}$).

- Although such an extensive analysis was not replicated for the new base run proposed by the reviewers, from the fecundity time series graphs displayed under ToR 2 of this report it is again concluded that the stock is not overfished relative to the fecundity threshold (be it the current threshold, $FEC_{15\%}$, or the revised threshold proposed by the Assessment Team, $FEC_{20\%}$); the stock is also estimated to be currently above the fecundity target (be it the current target, $FEC_{30\%}$, or the revised target proposed by the Assessment Team, $FEC_{39\%}$).

ToR 4c: 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?

- From the base run originally proposed by the Assessment Team, the sensitivity runs and the uncertainty evaluations (via MCB), it is concluded that the stock is not undergoing overfishing relative to the fishing mortality threshold ($F_{15\%}$) or the revised threshold proposed by the Assessment Team ($F_{20\%}$). The stock is also estimated to be below (with more than 50% probability) the original target reference point ($F_{30\%}$) and the revised target proposed by the Assessment Team ($F_{36\%}$).

- Although such an extensive analysis was not replicated for the new base run proposed by the reviewers, from the fishing mortality time series graphs displayed under ToR 2 of this report it is again concluded that the stock is not undergoing overfishing relative to the fishing mortality threshold (be it the current threshold, $F_{15\%}$, or the revised threshold proposed by the Assessment Team, $F_{20\%}$); the stock is also estimated to be currently below the target F (be it the current target, $F_{30\%}$, or the revised target proposed by the Assessment Team, $F_{39\%}$).

ToR 4d: Is there an informative stock recruitment relationship? Is the stock recruitment curve reliable and useful for evaluation of productivity and future stock conditions?

- The stock assessment document notes that an attempt was made to fit a Beverton-Holt stock-recruitment relationship within the assessment, but that the steepness parameter always ended up on a bound near 1. Annual recruitment values are estimated in the assessment from log-normal distributions centered at a Beverton-Holt curve with steepness parameter fixed at 0.99 and assuming a large coefficient of variation. In practical terms, this is similar to estimating a free recruitment parameter for each year and seems like an appropriate approach. Within the range of stock abundance values experienced during the assessment period (1955-2013) there seems to be no clear sign of recruitment increases or decreases depending on stock fecundities. In conclusion, estimating an informative stock-recruitment relationship does not seem possible based on the current data.

- If applying a stock-recruitment relationship is needed (for example, for Management Strategy Evaluations or for medium-term projections), I suggest using a hockey-stick functional form with breakpoint at the lowest observed total stock fecundity, or at a value a bit higher than the lowest observed fecundity for extra protection. Using this type of stock-recruitment function will provide some protection against overoptimistic recruitment assumptions at low stock abundances.

ToR 4e: 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 current estimates of fishing mortality and fecundity reference points are based on 15%SPR for thresholds and 30%SPR for targets (SPR is interpreted as fecundity-per-recruit for this stock). Because of the substantially revised stock perspective from the 2014 stock assessment benchmark, the stock assessment report proposes that new reference points be set, taking into account the current understanding of stock development and to be in line with sustainability; the proposal is to base the threshold and target reference points on the maximum and median F (at age 2), respectively, estimated to have occurred during 1960-2012 (which correspond to 20%SPR and 39%SPR when calculated from the new base run proposed by the reviewers). In the absence of more clearly specified management objectives for this stock, this proposal seems sensible.

- Fishing mortalities around the proposed target F would be expected to keep stock fecundity and recruitment within the observed historic range with high probability (assuming no major changes happen in the ecosystem relative to the situation generally experienced during 1960-2012).

- Concerning the calculation of the reference points, because selectivity-at-age on the stock as a whole (considering the combined removals of all four fleets) is changing from year to year, the average F over a range of highly selected ages (such as 2-4) is likely to provide a more consistent representation of fishing pressure across different years than F at a single age. Therefore, I suggest considering calculating the maximum and median historical F based on a range of highly selected ages (e.g. 2-4). Note that this will likely lead to %SPR values for the reference points somewhat different from 20% and 39%, which were obtained based on maximum and median F at age 2.

