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

Double Tree Hotel Beaufort
North Carolina
December 9-11, 2014

E J Simmonds

Table of Contents

1. Executive summary

The meeting to review the assessments of Atlantic menhaden was held in Beaufort, North Carolina December 9 to 11, 2014.

The input data, BAM model specification, and results for the Atlantic menhaden assessment were presented over the first day of the meeting. The data in the assessment was evaluated and found to be suitable for the assessment with the exception of the length frequency data for the NNAD index. The methodology was evaluated and found to be sound, robust, of a high standard and fully in accordance with standard practice. One minor modification to the assessment was proposed, due to the quality of the length data in the NAD survey and its influence on the assessment.

The uncertainties in the assessment, and their potential consequences, are well addressed: the sensitivity analyses support the conclusions of the assessment and show that these are robust to a number of plausible alternative model settings and biological hypotheses. The analysis of uncertainty is well developed with only two minor caveats, maintaining fixed weighting and covariance amongst variables which would lead respectively to under and overestimation of the uncertainty.

It is concluded that the estimates of biomass, abundance, and exploitation rate are reliable and consistent with the input data and population biological characteristics. The stock is seen to be not overfished relative to fecundity reference points and above target reference points. The stock is not considered to be undergoing overfishing relative to fishing mortality threshold reference points and F is considered to be below existing and proposed fishing mortality management targets.

There is no evidence for a stock recruit relationship within the historically explored range of SSB/fecundity. Alternative reference points are proposed, the limit reference points appear to be appropriate for the known dynamics of the stock. The proposed target reference points are considered acceptable but it is hard to link them directly to the management objectives given in Amendment 2 of the management plan.

Improvement in the survey data collection is needed, particularly to obtain reliable aging for the NAD survey.

The discussion of ecological reference points was considered useful. It was noted that the most important aspect for menhaden was the provision of ecosystem services in the context of menhaden as a forage fish. In this context it is suggested that following simpler methods initially would be the best way forward.

2. Background

SEDAR 40 provided compilation of data, an assessment of the stocks, and an assessment review conducted for Atlantic menhaden. The CIE peer review panel is responsible for ensuring that the best possible assessment has been provided through the SEDAR

process. The stock assessed through SEDAR 40 is 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 to Appendix 2. The agenda of the panel review meeting is attached in Annex 3 to Appendix 2 and the participant list is in Appendix 3.

Three CIE reviewers conducted an impartial and independent peer review during the SEDAR 40 review scheduled 9 to 11 December 2014, The CIE reviewers are required to have the necessary qualifications to complete the tasks in accordance with the SoW and ToRs (Appendix 2). The CIE reviewers were required to have expertise in stock assessment, statistics, fisheries science, and marine biology sufficient to complete the tasks of the peer-review.

3. Description of the reviewer's role in the review activities

I participated in all aspects of the review, paying particular attention to: the input data from surveys, which had been substantially reorganized since the previous stock assessment; and the sensitivity analyses and MC analyses.

4. Findings by ToR

The report is organized following the 8 Terms of Reference listed in Annex 2 to Appendix 2. Section 4.1 deals with the input data, Sections 4.2 and 4.3 the assessment and uncertainty, Section 4.4 the state of the stock and reference points, Section 4.6 and 4.7 research and modelling recommendations. Finally Section 4.8 provides a brief discussion of ecological reference points. The agenda for the meeting and the list of participants who attended the review are given in Appendix 3.

4.1. Evaluation of the data used in the assessment

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 input data for the assessment are discussed in detail, below and the conclusions to the questions posed in ToR 1 are provided together at the end of the this section.

Three main sources of data are used for the menhaden assessment:

> commercial catch data assembled as 4 fleets, reduction fishery north and south, bait fishery north and south, including both landings and proportions at age data.
>
> fishery-independent surveys carried out mostly by individual States and used to provide three composite indices: Juvenile Abundance Index (JAI) young of the year representing age 0, and two 1+ composite indices one for the south, Southern Adult 1+ Index (SAD), based on two surveys and one for the north, Northern Adult 1+ Index (NAD), based on seven surveys. Abundance index values and length frequency distributions (LFDs) are available for the SAD and NAD composite indices.
>
> Historic tag data from a program run in the 1970s.

These data were explained in detail in presentations by the Assessment Team (AT).

The catch data are assembled in four fisheries based on area and type of fishery. The methods for assembling the data were well explained in the assessment report. The major (reduction) fishery is well sampled and the age data appears to be sufficient to allow allocation of catch to age and to obtain good estimates of fishery selection at age in the assessment model. The smaller bait fishery historically appears to be less well sampled but sampling of this is considered sufficient as the historic contribution the total catch is small.

Trends in the magnitude of the two types of fishery have been relatively consistent over the last two decades, with closure of reduction factories and increases in the bait fishery. Only one reduction factory remains, and the view is that it is unlikely that new factories will open in the near future. Therefore for the future there is a need for more age samples in the bait fishery, and this has been recognized, and the AT are encouraged to arrange better sampling to improve data collection for this bait fishery. The importance of the bait fishery while still currently less than the reduction fishery may be expected maintain its current contribution or increase in its relative magnitude in the future.

Length and age data from the fisheries are used to estimate growth rates that are then used within the Beaufort Assessment (BAM) model to provide a numbers at age and length for estimation of total fecundity and also for use in the estimation of length based selectivity in the fishery independent surveys (see below). The growth estimation method includes a bias correction for fish that are not sampled fully in the fishery. This bias depends on the estimated selectivity in the fishery, which is an output from the assessment. The currently estimated fishery selectivity is relatively stable so the bias correction is thus relatively stable, but it would be preferable to consider inclusion of the growth model within the assessment to better integrate the model.

The three survey indices are derived from a subset of State surveys using a hierarchical weighting process to give composite indices (Conn 2010). The methodology appears to be appropriate. The methodology does not explicitly account for the spatial diversity observed in the menhaden stock, but simulation testing suggests that it is robust to the type of temporal and spatial heterogeneity that is considered to be present in the survey data. The NAD survey exhibits greater heterogeneity than the SAD and deserves some further consideration (see below).

A very substantial part of the discussions that took place during the review concerned the survey indices and their LFDs. A summary of the main points in the discussion follows:

The LFD data in the composite were assembled by treating each sampled fish equally. This appears to be a reasonable assumption for the JAI and the SAD indices, either because in the case of the JAI the size range is selected to give an age 0 index, so the size distribution is less important, and for the SAD the weighting of the two survey time series in the composite index are quite similar, so the weighting is more coherent with the survey weights. However, for the NAD survey this assumption may be less robust due to the increase in heterogeneity in LFD among the surveys and the wider range of weighting factors among the survey time series (see below).

