Center for Independent Experts (CIE) SEDAR 69 Atlantic Menhaden Assessment Review

Charleston, South Carolina November 4-8, 2019

External Independent Peer Review

By

Kenneth T. Frank, Ph.D., FRSC

Prepared for the Center for Independent Experts

February 2020

Table of Contents

1. Executive summary

The workshop for the single-species Atlantic menhaden peer review combined with the Atlantic menhaden Ecological Reference Points peer review was held in Charleston, South Carolina from 4-8 November 2019. The workshop was structured by formal presentations given by the Stock Assessment Subcommittee (SAS) and the Ecosystem Reference Points (ERP) working group, in-depth discussions associated with each of the several terms of reference (ToRs) and, at the end of each day, public comments.

At the start of the workshop, presentations by the SAS focused on the assessment of Atlantic menhaden involving outcomes of previous assessments, regulatory history of the stock, the input data from the commercial reduction and bait fishery, and treatment of fishery-independent indices of abundance. The Beaufort Assessment Model (BAM), a forward projecting statistical catch-at-age model (SCAA), was used to determine stock status and to provide management advice. This model has been used in past assessments of menhaden, both in the Atlantic (SEDAR 40) and Gulf stock (SEDAR 63). There were two major changes in the model since SEDAR 40 involving both the estimation of fecundity and natural mortality (M). The panel concluded that the rationale for these changes was clearly presented and agreed that the new methodologies better represented the life history attributes of menhaden.

Assessment model results were presented along with sensitivity runs, uncertainty analyses, and projections. Stock status was assessed against two reference points: one based on historical fishing mortality (F) and the other on fecundity. Both the terminal year target and threshold levels for each reference point were below the benchmark and on this basis it was concluded the stock was not overfished and over-fishing was not occurring. The panel agreed that the BAM model was an appropriate tool for assessing this stock and the conclusions regarding stock status. Concerns were raised by the panel for further improvements in the assessment that were considered recommendations for future research (e.g., developing a targeted fishery independent survey for menhaden, exploring alternatives to the current optimizer in the BAM model, improvement in collection of large fish).

The remainder of the workshop was devoted to the assessment of Ecological Reference Points (ERP) for Atlantic menhaden. Given the central position menhaden occupy in the food web as a forage fish species, it is considered essential to develop models to estimate population parameters of Atlantic menhaden (e.g., F, biomass, abundance) that take their pivotal role into account as well as to determine associated reference points and total allowable catch. Five ERP models were reviewed starting with simple (two surplus production models – no predators specified) to complex (a multi-species SCAA called VADER), and two versions of Ecopath with Ecosim or EwE (NWACS –MICE and NWACS-Full) with specific predators included. The panel agreed with the strategy of using input datasets (fishery independent indices, total catch, and both fishery dependent and independent age and length data) directly from previously vetted and approved stock assessments for modeled predator species. This streamlines the process for multispecies and ecosystem modeling by relying on existing processes to review input data. Diet data as input to the models was considered adequate but future improvements were recommended. While all five models could address sustainability for menhaden fisheries, only

VADER, NWACS-MICE and NWACS-FULL could address menhaden predators or their fisheries. Further, only the NWACS models directly model menhaden effects on predators as well as predator impacts on prey, and the NWACS-FULL model was difficult to update within required management time frames. The panel thus agreed with the ERP working group conclusion that NWACS-MICE is best able to address the full suite of management objectives when combined with BAM, which best captured menhaden population dynamics. The models are able to provide a quantitative representation of system and predator-prey dynamics and are therefore ready to be used to provide management advice.

2. Background

The process of conducting the review began with reading all of the material provided, which ranged from various technical documents associated with the Atlantic menhaden benchmark assessment and Ecosystem Reference Point assessment to publications in the primary scientific literature (Appendix 1). A Pre-Workshop webinar call was held on October 25, 2019 during which two power-point presentations were given to participants with an opportunity for the panel to ask questions regarding the material provided.

The review workshop took place in Charleston, SC and lasted five days from 4 -8 November 2019. Each day began with a series of presentations made first by members of the SAS and later by the ERP working group with follow-up questions and discussion after each presentation. All of the power-point presentation files were made available to the Panel during the course of the review.

The Chair of the review Panel was Dr. Michael Jones who facilitated the meeting and made sure that all the terms of reference (ToRs) were reviewed by the Panel. He also led the preparation of the Peer Review Panel Summary Report. Drs. Sarah Gaichas, Daniel Howell, Kenneth T. Frank, and Laurence Kell served as independent and impartial reviewers. The CIE reviewers (Howell, Frank, Kell) were requested to participate fully in the Panel review during the five day meeting and then to complete an individual, independent peer review report. The reporting format and content guidelines were to be developed in accordance with the requirements specified in the Statement of Work and terms of reference (Appendix 2); reviewers were not required to reach a consensus. CIE Reviewers contributed to the Peer Review Panel Summary Report.

Atlantic menhaden is an abundant, small-bodied, planktivorous clupeid species that occupies a central position in the food web of the mid- and south Atlantic region as a forage fish, linking the lower and upper trophic levels of the food web by grazing on planktonic organisms such as algae and zooplankton and serving as prey for fish such as striped bass, weakfish, and bluefish. Across this geographic range menhaden are managed as a single unit stock exhibiting somewhat regular seasonal migration patterns. The species supports a large-scale, commercial fishery having two major components: a purse-seine reduction sector that harvests fish for fish meal and oil and a bait sector that supplies bait to other commercial and recreational fisheries. The first total allowable catch (TAC) for Atlantic menhaden was implemented in 2013 at 170,800 mt, which was progressively increased to 216,000 mt in 2017. Since the TAC was established, reduction landings have ranged from 128,900 to 143,500 mt and bait landings have ranged from 37,000 to 45,500 mt. Management of Atlantic menhaden has traditionally not formally considered ecological interactions among species, but multi-species modeling approaches have matured to a point where meaningful management advice can result from their consideration.