- The Assessment Team recommended calculation of reference points based on the average biological parameters for the period 1955-2013 and the average fishery selection of the last three years. I agree with the use of recent fishery selectivity for reference point calculation, as I expect the recent situation in the fishery is more likely to be relevant for the next few years. In principle, the use of a longer time series average for biological parameters also seems reasonable, although one would need to carefully examine if there have been clear changes in productivity or in other biological parameters that may make a more recent average more relevant for the next few years. In this respect, for the menhaden stock it may also be useful to check how the recent productivity/biological features might influence the reference points.

ToR 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.

No minority report has been filed.

ToR 6. Review the Technical Committee's recommendations on research, data collection, and assessment methodology and make any additional recommendations or prioritizations, if warranted.

These recommendations are listed in Section 9 of the stock assessment document and were presented during the review meeting. The reviewers agreed with the recommendations overall and provided some additional comments, of which I highlight the ones I consider most salient here.

In terms of data collection:

- There was agreement that developing a coast-wide fishery-independent index of adult abundance at age should be a top priority.

- The reviewers also noted that obtaining age compositions of the NAD and SAD composite indices would be very helpful in order to better estimate selectivity-at-age for these indices.

- I also suggest trying to improve the length composition data of NAD and SAD indices (particularly if age composition data can not be obtained for them).

In terms of assessment methodology:

- There was agreement that conducting a Management Strategy Evaluation to evaluate harvest strategies and possibilities for reference points should also have high priority.

- Estimating growth inside, rather than outside (as done as present), the stock assessment model was considered worth investigating (if length composition data continue to be used in future assessments), although it was also noted that this may lead to a substantial increase in model complexity. If the Assessment Team decides to pursue this avenue, the BAM model platform could be extended to incorporate this feature (which has the advantage of the Assessment Team already having great expertise in BAM); alternatively, a platform that already allows for this feature (such as SS3) could be used.

- Some reservations were expressed about the recommendation to develop a seasonal spatially-explicit model incorporating movement, regarding the difficulties this approach entails (e.g. in terms of data needs and overall scientific effort) versus the added value it may offer for informing menhaden management. Since the Assessment Team presented this as a long-term recommendation, I suggest that this be again reconsidered after some years.

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

I note here main points which I think would be useful to investigate for the next assessment. Note that I think of these as points to investigate for the next benchmark assessment, not in an update assessment.

- Explore the possibility of obtaining representative age composition data for the NAD and SAD composite indices.

- Explore more appropriate ways of deriving length compositions of the NAD and SAD indices.

- If the tasks in the two previous bullet points are not possible, explore model configurations that do not require the use of (age or length) composition data for the NAD and SAD indices. Evaluate the robustness of assessment results to alternative model configurations (such as assumptions about the indices' selectivities-at-age) that may be considered plausible.

- Assuming that deriving appropriate age or length compositions for NAD and SAD is possible, find ways of improving the fit of the model to their age or length compositions.

- The latter bullet point may benefit from the estimation of (time-varying) growth within the assessment model (assuming length compositions remain in the model). However, this could substantially increase model complexity. A relatively simple alternative may be to allow some flexibility (e.g. using a constrained prior distribution centered at the values estimated outside the assessment) in the growth parameters used in the fit to the length composition data.

- Another possibility for improving the model fit could be to model fleet and index selectivities according to age-specific parameters (while assuming the same selectivity for a group of older ages) instead of pre-selecting functional forms (logistic or double-logistic in the current assessment).

- Along similar lines, selectivity could be allowed to vary from year to year in a relatively constrained manner (e.g. with some kind of constrained random walk) instead of using time blocks.

- I suggest exploring the use of more realistic estimates of M. Given menhaden's role as a forage species, using a time-varying M (responding mainly to predation changes) would seem very appropriate (although this is of course conditional on being able to obtain realistic estimates of such a time-varying M).