To evaluate the extent of the issues regarding LFDs the review the Panel requested information on the standardization procedures applied to the separate indices before they were combined into a composite index. The main aim was to first ensure that the variables used for standardizing the indices were definitely representative of catchability and not of abundance and secondly to consider the weights applied to the individual surveys to make the composite indices. The AT prepared a presentation explaining the Data Working Group decision tree in terms of the criteria used for potential inclusion of a survey and the subsequent standardization method applied to the survey index. The Data Working Group decision protocol emphasized that only variables that might affect catchability (not abundance) should be considered in the standardization. Graphs of all surveys used to form the YOY, SAD and NAD indices were presented before and after standardization; the changes due to standardization were, with some minor exceptions, not substantial. Changes to NY Peconic Bay Trawl in the last 5 years, VIMS Trawl Survey in the early 7 years and the GA Trawl Survey throughout the time period are the only major changes. In the future the DW might like to explore the sensitivity of these more extreme changes to the underlying assumptions, if they are driven by individual data points in those surveys, and the prediction intervals on the correction applied. It was noted that an investigation of between index correlations was generally supportive, showing mostly positive correlations among geographically closely linked surveys. It was also noted that positive correlation is not a necessary condition for the inclusion of a data set, as migration from one survey area into the next survey would imply negative correlation. Some of the simulations tested in Conn (2010) show this feature. Overall the

approach was judged satisfactory; and that the survey selection protocol was sensible and the index values could be used in the assessment.

In order to get a better understanding of how Conn's procedure (Conn 2010) combines the separate surveys into a composite index and how this would compare to combining the surveys with weights based on their areal extent. The Panel requested information on the weights that different surveys had received in the composite YOY, SAD and NAD indices. The AT produced these values (as averages over time). For SAD, the two surveys in that composite index received very similar weights. For NAD, the largest weight was on VIMS and the spread among the weights of the component surveys close to a factor 10. For YOY, there were differences between the weights of the component surveys, but not as big as in NAD. Some concern was noted regarding the spread of weights for the NAD surveys. An approach based on areal extent of the component surveys in the JAI was presented by the AT and used in the sensitivity tests. An attempt was made to produce a similar areal based composite index for the NAD index but there was insufficient time during the review meeting to do this with sufficient certainty and this approach was not pursued further. The AT might like to examine this approach as a further sensitivity test for the future. There was some concern that the quick method applied to create this area based composite had not correctly accounted for the period with one survey relative to the period with several surveys.

Part of the sensitivity analysis for the assessment model included replacement of the JAI index based on the Conn method with an area weighted alternative using the same survey data set. The CVs of the index by year was compared and the Conn weighted index appeared to have substantially reduced CVs relative to the area weighted index. When included in the assessment this option had reduced influence on the assessment and the option was not selected for the base run. This decision is endorsed.

It was concluded that overall the procedures used to produce the JAI, SAD and NAD composite indices were appropriate and it was acceptable to use these in the assessment model.

The Panel also explored the length compositions (LFDs) used for the SAD and NAD composite indices. This was motivated by both the diversity of mean length compositions among the individual State surveys (Figure 4.1 below) and a misfit in the assessment results between the observed and model-predicted LFDs (Figures 7.1.18 and 7.1.19 in the assessment report document). The AT additionally presented a plot of the LFDs of the 7 separate indices that go into NAD (averaged over the years available for each of the indices), which indicated substantial heterogeneity between the LFDs of different component surveys (Figure 4.1). The AT also explained that the LFDs of the composite index had been formed by direct combination of the lengths observed in the component

surveys without applying any weighting either to the within survey catches or the Conn survey weights. The Panel requested exploration of several alternative model assumptions in relation to these LFDs due to: the less than ideal method for assembling the LFDs; the observed heterogeneity in the length compositions of the component surveys; the fact that they were assembled with equal weight per sampled fish whereas the weights in the Conn's method were very different from uniform for NAD survey; and the misfit observed in the assessment model results to the length compositions of NAD and SAD (with the potential impact on other aspects of the assessment results, e.g. population abundance estimates). The diverse weighting among surveys in the NAD implies a similar diversity in contribution to abundance in the area. Applying similar abundance weighting to the survey LFDs would provide a match between abundance and length. Following testing of several model options (see assessment section) the Panel recommended that the LFDs of NAD and SAD should be down-weighted in the stock assessment (with respect to the weights selected in the base run proposed by the AT; more details later in this report).

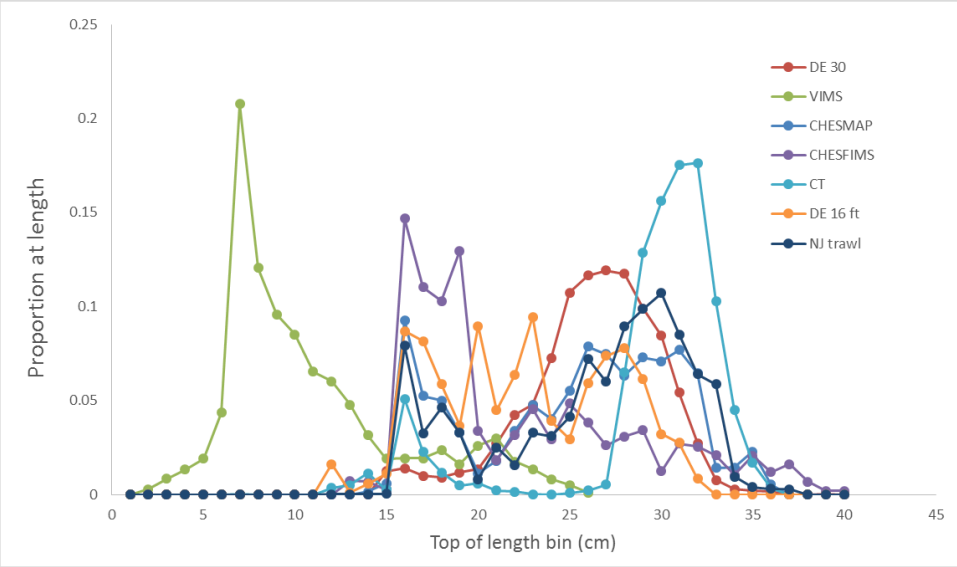

Fig 4.1 Average length compositions by survey for those surveys included in the NAD composite survey, LFD combined over years. (The lower mode in VIMS was allocated to the YoY composite survey and not included in the NAD)