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

Dr. Kenneth T. Frank (KTF) participated in all aspects of the review including reading all of the documents listed in Appendix 1 prior to the meeting, which were the foundation for the review, attending and participating in the 4-8 November 2019 panel review meeting in Charleston, SC, and assisting the Chair of the meeting with contributions to the summary report. KTF has extensive direct research experience in forage fish life history, ecology and population dynamics in the context of quantitative research on food chain dynamics and the dynamics of exploited marine ecosystems.

4. **Findings by Terms of Reference (ToR) for the Atlantic Menhaden Single-Species Peer Review**

ToR 1 Evaluate the thoroughness of data collection and the presentation and treatment of fishery-dependent and fishery-independent data in the assessment, including the following but not limited to:

a. Presentation of data source variance (e.g., standard errors).

b. Justification for inclusion or elimination of available data sources,

c. Consideration of data strengths and weaknesses (e.g., temporal and spatial scale, gear selectivities, ageing accuracy, sample size),

d. Calculation and/or standardization of abundance indices.

The Atlantic menhaden assessment used both fishery-dependent and fishery-independent data. Fishery-dependent data came from the commercial reduction and bait fisheries. Consideration of six fishery-dependent indices was dismissed as in SEDAR 40 (2015) because of the lack of age/length data. Two additional indices based on CPUE, used to support ecosystem reference point (ERP) modeling, were not used in the current assessment either. A few concerns were raised regarding the collection, assembly and treatment of the fishery-dependent data from the commercial reduction and bait fisheries. Expanded sampling was recommended in the bait fishery given the deficiency of age-5+ fish and the top of hold sampling in the reduction fishery should be examined further to ensure accurate characterization of the total trip catch, not just from the last tow which is the current procedure. Placing observers on boats to collect at-sea samples from purse-seine sets, or collecting samples at dockside during vessel pump-out operations could address this sampling issue. The panel considered the estimation of fishery-dependent data source variances to be appropriate.

A total of 49 fishery independent surveys were considered for inclusion in the assessment as indices of young-of-year (YOY) or adult (age-1+) abundance. These were initially screened using a standard set of criteria and reduced to a total of sixteen. The surveys do not specifically target menhaden and many use a small-mesh trawl as opposed to purse seining which is the dominant component of the fishery. Length frequencies are collected from each survey to separate YOY from older fish (age-1+) but there are no ageing data; consequently, none of the surveys had a time series of numbers or biomass at age. The lack of such data limits the utility of the surveys to track the temporal development of cohorts. However, the Connecticut Long Island Sound trawl survey, where 0-group vs 1+ were regressed on one another suggested some level of internal consistency in the survey. In this case, the YOY index was reasonably well correlated with the Age-1+ index the following year, suggesting both indices were picking up on a coherent population trend. However, such an analysis could only be conducted for this survey so that the reliability of the other surveys could not be assessed. Given the availability of length frequency data, it may be possible to assess cohort tracking through a modal analysis which forms the basis for a future research recommendation.

Cross-correlation analysis was used to assess the degree of coherence among the fishery-independent surveys. One would expect a high degree of correspondence in the index time trends given the unit stock assumption underlying the assessment. All pairs of correlation coefficients among the 16 YOY series (120 in total) were positive, with 14 being statistically significant at the $p < 0.05$ level. As might be expected, the significant correlations tended to occur among those surveys in close geographic proximity (on the order of 100 km or less). A similar result was evident in the analysis of the age-1+ indices (5 out of 28 correlations with $p < 0.05$). Given the life history characteristics of the species, the schooling nature of menhaden, and the large geographical scale of the surveyed population, the panel acknowledged that it was unrealistic to expect time trends in the indices to exhibit uniformly high correlation across the entire geographic domain of the surveys. Noteworthy is a recent cluster analysis of a subset of the YOY time series indices by Bucheister et al. (2016), which identified two broad geographical regions (Chesapeake Bay – CB, southern New England region - SNE) exhibiting inverse patterns of abundance. The CB indices were highest during the 1970s to early 1990s, whereas the SNE indices were high from about 1995 to 2005.

A research recommendation offered by the SAS for consideration in the longer term was the development of a coast wide, fishery-independent survey to replace or supplement existing indices. The design of such a survey could be informed by analysis of the existing non-targeted surveys by assessing the de-correlation spatial scale. It was noted that menhaden abundance from geographically adjacent surveys tended to be positively correlated. A more formal analysis would examine the correlation of all pairs of abundance indices from the sixteen surveys relative to their separation distances and then subsequently fit with an exponential decay model. A rapid (slow) decay implies local (large) scales are important in structuring the population. Typically, the spatial scale over which a property is coherent is commonly referred to as the e-folding scale and is expressed as the distance at which the correlation coefficient falls to e^{-1} or 0.37 (e.g., see Frank et al. 2016 for a fisheries application). This methodology could also be applied to some of the larger scale surveys by comparing correlations among survey strata (e.g., SEAMAP). The results of these types of analyses could provide useful information for the design of the future coast wide survey (i.e., by addressing the question of what is the minimum sampling intensity or station density necessary to adequately survey the population). Similar analyses could be done with the environmental data collected during each of the surveys, which could provide insights into the structuring mechanisms influencing menhaden distributions.