- Related to the above, I would suggest continuing exploring the development of a multispecies model that can take the main predator-prey interactions into account (see discussion under ToR 8).

- Finally, I suggest conducting an in-depth evaluation of reference points using Management Strategy Evaluation.

ToR 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.

- Annex E describes methodological options for developing ecological reference points (ERP) for menhaden taking into account its role as a forage fish. It also presents ecosystem monitoring and modeling approaches that could be used to support ecosystem-based management more broadly. The approaches can be grouped into: ecosystem indicators, nutrition reference points, production models, single-species age-structured models, and multispecies models (see Table 1 of Annex E).

- One of the difficulties often encountered with this type of task is the lack of specific management objectives, apart from some generally stated goals. Any evaluation involving multiple species or aspects of the ecosystem inevitably has to deal with identification of trade-offs and prioritization among different combinations of objectives and outcomes. It is not a simple task for anyone, but in my opinion it is relevant to tackle this issue for menhaden, given its role as a forage fish as well as providing for important directed fisheries. Table 2 of Annex E usefully identifies types of objectives on which each of the approaches could inform.

- *Ecosystem and nutritional indicators:* as Annex E states, these do not directly provide quantitative reference points for the menhaden stock, but can be used to provide information on several aspects of the ecosystem and they may help in some cases to guide management decisions. Overall, no clear methodologies for setting reference points for such indicators appear to be presently available, so reference levels will be to a certain degree subjective and based on (scientific) expert judgment of what indicator levels may be considered to reflect 'desirable' and 'undesirable' ecosystem conditions.

- *Production models*: Annex E presents two interesting approaches that can help progress towards the objective of accounting for menhaden's role as a forage fish. Both approaches can address the objective of a sustainable menhaden fishery in the light of forage pressure. However, if the management objective includes supporting predator species at preferred biomass levels, it seems to me that the Steele-Henderson model would be more useful; this is because it specifically uses predator biomasses as model inputs and would hence allow to inform management policies considering different balances in the prioritization between fishing mortality and predation mortality at alternative levels of predator biomasses. Of course, for such an approach to be useful the predator field and functional responses must be adequately characterized in the model, as otherwise results could be misleading. For this reason, Annex E suggests that both approaches are taken forward in tandem, so that the second approach can to some extent act as a cross-check of the results obtained from the first approach. This seems sensible to me.

- *Age-structured single-species approaches*: The first approach presented is based directly on the current assessment and considers the possibility of using more conservative reference points, following the principles advocated for forage fish in recent work such as Pikitch *et al.* (2012). I agree with Annex E that these results are best interpreted in a general sense and that it is difficult to know if it would be applicable to menhaden in this region. The other approaches are specific to the menhaden situation: one aims to estimate a time-varying M (using an index of predator consumption) within a single-species menhaden assessment model; another one is a hybrid that uses multispecies MSVPA-X to calculate predation mortalities on menhaden at agreed predators' threshold levels and assuming F=0 for menhaden, and then combines these M values with projections and reference points calculations in a single-species context.

  Some ideas related to linking predation mortality values to predator biomasses were discussed by the ICES multispecies assessment group WGSAM in their 2011 meeting and may be useful in this context (see ICES 2011, Chapter 8).

  Several relevant caveats about the hybrid approach were identified in Annex E and I agree with them (one of these caveats is that direct comparability of the BAM assessment results, based on time-invariant M, with projections based on M derived from MSVPA-X may not be entirely appropriate and needs to be further examined; another aspect is that the M_2 value caused by predators at the pre-specified predator biomass levels may change depending on the menhaden biomass level, and I do not think this is considered in this approach).

- One caveat about all modeling approaches discussed so far is that none of them considers how predator abundance or biomass may change in response to changes in prey biomass; only true multispecies models can incorporate this aspect. If these feedbacks are expected to be important, then I suggest that an appropriate multispecies model should be developed (possibly in parallel to some of the simpler approaches discussed above).