Given the importance of the NAD and SAD indices in future models it would seem worthwhile to further explore the use of the individual surveys to give coast-wide indices of abundance at age. The approach by Conn (2010) is primarily aimed at dealing with multiple indices of abundance all assumed to represent a single stock. The method accounts for series with different years of observations. The method has been tested to be

robust for surveys such as those encountered here, i.e. non overlapping surveys which cover only part of a population. The method is expected to provide a combined index based on the statistical properties of the individual surveys. For the SAD index with only two surveys with almost equal weights the method can be expected to be robust. For the JAI, again the small range of survey weights suggests a reasonably robust overall index. There is some concern for the NAD which has a wider range of weighting, raising some doubts regarding the efficiency of the method. In the presentation of the assessment data mention is made of a GAM method for combining the indices, but details were not provided. The objective of the index is to characterize a part of the area in terms of abundance at age or length. The constituent parts are individual surveys that could be expected to give average catch rates for each surveyed sub-area. If all the area was surveyed then a suitable index would be average survey catch rates raised to the area covered by the survey. Given that the available data currently used are based on trawl surveys (ignoring pound nets and seines) and are, in effect, a subset of this potential coast-wide survey, then the individual area weighted surveys can be considered observations of a part of a total index. The unknown index could be considered as a model index estimated from the parts of the whole under some constraint, such as the proportion of the population in each survey area being consistent over time. Such an approach could be modified to account for different sampling variance. This approach differs from Conn (2010) in that it explicitly considers that the contributions should be weighted by area, and the contributions depend on the relative abundance in each survey area. Other aspects are similar, the method is tolerant of missing surveys. This approach has been used to combine multiple larvae surveys first by Patterson and Beveridge (1995) and refined with a more appropriate error distribution by Gröger et al (2000). If the trawl surveys used in the NAD are sufficiently similar it may be worth exploring this area based method as an alternative.

In summary the following conclusions regarding the questions in ToR 1 are:

ToR 1 a. Are data decisions made during the DW and AW justified (i.e. sound and robust)?

For the commercial data the methods used to assemble the catch and age composition data are justified.

For the survey data, it is agreed overall, but there are some concerns regarding the composite NAD index and in the length compositions of NAD and SAD indices.

b. Are input data series reliable and sufficient to support the assessment approach and findings?

Taken as a whole, the data used in the assessment are considered reliable and sufficient to support the assessment approach and findings when used with appropriate weight.

c. Are data applied properly within the assessment?

For the reasons discussed above, it was recommended that the length frequency distributions of NAD and SAD be down-weighted in the stock assessment (with respect to the weights selected in the base run proposed by the AT; more details later in this report), other data are considered weighted appropriately.

d. Are data uncertainties acknowledged, reported, and within normal or expected levels?

The report is considered to provide appropriate information on uncertainties in the input data and, with the exception of the survey length compositions for the composite adult surveys, these are within normal expectations.

4.2. 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.

These ToR are discussed together.

The stock assessment model used for Atlantic menhaden is the Beaufort Assessment Model (BAM). BAM is a statistical catch at age model, a model type used commonly for many statistical fish stock assessments worldwide. BAM has previously been used in many SEDAR assessments (e.g., Cobia, Gulf menhaden, and red grouper). The configuration of BAM was set up to match the available data for Atlantic menhaden. During the review the AT demonstrated a high level of competence with the modelling software as they were able to modify the model configuration. The BAM for Atlantic menhaden is thoroughly documented both in mathematical terms and by openly sharing its source code (appendix C of the assessment report). This allows for review at the most detailed level if desired. Taken together this gives confidence in the model being scientifically sound, robust, and appropriate for the available data.

The predicted removals from the four fleets closely match the observed, which is expected, as the model is specifically configured with fixed low CV for error in observed catch by fleet. This model assumption is reasonable for situations where catch is well monitored in magnitude, which appears to be the case here. In addition to the catch abundance data age proportions based on scale reading are used to give catch at age data. There has been limited verification of age reading, though this analysis is supportive of relatively reliable ageing. However, a more comprehensive age checking

program/workshop might be helpful along with development of aging for the surveys (see section 4.1). The quality of age data is supported by the predicted population age compositions which appear to capture most of the main features in the observed age compositions derived from the catch data. (Figures. 7.1.5-8 in the assessment report).

The model is configured with dome shaped selectivity for the four fisheries, the JAI and the SAD indices, only the NAD index is assumed to have increasing selection with length (age) asymptotic at the largest lengths or oldest ages. Given the observed spatial distribution of the length distributions, with the largest oldest fish found in the northern areas this appears to be a reasonable assumption for the three surveys. The use of domed selectivity in the fisheries which are not spatially homogeneous is well supported by the cited references (e.g., Sampson and Scott 2011).

A minor indication of a shift in selectivity is observed at age 2 around the year 2003 for the northern bait fleet (Figure. 7.1.11). It is also noted that residuals in the last years look large relative to earlier years in 3 out of the four fleets, which suggests the selection assumptions need to be monitored carefully in future years. Overall the similarity of the fitted catch at age selection across the time blocks suggests determining selection in the historic fishery is not a critical issue for the assessment.

Predicted abundance indices for the three combined surveys (JAI, NAD, and SAD) are in agreement with the observed (7.1.13-15 in the assessment report). The age of the YoY index (JAI) is derived from a length cutoff, implemented at different length throughout the year, this appears to be a good method for giving an age index. However, no age data are available for the NAD and SAD indices and the model was not able to predict well the length compositions of the two adult combined surveys (7.1.16-17 in the assessment report). These length compositions are mainly required to inform the age specific selectivity used to predict the abundance indices from the two adult surveys, this age specific selectivity may be approximated even when the detailed length compositions are not well matched. However, the length compositions also influence the whole assessment, this is discussed further below. The different components of the likelihood were weighted according to the approach of Francis (2011). This method is generally considered to be appropriate, but it does imply a slightly arbitrary weighting between those parts of the likelihood based on log normal errors, where the variance is used and those based on multinomial components where the number of samples is used more explicitly in the likelihood. However, for the final model selected to give the state of the stock this may not be of great importance as the multinomial component is further down-weighted (see below).

The overall conclusion is that the Base model in BAM as developed by the AT was configured properly and used consistent with standard practices, however, one concern remained; the mismatch between model predicted and observed length compositions in

the SAD and NAD surveys. Examination of the input data had also raised concerns that the length compositions may not be well specified (see section 4.1).