The multiple fishery-independent indices were combined into regional composite indices using a hierarchical modeling method developed by Conn (2008). This method was also used (and reviewed) in the previous assessment (SEDAR 40, 2015) for Atlantic menhaden. Four composite indices were developed: a northern adult index (NAD), a mid-Atlantic adult index (MAD), a southern adult index (SAD), and a young-of-year abundance index (YOY). As in past menhaden assessments, the use of the Conn (2008) method was discussed, specifically in terms of the mechanics of the procedure to combine surveys and exploration of the use of alternative weighting schemes. During the current review the use of the VAST (Vector-Autoregressive Spatio-Temporal) package as an alternative to the index standardized methodology by Conn (2008) was recommended. It is available in the R statistical environment and may be a superior choice as to how best to combine multiple indices of menhaden abundance.

Many large-scale commercial fisheries targeting small pelagic fish species (e.g., Atlantic and Pacific herring, capelin stocks in the Barents Sea and Iceland) employ acoustic technology to survey the population. While this method alone has limitations, when used in conjunction with other survey methods, it can provide useful data on biomass trends and distributional patterns. Despite these potential advantages, it appears acoustics may have limited application to menhaden given their occupancy of relatively shallow, highly turbid waters associated with nearshore and estuarine habitats. Additional survey strategies were discussed including utilization of acoustic surveys during winter in deeper, offshore waters and the use of aerial surveys given the ease with which highly concentrated menhaden schools can be detected by aircraft.

The panel briefly discussed the information on habitat conditions included in the single species assessment report, but did not devote much time to this because it does not currently play a significant role in the assessment. It was noted that the NEFSC Ecosystem Context Assessment data are of limited value because menhaden are rarely found in this trawl survey, and the panel encouraged the SAS to consider offering recommendations on how this survey might be made more relevant to menhaden. For both single species and ecosystem based assessments in the future, relevant information on environmental covariates that could affect menhaden productivity would be extremely valuable.

In summary, the ToR for the evaluation of the thoroughness of data collection and the presentation and treatment of fishery-dependent and fishery-independent data in the assessment were considered acceptable including the presentation of data source variance (e.g., standard errors), the justification given for inclusion or elimination of available data sources, acknowledgement of the data strengths and weaknesses (e.g., temporal and spatial scale, gear selectivities, ageing accuracy, sample size), and the calculation and/or standardization of abundance indices.

ToR 2. Evaluate the methods and models used to estimate population parameters (e.g., F, biomass, abundance) and biological reference points, including but not limited to:

a. *Evaluate the choice and justification of the preferred model(s). Was the most appropriate model (or model averaging approach) chosen given available data and life history of the species?*

The panel agreed that the BAM single-species Statistical Catch at Age (SCAA) model was an appropriate tool for assessing this stock. The base model was used in previous Atlantic menhaden assessments (SEDAR 40 and a 2017 Update) and has been used for the Gulf stock (SEDAR 63) so that the assessment team is well aware of its suitability as a model framework for this species. The model fit and performance was considered acceptable based on direct fits to the data (with error bars where appropriate) and presentation of bubble plots; selectivities were shown, and overall modeled populations and F presented. Surplus production models were presented but were not considered as an alternative to the BAM model.

The main concern with the BAM model not covered in ToR 2.c (below) was with the optimization of the solution. Some evidence was presented and investigated further during the meeting that there were occasions where the optimizer failed to converge to the appropriate solution. The panel recommended the use of a “Jitter” analysis to increase confidence in the final optimized solution. In such an analysis the effect of varying input parameters on model results is assessed. A stable model should converge on a global solution across a reasonable range of input parameters. The panel considered that such an analysis was necessary and sufficient to address the concerns, and that the optimizer (with a Jitter analysis) was suitable to support the stock assessment.

A second concern was over the lack of data on the large fish, which is problematic for a SCAA. Collecting better data on the larger fish is discussed further in a later ToR, where it is identified as an important data collection recommendation for this stock.

b. *If multiple models were considered, evaluate the analysts’ explanation of any differences in results.*

Three different single species models were presented, the BAM single species Statistical Catch at Age (SCAA) and two surplus production models (time-varying r and Steele-Henderson). The two surplus production models, utilizing fishery-dependent data, agreed with the BAM model about the overall trend of increasing Atlantic menhaden population size and decreasing exploitation rates over the last 30 years. The assessment team’s choice of the BAM over the surplus production models was considered appropriate and well justified, and the noted deficiencies of the available CPUE time series indicate that adequate data do not exist to use surplus production models to assess this stock.

c. *Evaluate model parameterization and specification (e.g., choice of CVs, effective sample sizes, likelihood weighting schemes, calculation/specification of M, stock-recruitment relationship, choice of time-varying parameters, plus group treatment).*

There were changes to the model since the previous review in SEDAR 40. The biggest of these were the change in fecundity estimation (nearly an order of magnitude increase in fecundity at age across all age groups) and the change in estimation of M (increases in M for all age groups (0-6y)).

The rationale for the change in fecundity was clearly presented and was summarized in an unpublished document by R. Latour and J. Gartland (VIMS, Gloucester Point, VA) entitled the *Reproductive biology and fecundity of Atlantic menhaden*. The panel agreed that the new methodology better represented the biology of the spawning of the menhaden and was therefore a better reflection of reality than the previous approach based on Lewis et al. (1987). The panel noted that while this change obviously scales the estimates of population-level fecundity (the metric reported as an alternative to spawning stock biomass), it has little impact on the management outputs of the model because the management targets are similarly scaled upward.