- *Multi-species and ecosystem models*: I am in no doubt that both approaches could be useful and I would not want to discourage anyone from developing a particular approach. However, my feeling is that a "simple but sufficiently complex" multispecies model would be the most useful option for management purposes. Such a model should be able to account for the main interactions, so as to achieve a level of realism that makes it useful in management, but without making it any more complex than necessary to achieve this. This suggests to me pursuing the development of a multispecies statistical catch-at-age model. Achieving a useful balance between simplicity and realism is obviously not easy, but that is what I would ultimately be aiming to do. An appropriately constructed multispecies model should help in getting an understanding of the main trade-offs that may be encountered between yield and abundance of a suite of alternative species, and this should be useful to inform subsequent development of reference points in line with management priorities.

- The selection of specific reference points requires agreement on the management goals and on how trade-offs will be evaluated if contradictory goals are identified. Whereas much of the ground technical work to support the development of ERPs can probably still proceed without formal agreement on a specific set of management objectives, I expect an iterative process of dialogue with managers will be needed (likely also involving stakeholders at some stage, although this will also depend on governance frameworks in the USA, with which I am not familiar).

ToR 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.

This has been done together with the panel's chair and the other two CIE reviewers. A final panel report will be submitted shortly.

Appendix 1: Bibliography of materials provided for review

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 (corresponding Excel documents found in 'S40-4_FI Index Diagnostics' folder)
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. (Conn, P.B. 2010. Canadian Journal of Fisheries and Aquatic Sciences 67:108-120)
SEDAR40 – 9	A proposed, tested, and applied adjustment to account for bias in growth parameter estimates due to selectivity (Schueller, A.M., Williams, E.H., Cheshire, R.T. 2014. Fisheries Research, 158, 26-39)

The following documents are also cited in my report:

ICES.2011. Report of the Working Group on Multispecies Assessment Methods (WGSAM), 10–14 October 2011, Woods Hole, USA. ICES CM 2011/SSGSUE:10. 229 pp. Report available online at http://ices.dk/sites/pub/Publication%20Reports/Expert%20Group%20Report/SSGSUE/2011/WGSAM2011.pdf

Francis, R.I.C.C. 2011. Data weighting in statistical fisheries stock assessment models. Canadian Journal of Fisheries and Aquatic Sciences 68:1124-1138.

Pikitch, E., Boersma, P.D., Boyd, I.L., Conover, D.O., Cury, P., Essington, T., Heppell, S.S., Houde, E.D., Mangel, M., Pauly, D., Plagányi, É., Sainsbury, K., and Steneck, R.S. 2012. Little Fish, Big Impact: Managing a Crucial Link in Ocean Food Webs. Lenfest Ocean Program. Washington, DC. 108 pp.

Appendix 2: Copy of CIE statement of work

Attachment A: Statement of Work for Dr. Carmen Fernandez

External Independent Peer Review by the Center for Independent Experts

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 cannot 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 December 24, 2014**, 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.

<u>Tuesday</u>

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

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

<u>Wednesday</u>

8:30 a.m. – 12:00 a.m.	**Panel Discussion** **-** *Continue deliberations* *- Review additional analyses*	**Chair**
12:00 a.m. – 1:30 p.m.	**Lunch Break**	
1:30 p.m. – 3:30 p.m.	**Panel Discussion** *- Continue deliberations* *- Review additional analyses*	**Chair**
3:30 p.m. – 4:00 p.m.	**Break**	
4:00 p.m. – 6:00 p.m.	**Panel Discussion/Panel Work Session** *- Recommendations and comments*	**Chair**

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

<u>Thursday</u>

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

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

Appendix 3: Panel membership or other pertinent information from the panel review meeting

Panel members (alphabetical):

- Carmen Fernández, Spain
- Michael Jones, USA (Panel Chair)
- Anders Nielsen, Denmark
- John Simmonds, United Kingdom