This problematic mismatch between predicted and observed length compositions was further investigated by the review panel and assessment team. Although the primary purpose of the length compositions was to allow estimation of survey selectivity at age, there was concern that the misfit was possibly biasing other estimated quantities from the model (e.g., stock numbers). The Base model assessment showed a wide CV in length at age, which appeared to be smearing length at age in the model. There was concern that the wide CV was the result of the model attempting to fit the poor length composition data in the NAD and SAD surveys. A sensitivity run was requested where the CV around the growth function was set to half of its estimated value, the results showed that the fit between observed and estimated length compensation changed, but that also that stock sizes were influenced, raising concern that the stock numbers were indeed dependent on the length composition data in the adult surveys. Several runs were tested to explore possibilities:

1 Remove length compositions, and fix NAD/SAD selection
2 Remove length compositions, and estimate NAD/SAD selection
3 Down-weight length comps by a factor of 10, estimate NAD/SAD selection
4 Down-weight length compositions by 20, estimate NAD/SAD selection

Runs 2 and 4 failed to converge, run 1 and 3 converged and gave similar estimates of year classes, fecundity and F. Based on these results the best option appeared to be down-weight the length comps, taking some guidance from the length in the NAD and SAD but limiting its influence on the numbers at age. Down weighting by a factor of 10 was pragmatically chosen as the new base run. Changes in year class strength are observed between the base model and the revised model. There is a slight increase in variability in recruitment, which fits with the reduction of smearing across length classes and reduced influence of the survey length compositions. Overall the change in estimated fecundity and F in the recent years is minor. Such a change with only minor consequences increases the perception of a robust assessment. Although the change to the model and the conclusions on the state of the stock were minor, it was considered that this model formulation was preferable for the future, so should be recommended.

It is recommended that future versions of the model maintain the down-weighting, do not use the length composition data at all, or find better ways to determine them from the survey data. One possibility to explore is to use the individual surveys in the assessment (instead of the combined index). Using the individual surveys would allow individual selection curves, which would better reflect that these surveys are coming from different areas and times, however, it will also require more parameters to be estimated. At this stage it is not possible to select a preferred method, either in the survey data preparation

or in the model, however, from the age/length composition in the surveys it would seem to be an area worth pursuing.

The assessment report indicated that a stock synthesis (SS3) model was configured for the stock, but the model was not presented at the review meeting. The assessment team had determined that seasonality had to be included first. It is possible that this model may have utility for estimating the growth internally but may have less flexibility given the local expertise with the BAM model. As the model was not presented no further comment can be made.

The AT provided an extensive sensitivity analysis of the model. With the exception of sensitivity to alternate assumptions on natural mortality, the conclusions on current F and fecundity are very robust and give a high degree of confidence in the conclusions of the assessment. Where different M values are tested these should also influence the reference points, and it might be better to consider change in M with respect to reference points rather than in absolute terms. It is expected that in this case the influence of M will be seen to be less important than that illustrated in Figure 7.4.1.57 in the assessment report.

In addition to natural mortality, the assessment was also sensitive to leaving out the NAD, however, exclusion of this index did not change the conclusion to overfished or overfishing status.

Conclusions to stock assessment method

a) Are methods scientifically sound and robust?

   The methods are sound and robust given the suggested minor change to the baseline model

b) Are assessment models configured properly and used consistent with standard practices?

   The methods used are consistent with standard good practice.

c) Are the methods appropriate for the available data?

   The methods are fully consistent with the type of data available, though alternative ways to use the survey length data should be investigated.

d) If multiple models or model configurations were considered, evaluate the explanation of any differences in results and justification of a base model.

   The configuration of the base is well described and the arguments well supported with sensitivity analyses, a minor change and alternative model was selected as the final base model.

4.3. Consideration of uncertainties in the assessment.

ToR

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 assessment team has put a lot of effort into investigating the uncertainties. In addition to the sensitivity analyses discussed above, which are used to evaluate model robustness, uncertainty is explicitly considered through bootstrapping.

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

Instead the assessment team used a parametric Monte Carlo Bootstrap (MCB) method, where the data and fixed inputs (including natural mortality) were sampled using relevant uncertainties. For each of 1000 complete sampled sets of input the model was re-estimated, which results in a simulated distribution of all estimated quantities. This approach correctly propagates the uncertainty through the nonlinear model equations to the quantities of interest.

Two minor concerns about the implementation details of the approach were raised.

1) The assigned weights of the likelihood components were set fixed in the model and were kept fixed and not part of the sampling.

2) All quantities were sampled independently. For instance the two parameters of a logistic function were each simulated uniformly from their 95% interval. If these model parameters are negatively correlated, then sampling them independently will result in unlikely pairs (and hence unlikely logistic curves). It is suggested to use the joint distribution where available.

These two aspects are most likely to result in opposite changes to the estimates of uncertainty, the first has a potential to underestimate the second to overestimate the uncertainty.

In addition to the MCB method the assessment team had prepared a wide range of sensitivity runs. These included: leaving out entire data sources, including ageing uncertainties, changing an index calculation method, different assumptions about natural mortality, different weighting of likelihood components, and different time varying

assumptions. The assessment results were seen to be robust to most alternatives, and to react as expected to others (fig. 7.4.1.1-77). The results were most sensitive to changes in assumed natural mortalities and omitting the NAD index.

Finally a retrospective analysis was presented, where the last 1, 2, 3, or 4 years of data were left out to demonstrate that the final years estimates are not severely biased. For the estimates of fishing mortality and biomass a minor systematic retrospective bias is seen, in recruitment and biomass, but it is small and only supported by four points. The analysis suggests a progression of changes in survey Q which is consistent with a retrospective error. Overall this error is small.

Overall it is concluded that the method used to evaluate uncertainty reflects and captures the most significant sources of uncertainty in the population, data sources, and assessment methods, and that the implications of uncertainty on technical conclusions are clearly stated (Figure. 8.3.2.1-12).

4.4. 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?

The five separate aspects of this ToR are dealt with separately in this section.

Estimates of biomass, abundance, and exploitation rate

ToR4.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?

The review group considers that the assessment provides reliable estimates of biomass, abundance and exploitation rates. A range of sensitivity analyses support the view that the results are robust to a range of plausible alternative assumptions. A major sensitivity is in the estimated recruitment and biomass when uncertainty in M is considered. However, this sensitivity is to be expected and is comparable to other assessments.