The change in estimated age-varying M had a larger impact on the model results. Age-varying natural mortality (M) was estimated using methods reported in Lorenzen (1996) and scaled to a recent analysis of historic tagging data (Lilijestrand et al. 2019a, 2019b). The panel noted that the historical tagging data were from the 1960s, but accepted that the conditions to repeat the experiment (processing plants spread out along the coast) no longer existed. Comparing the M-at-age estimates with other similar fish species provided some reassurance on the estimated values, and led the SAS to conclude the higher estimates were not unreasonable for this species. Estimating M was acknowledged to be extremely difficult in stock assessments and therefore the panel accepted the revised methodology.

Time varying blocks of selectivity were chosen, and the justification was clearly explained and appeared valid. The choice of selectivity form (logistic or flat-topped rather than dome shaped) for the NAD index (a composite index of adult menhaden based on three fishery-independent surveys in the northern portion of the stock range) was highlighted as an area of concern. The panel concluded that the logistic selection for the NAD was the appropriate choice at present, but recommended a re-evaluation of the available data aimed at identifying an index series that provided better coverage of the larger fish. Development of a survey to sample these larger fish directly was recommended by the panel.

Related to the issue of large fish, the panel noted a trend in the length-weight relationship for the larger fish. The predicted weight for larger fish (250 mm fork length) exhibited a declining trend during the past decade. However, given the data available, the SAS argued this could simply be due to a change in sampling, both spatially and temporally, for the larger fish. The panel accepted this position. The panel therefore does not recommend that this trend be accounted for in the model. Improvements in future data collection of larger fish would address this issue.

The panel noted that a modified (Dirichlet) likelihood scheme was used for this assessment, and that the change in methodology for computing the likelihoods had an impact on the final solution. The new likelihood weighting scheme for age and length compositions was considered appropriate by the panel. In principle, the new scheme is an improvement, as the Dirichlet is designed to be self-weighting, perform better with correlated data, and reduce the effect of outliers. The Dirichlet was also presented as being increasingly used in other US fisheries assessments. The panel recommended that the technical team compare the changes they obtained for menhaden with changes in other stocks where this change in likelihood calculation has occurred.

In summary, this is clearly a mature assessment model, and while noting that further research to improve the model is warranted (see ToR 8), the panel concluded that this represents a suitable and viable basis for giving advice for this stock.

ToR 3. Evaluate the diagnostic analyses performed, including but not limited to:

a. Sensitivity analyses to determine model stability and potential consequences of major model assumptions

Numerous sensitivity runs were presented. They were constructed by including or excluding surveys, either singly or excluded in combination. Mostly this showed little sensitivity to choice of tuning series, which might be expected given the lack of trend in most of the surveys. The exception was the NAD, where leaving this out had large impacts on the model results because this was the only series with a logistic selectivity function. The sensitivity runs did serve to highlight this feature of model tuning, and further sensitivity runs based on giving the NAD a dome-shaped selectivity confirmed that the selectivity was the driving factor, rather than the inclusion or exclusion of the NAD data series.

Further sensitivity tests on the choice of M showed that the absolute model estimates were sensitive to the choice of M, as would be expected, but that the trends and status determination were robust to the different M values examined.

A sensitivity test on the method used to compute misfit scores for the likelihood components was presented, and demonstrated quite a large change in modelled population. A recommendation for research would be to identify other stocks where similar changes have been made in the likelihood methodology, and compare the changes encountered here with those obtained in other assessments.

There is an issue over the level of detail to be provided on the outcomes of the sensitivity runs in future reviews. It is suggested that the diagnostics in the current report be retained, but that in addition a full suite of diagnostics (including the parameter estimates and likelihood components) be made available on line for examination prior to and during the review.

In general, the panel concludes that a reasonable range of modeling assumptions and dataset choices were examined in the sensitivity tests. These tests were successful in identifying issues around model stability that should be monitored during further use of the BAM assessment model.

b. Retrospective analysis

Retrospective analyses were completed by running the BAM model in a series of runs sequentially omitting years 2017 to 2015. The analysis was limited to these years to avoid changing model assumptions. The analysis revealed little systematic pattern but did highlight an instability, where the 2014 retrospective was markedly different from the 2013 and 2015 values. In this context the retrospective runs were useful for validating that such an outlier is not occurring in the terminal year of the assessment. Given the issue with fitting highlighted under ToR 2, above, it is important that the convergence of the retrospective runs also be validated using a Jitter analysis.

ToR 4. Evaluate the methods used to characterize uncertainty in estimated parameters. Ensure that the implications of uncertainty in technical conclusions are clearly stated.

Two methods were used to assess uncertainty for the BAM: a Monte Carlo Bootstrap (MCB) that focused on uncertainty in natural mortality and fecundity, and an MCMC analysis that evaluated uncertainty in the model estimated parameters. The MCB analysis revealed greater uncertainty in estimated population and fishery outputs (biomass, fishing mortality, recruitment). Uncertainty in the model estimated parameters (MCMC outputs) had a substantially smaller effect on these outputs but was developed for illustrative purposes and not for use in stock projections. Evaluation of stock status with respect to reference points was robust to the uncertainty scenarios presented.

The panel concluded that the methods used to characterize uncertainty in estimated parameters were appropriate and reasonable and appreciated the use of two methods for propagating uncertainty that addressed distinct sources of uncertainty. The panel had suggestions for future work on model uncertainty including i) examination of the relative contributions of uncertainty in M and uncertainty in fecundity on the MCB outputs, particularly with respect to their influence on stock status relative to reference points, ii) consider ways to combine the two methods into a single uncertainty analysis, and iii) investigate the cause of the bimodal distribution of outcomes depicted for F target and F threshold in Figure 247 from the MCB simulations where the lower mode suggested some simulations produced very low estimated F values.

Tor 5 Minority Report

There was no minority report.

ToR 6. Recommend best estimates of stock biomass, abundance, and exploitation from the assessment for use in management, if possible, or specify alternative estimation methods

The panel concluded that the estimates of biomass, abundance and exploitation rates presented for the Base Run of the BAM model reported in the Single Species report are the best available estimates of these quantities.

ToR 7. Evaluate the choice of reference points and the methods used to estimate them. Recommend stock status determination from the assessment, or, if appropriate, specify alternative methods/measures.

The panel discussed the choice of single species reference points, one based on fishing mortality and the other fecundity (proxy for SSB) and the methods used to estimate them. Current target and threshold fishing mortality reference points were based on the mean and maximum fishing mortality rates respectively for ages 2 to 4 during the period 1960-2012. During this period the stock was without management constraints and had not collapsed. Reference points for reproductive output were the fecundity (number of maturing or ripe eggs) estimates associated with the fishing mortality target and threshold estimated from the BAM. Calculations were based upon the estimated selection pattern of landings across all fleets and areas, and the assumed time invariant M-at-age and 1:1 sex ratio. Uncertainty in the derived reference points was estimated by two approaches, namely a parametric Monte Carlo bootstrap (MCB) procedure in which the input values of M and fecundity were resampled (MCB), and a MCMC analysis to estimate parameter uncertainty.

The panel agreed that the methods applied were sound and also agreed with the SAS decision not to base reference points on FMSY calculations for this species. Finally, the panel noted that while these single species reference points were appropriate in the context of a single species assessment, they eventually should be replaced by Ecological Reference Points (ERPs) informed by the multi-species analyses, as discussed at length in the ERP ToRs.

The capacity of the stock assessment model to forecast the future state of the resource is of obvious importance but there was no associated ToR explicitly considering an evaluation of the methods used for stock projections. Due to the assumed high level of natural mortality, future stock biomass will be largely driven by year-class strength and hence recruitment. There is not a stock and recruitment relationship and recruitment has been relatively stable (CV of 30%), with no sign of extreme recruitment events. Such a low CV for recruitment could be considered somewhat anomalous given that many small pelagic species exhibit so called “boom and bust” dynamics (e.g., see Alheit and Peck, 2019). However, the broad spatial and temporal pattern of spawning and the indeterminate batch fecundity characteristic of the species may serve to reduce the risk of recruitment failure in any given year. The SAS used a non-linear time series (NLTS) approach for projections, which in the absence of a stock recruitment relationship, was considered appropriate by the panel.

There is a relatively long lag between the last year in the assessment and the years for which TACs are being set, e.g., the last year in the current assessment is 2017, which will be projected for reported landings in 2018 and preliminary estimates for 2019. The TAC will then be set for 2020 to 2022. Given the high level of natural mortality, the stock in 2020 through 2022 will be dominated by year-classes not estimated by the current assessment. Therefore, an evaluation of prediction skill is important, and the panel suggested that the SAS consider this using a hindcasting approach. The panel also suggests that the SAS examine estimates of surplus production (SP) obtained from BAM to provide a check on whether predictions of changes in biomass (Bt+1 - Bt) can reliably be made based on catch and Bt. Answering whether similar B levels exhibited similar SP at different historical times is a check on whether there has been non-stationarity in production processes (Walters et al. 2008).

Finally, the panel encourages the SAS to explore use of empirical indicators to monitor stock status and performance. These could be used to examine spatial and temporal trends in stock demography (e.g., the relative abundance of large individuals that may make a major contribution to spawning potential, age or length structure of the population, length and or weight at age, condition factors, etc.). This is important since changes in growth have been seen in the past and natural mortality is substantially higher than fishing mortality.

ToR 8. Review the research, data collection, and assessment methodology recommendations provided by the TC and make any additional recommendations warranted. Clearly prioritize the activities needed to inform and maintain the current assessment, and provide recommendations to improve the reliability of future assessments.

A number of short and longer term recommendations were made for research, data collection, modelling and management by the technical committee (TC). The panel was in general agreement with these recommendations. The panel offered the following supplementary comments.

In the short term, the panel strongly encourages a thorough exploration of options for adding a survey or fishery dependent dataset that more representatively samples the larger, older fish in the population. One option might be expansion of sampling from the bait fishery. The panel also suggested further exploration of the size frequency data from the existing fishery independent surveys to assess the internal consistency of the surveys (i.e., cohort tracking) as a further screening tool for inclusion/deletion of surveys in future assessments (see ToR 1). Collection of age data for the existing fishery independent surveys is also considered a priority and has been recommended in the past (SEDAR 40). The panel also supported the TC's recommendations for the development of a coast-wide fishery-independent index of abundance-at-age given that none of the existing fishery-independent surveys are specifically directed towards menhaden.

One of the recommendations under assessment methods was to conduct a Management Strategy Evaluation. The panel agrees that an MSE would be valuable for evaluating the robustness to uncertainty of reference points and control rules informed by the single species assessment. Another potential benefit would be to evaluate the benefits of improved data collection and biological sampling. However, decisions about how to structure the MSE will require careful thought to avoid progress on management being impeded by a process that could take several years and require a large commitment of resources. The panel recommends that if an MSE is to be undertaken, the effort be framed in the context of Ecosystem Reference Points rather than single species management, as discussed in the review of the ERP report.