The review group paid particular attention to the sensitivity of the assessment to the newly derived fishery independent indices (NAD, SAD and JAI). The sensitivity analyses in the assessment report provide a good indication of the sensitivity of the assessment to inclusion or exclusion of each of these indices. The assessment showed some sensitivity to the NAD index (likely related to the fact that it is the only dataset with asymptotic selectivity in the model). It was noted that the base run proposed by the AT fitted rather poorly to the length composition data. It was concluded that the length compositions for these indices were not representative of the populations in the total area represented by these abundance indices (see section 4.1 for a discussion of the input data). Following exploration of a number of different configurations of the BAM model (See Section 4.2), it was concluded that running the model without including the length composition data was preferable. However, problems were encountered with model convergence if all the length composition data were removed; therefore, an alternative parameterization with down-weighted length compositions for the NAD and SAD indices was selected as an agreed baseline assessment. This change in configuration from the AT base run did not change the conclusion on the state of the stock.

Stock Status

ToR4.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?

Based on the reviewed agreed BAM assessment baseline run, the sensitivity runs, and estimated uncertainty in the assessment (Monte Carlo bootstrap runs), the review group agrees with the AT's conclusion that the stock is not overfished relative to either the original biomass threshold reference point ($FEC_{15\%}$) or the revised biomass threshold reference point proposed by the AT ($FEC_{20\%}$). The stock is also estimated to be above (with more than 50% probability) the original target reference point ($FEC_{30\%}$) and the revised target point proposed by the AT ($FEC_{36\%}$).

Exploitation status

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?

Based on the reviewed agreed BAM assessment baseline run, the sensitivity runs, and estimated uncertainty in the assessment (Monte Carlo bootstrap runs), the review group

agrees with the AT's conclusion that the stock is not undergoing overfishing relative to either the original fishing mortality threshold reference point ($F_{15\%}$) or the revised fishing mortality threshold reference point proposed by the AT ($F_{20\%}$). The stock is also estimated to be below (with more than 50% probability) the original target reference point ($F_{30\%}$) and the revised target point proposed by the AT ($F_{36\%}$).

Stock Recruitment Relationship

ToR4.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 SAS stated that they tried to fit a Beverton-Holt stock-recruitment curve; however, the steepness parameter always ended up on a bound near 1.0. Given the interim reference points, the AT decided to fix the steepness value at 0.99, which allowed for the estimation of a median recruitment and estimated annual deviations. A sensitivity analysis examined sensitivity of the state of the stock to shallower slope S-R relationships and concluded that the state of the stock was not influenced by this decision. This conclusion is supported by the review group. Given that there are no indications of reduced recruitment at either low or high biomass (within the observed historic range), the use of median recruitment is reasonable.

The use of an S-R relationship is particularly important in the context of conducting Management Strategy Evaluations (MSE), where simply assuming median recruitment with process error independent of biomass would not be a precautionary approach. Within an MSE, consideration might be given to the use of a hockey-stick S-R function with a breakpoint at the lowest observed fecundity. While not biologically realistic in all aspects, such an approach has the advantage of assuming a conservative slope to the origin and no dependence of recruitment on fecundity at higher biomass.

Reference points

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 Assessment Report states that 'the TC does not recommend that the current, interim SPR-based overfishing and overfished definitions continue to be used for management.'

The TC recommends that the Atlantic menhaden Management Board adopt SPR reference points based on the maximum *F* value experienced at age-2 during the 1960-2012 time period as the threshold and the median *F* value experienced at age-2 during the 1960-2012 time period as the target, along with the associated FEC values.

The following observations on the choice of single species reference points are provided:-

The use of an age-2 metric for the fishing mortality may not be a good choice. Although the assessment uses fixed selection for the recent period for each of the four fishing fleets, the distribution of catch among these fleets has changed in recent years and may be expected to change into the future due to a number of features. The bait fishery has increased and the reduction fishery declined in recent years. If the shift towards the bait

fishery were to continue, this would result in further changes in selection across the combined fishery. Additionally, although the TAC allocations between States may be relatively fixed, different States fish different combinations at age, implying different partial Fs at age and variation in the distribution of F across ages among years. It is therefore recommended that mean F over several ages, such as ages 2-4, may be a better metric to represent fishing pressure under changing selectivity. The reviewers note that application of the method proposed by the AT to derive the new reference points based of mean F(ages 2-4) instead of F(age 2) will likely lead to %SPR values different from $F_{20\%}$ and $F_{36\%}$ (which were obtained based on F(age 2)).

The AT recommend calculation of reference points based on the changes in the stock in the time period 1960 to 2012, average biological parameters for the period 1955-2013 and average fishery selection based on the last three years.

The use of recent fishery selectivity for reference point calculation is well founded, based on the perception of the long term trend in the selectivity of the fishery (when considering the total fishery on the stock, i.e. all fleets combined). It is evident both from the assessment output that changes have occurred and the perception that these changes are unlikely to be reversed seems reasonable. The exact choice of period for selectivity (3 years or some other recent period) does not currently appear to be critical given the selectivity assumptions in this model (selectivity-at-age is fairly constant since about 2006).

The use of long term biological data to evaluate reference points is appropriate and the removal of the few years at the start of the series seems reasonable given the uncertainty in the model and the very different exploitation rates that the model gives in the first years. The use of the full time series appears to be an appropriate choice for limit (threshold) reference points, where the objective is to obtain estimates of exploitation rates and fecundities to avoid. However, recruitment during the last 20 years has mostly been below average and the growth is currently different from that observed in the middle of the time series. When considering target reference points which are applicable for use in the near future (10 years) it might be useful to check how the recent productivity might influence the biomass/fecundity related target reference points. It might be more reasonable to use recent recruitment and growth when considering target reference points.

The AT have proposed reference points related to historic exploitation levels (since 1960), rather than any other criteria. In the absence of any specific alternative agreed approaches for this stock the review group considers this is reasonable.

Some information on management goals is given in Amendment 2 to the Interstate Fisheries Management Plan for Atlantic menhaden (2012), which states that the goal 'is to manage the Atlantic menhaden fishery in a manner that is biologically, economically, socially and ecologically sound, while protecting the resource and those who benefit from it. When fully implemented, the Amendment is designed to minimize the chance of a population decline due to overfishing, reduce the risk of recruitment failure, reduce impacts to species which are ecologically dependent on Atlantic menhaden, and minimize adverse effects on participants in the fishery.'