With respect to research recommendations regarding assessment methods, the panel noted that while the Automatic Differentiation optimizer in ADMB is fast and therefore efficient, it is vulnerable to false convergence problems. It is therefore important the performance of the optimization be examined carefully before presenting model solutions. There were instances with the single species BAM model (see ToR 2) where diagnostics suggested an optimization failure. The panel therefore strongly recommends the Jitter analysis be performed on any model solution. This involves running a large number of optimizations, each one with slightly different starting parameters, to increase confidence that the final solution represents the global optimum.

ToR 9. Recommend timing of the next benchmark assessment and updates, if necessary, relative to the life history and current management of the species.

The SAS recommended continuing the timing of benchmark assessments and updates for the single species assessment, with an update in 3 years and the next benchmark in 6 years. The panel supports this recommendation. The single species assessment model is "mature" and does not appear to require any substantial modifications that would warrant a benchmark sooner than 2025. Given the relatively short lifespan of Atlantic menhaden, and the unpredictability of future recruitment trends, it does not seem appropriate to extend the time between benchmarks beyond 6 years. The continuation of intervening updates of menhaden stock status as needed is also advised as was initiated in 2017. If recommendations from this review regarding the fishery independent assessment of larger, older menhaden are implemented in the near future, it will take several years for a new index time series to be useful in the assessment model.

The panel notes, however, that with movement towards ecosystem-based reference points for Atlantic menhaden and consequently linkages between management strategies for several species of ASMFC concern, there will be large benefits in the future for synchronization of assessment updates and benchmarks among the key species in the models that inform ecosystem-based reference points. This may have implications for the timing of future Atlantic menhaden assessment updates. On the other hand, logistical constraints alone may make assessment synchronization difficult to achieve, particularly if the assessment team and the lead assessors have multiple stock assessment responsibilities.

5. Findings by TOR for Atlantic Menhaden Ecological Reference Points Peer Review

ToR 1. Evaluate the justification for the inclusion, elimination, or modification of data from the Atlantic menhaden single-species benchmark assessment.

Two long-term fishery-dependent indices of abundance for Atlantic menhaden were considered for inclusion in the ERP assessment (specifically for the two production models): a commercial reduction fishery CPUE index (RCPUE index) and the Potomac River Fisheries Commission index (PRFC) derived from the commercial bait fishery. Neither of these indices was used in the single-species assessment for a variety of well understood reasons. However, surplus production models require relatively long CPUE time series for model tuning.

The ERP WG decided to use the RCPUE index rather than the PRFC for ERP model base runs because of its larger spatial coverage, consistently recorded unit of effort, known variance structure, support from supplemental analyses that showed relatively strong correlations with other sources of data, and the ability to standardize the data through explanatory covariates (week, factory, vessel size), among other factors. The panel agreed with the choice of the inclusion of the RCPUE and exclusion of PRFC.

The panel also agreed with the use of BAM model outputs for tuning the EwE models, and the data used in the menhaden single species assessment for the multi-species SCAA model. Finally, the panel encouraged the ERP WG to consider the pros and cons of directly inputting single species values into the ecosystem models versus allowing the models to estimate them.

ToR 2: Evaluate the thoroughness of data collection and the presentation and treatment of additional fishery-dependent and fishery-independent data sets in the assessment, including but not limited to:

a. Presentation of data source variance (e.g., standard errors).

b. Justification for inclusion or elimination of available data sources,

c. Consideration of data strengths and weaknesses (e.g., temporal and spatial scale, gear selectivities, aging accuracy, sample size),

d. Calculation and/or standardization of abundance indices.

The panel generally considered data collection, presentation and treatment of additional fishery-dependent and fishery-independent data sets across all five presented ERP models thorough and appropriate. The panel was impressed by the amount of work and extent of data collation and treatment represented in the report. The models differ in complexity with some having minimal data requirement and few assumptions, and others with extensive data requirements and many assumptions. The latter group, intended to represent multispecies and full food web dynamics, thus has many data gaps to be filled and alternative treatments of data to be considered. The panel proposed research recommendations for future work based on these concerns.

Overall, the panel agreed with the strategy of using input datasets (fishery independent indices, total catch, and both fishery-dependent and -independent age and length data) directly from previously vetted and approved stock assessments for modeled predator species. This streamlines the process for multispecies and ecosystem modeling by relying on existing peer review processes to evaluate input data. The panel did note that none of the data sources for ERP models were shown with standard errors or other measures of variance. In future reporting, presentation of measures of variance would be helpful.

The review panel therefore primarily considered whether the process for inclusion/elimination of available predator stock assessment input data sources was appropriate, and applied all considerations above to data sources that had not been previously reviewed. This includes diet data, dogfish inputs for the VADER model, and additional EwE inputs. Stock assessment model outputs (F, SSB, recruitment, or other model-estimated quantities) were not considered to be data even if they were inputs to ERP models and the panel considered these to be previously reviewed information.

Selecting stock assessment fishery independent indices

A subset of assessment inputs (fishery independent indices) were selected to streamline the amount of input data for ERP models. The panel agreed that ERP models do not need to use all index datasets that are incorporated into single species assessment models as long as the most influential indices are retained for ERP models. This was achieved by asking relevant Technical Committees for each species to provide advice on the top 3 most influential indices plus alternate indices for use in sensitivity runs. The panel agreed with this approach.

Diet data

ERP models were developed on diet data from the Northeast Area Monitoring and Assessment Program (NEAMAP), the Chesapeake Bay Multispecies Monitoring and Assessment Program (ChesMMAP), and the Northeast Fisheries Science Center Food Habits Database (NEFSC FHD). These bottom trawl survey programs covered a large proportion of the Atlantic coastal shelf and provided 10 – 30 years of diet data collected with consistent methodologies. Other localized sources of diet data were included as well. The panel agreed that these sources combined represent a good basis for diet composition inputs across the spatial range of menhaden and key modeled fish predators.

The longest time series of diet composition data is from the NEFSC bottom trawl survey. The panel noted that the BAM model does not use this survey as an index due to the low numbers of menhaden caught in the survey. However, the survey does frequently capture key menhaden predators but with low numbers of menhaden in predator stomachs. The panel agreed that including the NEFSC trawl survey diet data is useful, but strongly supports the inclusion of additional diet data from NEAMAP and ChesMMAP as well. The panel suggested including other sources of diet data, if possible, from additional surveys encountering menhaden and their predators.

Both ERP models identified as candidates for future work would benefit from additional sources of diet data as part of further development. The VADER model requires time series of diet by age. The EwE models require a snapshot diet composition input which more flexibly accommodates diet data from single studies to supplement survey-collected stomachs. The full EwE model requires diet information from seabirds, marine mammals, and other species not sampled by fish trawl surveys. ERP modelers suggested that the Bayesian approach for allocating diet information to appropriate model age and year bins could be adjusted to use additional diet data “snapshots” from single studies as priors and the panel supported this suggestion. In addition, other methods to estimate diets from genetic barcoding, stable isotopes, and fatty acids could be explored, possibly based on short-term research projects to identify promising datasets and methods. The panel also recommended further sensitivity analysis to evaluate impacts of different diet inputs to models (see below).

Spiny dogfish data inputs to VADER

The one exception to the direct use of previously vetted assessment data inputs was for spiny dogfish in the VADER model. The spiny dogfish assessment is a female-only index-based assessment because there is no catch at age data from the fishery (which catches 95% females). The ERP working group reconstructed male and female indices from NEFSC bottom trawl survey data, and also extrapolated fishery catch at age for spiny dogfish from length at age in surveys combined with an age-length key to create a combined-sex age structured population dynamics model. The assumption that fishery catch at age is similar to survey catch at age was not examined in depth during this review, nor were sensitivities conducted to evaluate the effects of alternative assumptions. The panel recommended collaboration with spiny dogfish assessment scientists to evaluate the most appropriate data to support future multispecies modeling. The methods as outlined in the assessment report seemed reasonable to run the VADER model, but the panel noted that more rigorous review of these methods is required if the VADER model will be used as a primary ERP model or as an MSE operating model in the future.

Ecopath biomass inputs

Assessment model-estimated biomass was used as an input to the Ecopath portion of both EwE models. This was considered appropriate by the panel given that the EwE models are intended to integrate what is currently considered the best available information across assessed species and to reconcile reference points at the scales currently used in management by incorporating predator-prey dynamics. This acknowledges that the EwE software is not well suited to statistical estimation of the observation processes required to convert survey index data into the snapshot of total biomass required to initialize the dynamic food web model. However, the panel noted that any potential (currently undetected) bias in stock assessment estimated biomass is passed along to EwE when using this as an input, and remains an issue for any subsequent ERP analysis.

The panel noted that the EwE-MICE model was fit to index data (as described above for the VADER model), rather than stock assessment model output. The use of stock assessment output in the initial mass balance sets up the scale of the food web model while the dynamic predator prey interactions are estimated from the combination of index trends across all species/groups in the EwE MICE model. In contrast, the full EwE model is fit to stock assessment estimated biomass trends rather than index data. This approach relies on a stronger assumption of the accuracy of stock assessment estimated trends across all modeled species in estimating dynamic predator-prey parameters. The panel was unable to fully evaluate the differences between these approaches.

The panel suggested that uncertainty information for the datasets used, in terms of metrics of their estimated reliability and confidence, be included with future documentation of both EwE models. This information could be summarized in a common format for these food web models and then carried forward in sensitivity and/or uncertainty analyses.

Environmental data (discussed but not implemented at present)

While there are stated objectives to ensure menhaden sustainability in the face of a changing environment, most of the ERP models presented do not yet include environmental data because they were developed to address predator-prey interactions as a first priority. The panel agreed with this prioritization for developing ERPs, and suggests that environmental drivers can be evaluated and incorporated as appropriate in the future.

The VADER model has the capacity to include a temperature time series which affects consumption rates; however, this was not implemented in the ERP model runs reviewed by the panel. Nevertheless, the description of the temperature data product in the report is an appropriate starting point for incorporating environmental data into the VADER model.

There are many sources for environmental and oceanographic data that can be applied at multiple spatial and temporal scales for the region addressed by ERP models. The panel suggested that future work could first identify any broad environmental drivers likely to affect all species in the VADER and EwE-MICE models and then investigate which datasets best represent these drivers. In addition, the panel agreed with ERP modelers that converting spatially-explicit environmental information into time series for input into spatially aggregated models is challenging, and research into appropriate methods for handling environmental data is appropriate. One could also consider using broad-scale indicators of climate variability such as the North Atlantic Oscillation (NAO) index or the Atlantic Multidecadal Oscillation (AMO) as environmental input to the ERP models. The extent to which either of these indices could affect all species in the VADER and EwE-MICE models is not known, but the AMO has been implicated as one of the best environmental predictors of coast-wide recruitment patterns for Atlantic menhaden (Buchheister et al. 2016) and Simpson et al. (2016) reached similar conclusions using a different recruitment time series. However, a recent investigation of the relationship of the AMO to local observations of ocean temperature across four arctic/subarctic regions revealed generally weak relationships, which suggested there is little support for the use of the AMO as a strong proxy for *in situ* temperature series (Frank et al. 2018). There is an obvious need to consider how broad-scale indicators of climate variability are influencing stock level environmental conditions.