In the context of these objectives, if fishing mortality is around the proposed target reference point it can be expected that the stock will remain above the lowest observed biomass of the historic series with high probability; this satisfies the requirement to 'reduce the risk of recruitment failure' due to depleted biomass. This is because there is no indication from the observed historic population dynamics that recruitment has been lower due to lower fecundity. Fishing around the proposed target F would not be expected to lead to 'overfishing' with respect to the historic fishery. The other objectives are more difficult to define. The fishing mortality proposed as a target is likely to maintain a stock that will give managers some flexibility for minimising 'adverse effect on the participants in the fishery', though it is unclear if an alternative MSY-based reference point would be more useful, basing the exploitation in relation to MSY criteria, either at F_{MSY} for maximum yield, or below F_{MSY} for increases economic yield or greater likelihood of flexibility for managers to obtain reduced variability in catch. Any F target based approach is expected to produce variable TACs between years, given the expected variability in recruitment and the relatively high natural mortality associated with a forage fish such as menhaden. The 'impacts to species which are ecologically dependent on Atlantic menhaden' would appear to be provided for at current growth rates for current predators abundance at their current biomass levels, by the use of current values of M in the assessment and reference points. It is possible that current abundance of menhaden is not sufficient for unrestricted predator growth, if this is the case F may need to be reduced if additional ecosystem services are identified. However, should the biomass of predators change, consideration of different natural mortality rates may be needed. In this context, reference points for a forage species might be expected to change over time. For a discussion of other issues in setting reference points in a multispecies context see also the section on ecological-based reference points (ERPs).

4.5. Minority report

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 is needed as the Panel was able to reach consensus on the assessment and outputs. Any suggestions for development of the assessment or suggestion alternative analyses is included in the assessment section and Section 4.8

4.6. Review the Technical Committee's recommendations on research

The TC developed a set of "Research and Modeling Recommendations" that were categorized by time frame (short versus long term) and research type (data collection versus assessment methodology). In general the TC's recommendations can be supported. There was strong agreement that developing a coast-wide fishery-independent index of abundance is the top priority for data collection. Related to this was a suggestion that collection of age composition data for the existing fishery independent surveys should also be a high priority. The AT noted that this was reflected in the existing recommendations under item 1 in the short-term data priorities: "work with industry and states to collect age structure data and biological data outside the range of the fishery". Given the challenges, discussed elsewhere in this report, of using index length-frequency data to inform index gear selectivity in the model, it is concluded that to have direct estimates of survey age composition would be a very valuable addition to the assessment data.

It is suggested that conducting a Management Strategy Evaluation to evaluate the performance of alternative harvest strategies should be a high priority for the immediate future. Ideally the MSE should be informed by a structured Decision Analysis process (also listed as a research recommendation) that would both inform the MSE with respect to management objectives and options, and provide an opportunity for the MSE to be transparent for both stakeholders and decision makers. This should be developed with managers as a dialog process, discussing initial objectives and the metrics that could be used to check these; developing initial MSE approaches to illustrate the main issues; and discussing these with managers to further refine the objectives and metrics before presenting overall conclusions. Managers often need to see potential results before they can understand the options and refine their questions, so this iterative dialog, which can sometimes be time consuming, is necessary to draw out the critical issues and increase confidence in the process.

The panel expressed some reservations about the recommendations to "develop an integrated length and age based model" and to "develop a seasonal spatially-explicit model, once sufficient age-specific data on movement rates of menhaden are available". If the AT pursues the former, the panel suggested considering adapting the BAM to integrate length and age, rather than using a different modeling framework such as SS3. The panel cited previous experience with numerous challenges associated with developing spatial assessment models that explicitly incorporate movement, implying that the benefits (in terms of informing menhaden management) of pursuing this modeling strategy might not outweigh the costs (in terms of scientific effort). For Atlantic menhaden, the current approach suggests that migration northwards is age/size dependent. The estimates of abundance from the individual surveys suggest strongly that

this can vary significantly by year, making it hard to parameterize this robustly. While the historic fishery covered a wider area and season, the current fishery is temporally and spatially more restricted, so use of this to help to parameterize a spatially explicit seasonal model is likely to be more difficult than it would have been in the past. Increasing the coverage of the bait fishery would help, but this is probably needed just to ensure the current model is monitoring catch fully.

4.7. Guidance improvements in data or modeling for the next assessment

It was noted that two aspects of modeling are currently conducted in advance of the main BAM model.

- Growth model (if required in future versions of the model)
- Scaling of natural mortality based on tag data.

It is recommended that the Assessment group should investigate the potential for including these aspects as part of the assessment. The model currently uses growth / length as a basis for several aspects of the model: selectivity and fecundity. It may be possible to estimate selectivity at age using age data for the NAD and SAD survey indices, making the length estimation of less importance. If selection cannot be estimated through age information direct integration of estimation of growth could be included in the model. Length data are also used via the fecundity relationship to give total stock fecundity. This currently uses mean length at age, which has some bias associated with it, particularly with calculation of reference points, As growth changes over time it is important to maintain the model response to changes in growth over time and the use of length does provide this, however, use of weight at age may also provide this conversion and the relationship between mean weight and mean fecundity may have less potential for bias.

Currently the tag data are used externally from the BAM model to give guidance on natural mortality. This approach needs fleet selection and fishing mortality data, which are estimated using the BAM model. If fixed mortalities are to be considered as part of future modelling, consideration should be given to the inclusion of estimation of natural mortality using the tag data that are currently used externally. This would help to integrate the estimation process.

4.8. Feedback on the proposed ecological reference points

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.

The report (Appendix E to the assessment report) provides a discussion of a variety of aspects to be considered in the development of ecological reference points from phytoplankton upwards through the trophic levels. This analysis lays out some of the issues well but does not draw out what should be the next steps.

There is clearly a current need for ecological considerations in menhaden management, and despite the absence of specific objectives it is clear that there are management issues that require consideration of the "role" of menhaden in the food web/ecosystem. Some of the management objectives have already been discussed above in section 4.4e, and these are based on the guidance that is given in Amendment 2 to the Interstate Fisheries Management Plan for Atlantic menhaden (2012). This guidance consists mostly of single species objectives but they do include the requirement to "reduce impacts to species which are ecologically dependent on Atlantic menhaden". In this context the extent of the reduction is not indicated in the text, though the other objectives in Amendment 2 suggest that continuation of the menhaden fishery is also a requirement, implying reduction of human impacts to zero is not intended. This therefore implies part of the productivity of menhaden should be expected to be available for predators, and this should be accounted for in management.

Broad ecosystem indicators related to environmental conditions, even though those that may be experienced by menhaden, might be of interest for tracking important changes. However, it will be difficult to translate such general environmental indicators into reference points that could trigger or inform management actions. Species such as menhaden will be responsive to environmental change in terms of reproductive success, growth, mortality and distribution, but while prediction of these dependencies for some stocks has been successful for short periods it has proved difficult to give predictions that sustain over longer time, as the relationships break down. Observed relationships in the recent past often break down after a few years (Myers 1998). Relationships between temperature (or upwelling) and recruitment have shown some success but the interactions are often more complex than simple indicators can capture and the potential for breakdown is considerable. It seems unlikely that broad environmental reference points will have much utility for reference points for Atlantic menhaden.

Table 2 of Appendix E: potential management objectives and the approaches suggested to achieve each objective, suggests the types of objectives the AT felt each of the approaches would be able to address. Amendment 2 can be interpreted to imply consideration of menhaden management in the context of the stock itself and its predators. Considering Table 2 in this context, the single species models alone will not be able determine metrics that are obviously sensitive to provision for predators as part of ecological reference points. This implies that some of the other modelling approaches will be needed and will need to be explored. Two approaches are suggested: at a simple level the Steele-Henderson approach, or at a more complex level, multispecies models are needed to consider menhaden multispecies reference points. The Steele-Henderson approach may be the best starting point but multispecies models may be more informative.

Multispecies models tend to fall into two categories, i.e. MSVPA/MSSCAA or Ecosym/Ecopath. The first class of these could include menhaden and predators with a small range of interactions which can be parameterised at a population scale. This MSVPA/MSSSCAA approach would tend to fit well with the objectives as stated in Amendment 2. Ecosym/Ecopath provide a more complete range of interactions which conceptually will better represent the environment. These models require parameterisation of many trophic links. These may be difficult to parameterise robustly and predictions drawn from the models, when considered in the context of uncertainty, may provide little that is of useful guidance that would not be consistent with the MSVPA approach.

Multispecies models will not provide direct estimates of reference points, they will however, give indications of trade-offs between predator abundance and menhaden natural mortality. The information can be used to provide a framework to discuss the trade-off between forage fish exploitation and the exploitation / abundance of their predators. An example of such an illustration of the trade-offs for managers is given in a multispecies management plan evaluation for the Baltic Sea. This followed a formal consultation process first with review meetings of current management that did not involve multi-species issues, a scoping meeting (STECF2012a) which defined the work needed and follow up report meeting (STECF 2012b). The study is based on a basin scale single area multispecies model, which is parameterised for only a small range of species, herring and sprat as the forage fish and cod as the main predator. Some of the current issues and the main interdependencies are well described in Casini et al (2010) and Casini 2011. The fisheries which are dominated by cod fisheries are described in Bastardie et al (2010a and 2010b). The management of the five main pelagic stocks, four herring and one sprat stock, which form the forage fish in this area had already been evaluated in 2009 (ICES 2009). There was also some knowledge of environmental drivers and response to climate change in the Baltic (Mollman et al 2009 and MacKenzie et al 2007). All of this work was brought together under the STECF study (STECF 2012) which also

involved stakeholder and managers. The results for this study do provide managers with the sensitivity of the predators on the abundance of forage fish. This type of approach could potentially be used as a framework to consider suitable multispecies targets and limits reference points for the forage fish. The difficulties that were encountered in this relatively simple model were that the predation data were quite good for cod predation on sprat and herring, but sparse to characterise cannibalism of cod except at the basin scale, yet this was critical for understanding the considerations at higher cod biomass. Potential interactions such as cod egg mortality or density dependent growth of the forage fish were not explicitly included in the model. Both these effects might be expected to change the trade-offs and understanding of the implications of higher and lower exploitation rates. Currently in the Baltic cod are found to be growing slowly (ICES 2014), more slowly than any of the model predictions. There is some debate regarding the causes of this, the two main competing hypotheses are shortage of food, or parasite load. The first of these is not explained at the basin scale as sprat and herring are relatively abundant, but the effect might be dominated by local scale distributional changes, as the result of reduced area overlap and local depletion. Cod are currently occupying only part of the area of the Baltic Sea they previously occupied and the abundance of sprat and herring in this part of the area is low. For the parasites, the issue may be either causal or the effect of poor condition: seal parasites inducing poor growth, or poor growth resulting in greater vulnerability to parasites, the abundance of the parasites is linked to increases in seal populations in the Baltic. Thus this study gives some guidance regarding the type of information used to develop a multispecies management plan, both of a scientific nature and information for stakeholder involvement. However, the study does not provide direct ideas for reference points, rather it illustrates the difficulties that can be encountered and gives simple ideas of the trade-offs that are considered in this ‘simple’ case.

5. Panel review proceedings

I was impressed with the overall quality of this review and all who participated in it, I would like to thank all involved for their efforts. In particular I would like to thank the AT for a really excellent set of documents, well prepared and provided on time, and the AT presenters for their clear and well prepared presentations and the chairman for his work guiding the review and for the work assembling and editing the RP report.

All the data and assessment reports were provided on time. The overall quality of the assessment report and the work to support this was exceptional. The hard work of the AT both before and after the review should be commended.

6. Conclusion

The reports and presentations have provided an excellent basis to evaluate the performance of the assessment. The science reviewed was of a high standard and could be classed as ‘of the best scientific information available’. Comments given through this report should not be read as direct criticism of what has been done, rather ideas of areas for development. In retrospect one can always find room for improvement, and as such minor suggestions have been made throughout this report.

7. References

Bastardie, F., Nielsen, J. R., and Kraus, G. 2010a. The eastern Baltic cod fishery: a fleet-based management strategy evaluation framework to assess the cod recovery plan of 2008. – ICES Journal of Marine Science, 67: 71–86.

Bastardie, F., Vinther, M., Nielsen, J. R., Ulrich, C. and Storr Paulsen, M. 2010b. Stock-based vs. fleet-based evaluation of the multi-annual management plan for the cod stocks in the Baltic Sea, Fisheries Research, 101: 188–202.

Casini M., Bartolino, V., Molinero, J.C. and Kornilovs, G. (2010). Linking fisheries, trophic interactions and climate: threshold dynamics drive herring Clupea harengus growth in the central Baltic Sea. Marine Ecology Progress Series, 413:241-252.

Casini, M., Kornilovs, G., Cardinale, M., Möllmann, M ., Grygiel, W., Jonsson, P., Raid, T., Flinkman, J. and Feldman, V. (2011). Spatial and temporal density-dependence regulates the condition of central Baltic Sea clupeids: compelling evidence using an extensive international acoustic survey. Population Ecology, 53:511-523.

ICES 2009. Report of the Workshop on Multi-annual management of Pelagic Fish Stocks in the Baltic. 23-27 February 2009, ICES Headquarters, Copenhagen. ICES CM 2009/ACOM:38

ICES 2014 Report of the Baltic Fisheries Assessment Working Group. 12-19 April 2014, ICES Headquarters, Copenhagen. ICES CM 2011/ACOM:10

Gröger, J., Schnack, D., and Rohlf, N. 2000. Optimisation of survey design and calculation procedure for the international herring larvae survey in the North Sea. Archiv für Fischerei und Meeresforschung, 49: 103–116

Brian R. Mackenzie, Henrik Gislason, Christian Möllmann and Friedrich W. Köster 2007. Impact of 21st century climate change on the Baltic Sea fish community and fisheries. Global Change Biology,13:7, pp 1348–1367.

Möllmann, C., Muller-Karulis, B., Kornilovs, G., and St John, M. A. (2008). Effects of climate and overfishing on zooplankton dynamics and ecosystem structure: regime shifts, trophic cascade, and feedback loops in a simple ecosystem. ICES Journal of Marine Sciences, 65: 302–310.

Myers RA. 1998. When do environment-recruit correlations work? *Reviews in Fish Biology and Fisheries* 8: 285-305.

Patterson, K.R. and Beveridge, D.S. 1995. Report of the herring larvae surveys in the North Sea and adjacent waters in 1994/1995. ICES CM 1995/H:21

Sampson D. B. and Scott R. D., 2011, A spatial model for fishery age-selection at the population level Can. J. Fish. Aquat. Sci. 68: 1–10

STECF 2012 a Scientific, Technical and Economic Committee for Fisheries. Scoping for Impact Assessments for Multi-Annual plans for Baltic Multi-species and cod in

the Kattegat, North Sea, West of Scotland and Irish Sea (STECF-12-05). 94 pp. Scientific and Technical Research series – ISSN 1831-9424 (online), ISSN 1018-5593 (print) ISBN 978-92-79-24797-2 doi:10.2788/25628

STECF 2012b. Scientific, Technical and Economic Committee for Fisheries. Multispecies management plans for the Baltic (STECF-12-06). EUR – Scientific and Technical Research series – ISSN 1831-9424 (online), ISSN 1018-5593 (print) ISBN 978-92-79-24799-6 doi:10.2788/25697

Vinther, M and Lewy, P., 2012. SMS, a stochastic age-length structures multispecies models applied to North Sea and Baltic Sea stock. Working document to ICES WKMULTBAL, March 2012

Appendix 1: Bibliography of materials provided for review

Atlantic menhaden review workshop document list.

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

Background documents

SEDAR40 – 2	Fishery Dependent Indices
SEDAR40 - 3	Powerplant Impingement
SEDAR40 – 4	Fishery Independent Survey Standardization
SEDAR40 – 5	Fishery Independent Index Standardization Guidelines
SEDAR40 – 6	Atlantic Menhaden Stock Assessment Update – 2012
SEDAR40 – 7	Atlantic Menhaden Benchmark Stock Assessment and Review Panel Reports – 2010
SEDAR40 – 8	Hierarchical analysis of multiple noisy abundance indices. (P. Conn 2010)
SEDAR40 – 9	A proposed, tested, and applied adjustment to account for bias in growth parameter estimates due to selectivity (Schueller et al. 2013)

Appendix 2: Statement of Work

Statement of Work for John Simmonds

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

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

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

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

1) Conduct necessary pre-review preparations, including the review of background material and reports provided by the NMFS Project Contact in advance of the peer review.
2) Participate during the panel review meeting at the **Atlantic Beach, North Carolina** during **December 9-11, 2014.**
3) Conduct an independent peer review in accordance with the ToR (**Annex 2**) in Atlantic Beach, North Carolina during **December 9-11, 2014.**
4) **No later than 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: 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.

Appendix 3 Attendance and Review Group Agenda

Workshop Participants

Review Workshop Panelists

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

Analytical Representatives

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

Observers

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

Council and Comission Staff

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

Review Workshop Attendees

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

Workshop agenda

Tentative Agenda (Draft: 11/18/2014)

SEDAR 40 Atlantic Menhaden Review Workshop

Atlantic Beach, North Carolina
9-11 December, 2014

Tuesday

9:00 a.m. Convene
9:00 a.m. – 9:20 a.m Introductions and Opening Remarks Coordinator/Chair
- *Agenda Review, TOR, Task Assignments*

9:20 a.m. – 12:00 a.m. Assessment Presentations

-Regulatory History	*Mike Waine*
-Life History	*Amy Schueller*
-Commercial Reduction Fishery	*Joseph Smith*
-Commercial Bait Fishery	*Mike Waine*
-Indices of Abundance	*Amy Schueller*

12:00 a.m. – 1:30 p.m. Lunch Break
1:30 p.m. – 3:30 p.m. Continue Assessment Presentations

- Assessment Model and Results	*Amy Schueller*
- Reference Points and Stock Status	*Amy Schueller*
- Projection Methodology	*Amy Schueller*
- Research and Modeling Recommendations	*Genny Nesslage*

3:30 p.m. – 3:45 p.m. Break
4:00 p.m. – 5:15 p.m. Continue Presentations / Panel Discussion Chair
- *Continue presentations as necessary*
- *Identify additional analyses, corrections, etc.*

5:15 p.m. – 5:30 p.m. Day 1 Summary & assignments to analytical team Chair
5:30 p.m. – 6:00 p.m. Public Comment

***Tuesday Goals*:** Initial assessment presentations completed, sensitivity and base model discussion begun, additional analyses requested

Wednesday

8:30 a.m. – 9:30 p.m. Panel Comments **Chair**
- Initial panel comments on assessment
9:30 p.m. – 12:00 p.m. Analytical team report on additional analyses Analytical Team
Panel Discussion Chair
-*Continue panel deliberations*
12:00 a.m. – 1:30 p.m. Lunch Break
1:30 a.m. – 2:00 p.m. Ecological Reference Points Methods **Jason McNamee**

2:00 a.m. – 3:30 a.m. Panel Discussion **Chair**
- *Discussion on ecological reference points*
3:30 p.m. – 4:00 p.m. Break
4:00 p.m. – 5:30 p.m. Panel Discussion/Panel Work Session **Chair**
- Continue deliberations
- Recommendations and comments
5:30 p.m. – 6:00 p.m. Public Comment

***Wednesday Goals*:** Ecosystem reference points presentation completed, sensitivities and modifications identified, preferred models selected, projection approaches approved, report drafts begun

Thursday
8:30 a.m. – 10:30 a.m. Panel Discussion **Chair**
- Final analyses reviewed
- Projections reviewed **Chair**
10:30 a.m. – 11:00 a.m. Break
11:00 a.m. – 12:30 p.m. Panel Discussion or Work Session **Chair**
- Review Reports
12:30 p.m. – 1:00 p.m. Public Comment
1:00 p.m. ADJOURN

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