ToR 3: Evaluate the methods and models used to estimate Atlantic menhaden population parameters (e.g., F, biomass, abundance) that take into account Atlantic menhaden's role as a forage fish, including but not limited to:

a. *Evaluate the choice and justification of the recommended model(s). Was the most appropriate model (or model averaging approach) chosen given available data and life history of the species?*

b. *If multiple models were considered, evaluate the analysts' explanation of any differences in results.*

c. *Evaluate model parameterization and specification as appropriate for each model (e.g., choice of CVs, effective sample sizes, likelihood weighting schemes, calculation/specification of M, stock-recruitment relationship, choice of time-varying parameters, choice of ecological factors).*

The ERP report presented five models to estimate menhaden population parameters that account for its role as a forage species. The models ranged in structural complexity from a simple surplus production model with time-varying menhaden production (SPM-TVr) to a full food web model for the Northwest Atlantic continental shelf (NWACS-FULL). Intermediate models included a surplus production model that explicitly accounted for menhaden removals due to predation (Steele-Henderson), a multispecies statistical catch at age model with menhaden and 5 other species (VADER), and a scaled down food web model focused on menhaden and a subset of key predator and prey species (NWACS-MICE). The panel took this ToR to mean a review of the ERP models used to develop a tool to estimate reference points and not an evaluation of the predator stock assessments that provided inputs to ERP models.

The ERP report recommended a combination of the BAM model and the NWACS-MICE model to estimate Atlantic menhaden population parameters that take into account its role as a forage fish. The ERP report further recommended this combination as a tool for managers to use in achieving multispecies objectives, rather than recommending a specific F or catch level for menhaden at this time. The panel agreed with these recommendations, and justifications and caveats for methods and models were provided.

The ERP working group provided strong justification for choosing the BAM and NWACS-MICE models. This was based on their ability to provide information relevant to ecosystem management objectives specified in a 2015 stakeholder workshop as well as their technical merits. Objectives and performance metrics from the 2015 workshop were focused on sustaining menhaden to provide for menhaden fisheries and their predators, to provide fishery stability, and to minimize risk due to changing environmental conditions. While all presented models could address sustainability for menhaden fisheries, only VADER, NWACS-MICE and NWACS-FULL could address menhaden predators or their fisheries and none of the models are currently set up to address changing environmental conditions. Further, only the NWACS models directly model menhaden effects on predators as well as predator impacts on prey, and the NWACS-FULL model was difficult to update within required management time frames. The panel thus agreed with the conclusion that NWACS-MICE is best able to address the full suite of management objectives when combined with BAM, which best captured menhaden population dynamics.

The panel noted that the VADER model may also be useful in the future for addressing the specified ecosystem management objectives if prey-dependent dynamics can be incorporated for modeled predators. However, the panel recognized that this may be difficult, and that there is a lack of published examples where this has been done within multispecies statistical catch at age models.

The ERP report retained analysis of all models and compared results across them (including BAM). The panel appreciated the clear, comparative summary of each model. The panel agreed that all models showed generally similar recent trends and scale in comparable outputs (Age 1+ biomass, exploitation rate). This approach increases confidence that input data rather than model structure is largely driving model results, and argues for continuing to maintain a suite of supporting models with a range of complexity. Differences between the results were mainly attributable to structural assumptions: for example, surplus production/biomass dynamics models are not designed to track short term biomass changes that arise from inter-annual recruitment variability. Further useful comparisons explaining differences between the NWACS-MICE and FULL models were made in the report and in presentations during the meeting.

The panel's evaluation of model parameterization and specification for each ERP model are provided below. The suggestions for further model development and evaluation should be interpreted as constructive advice for future work, not additional work that is required before the two recommended models (BAM and NWACS-MICE) can be used to guide future management.

NWACS-MICE (selected model)

The panel found the overall specification of the NWACS-MICE to be reasonable given the different requirements of the EwE modeling framework from those of BAM or other statistical catch at age frameworks. There are two components to the specification: the static (Ecopath) model and the dynamic (Ecosim) model. The static model initializes the time dynamic model, which is then calibrated using sum of squares fits to time series of biomass and catch for multiple species.

The NWACS-MICE static (Ecopath) model parameterization used information from regional databases and stock assessments as available; this is appropriate and is discussed in detail under ERP ToR 2. In particular, the panel supported the decision to use biomass accumulation terms, which does not force the food web model to start in equilibrium. Some parameters were used directly or aggregated from NWACS-FULL (such as diet imports for predators); therefore, these models should continue to be reviewed and updated together. The main issue noted with the static model parameterization related to low estimates of EE for menhaden age groups resulting from B and P/B inputs from the BAM for menhaden combined with diet and other inputs for predators, even when considering that only a subset of predators were included in the model. This was addressed in a sensitivity run (see ToR 5). Ecotrophic efficiency (EE) is defined as the proportion of the production that is utilized in the modeled ecosystem and accounted for by fishing, predation, migration, and biomass accumulation.

The panel had suggestions for future work on the static model parameterization; some can be addressed by sensitivities and some can contribute to uncertainty characterization (see ToRs 5 and 6).

- Biomass (B) parameters based on stock assessments may change substantially between assessments; therefore, sensitivity of the static model to changes in these inputs over a historical range of assessments for key species should be evaluated.

- There are multiple methods for estimating P/B and Q/B ratios based on empirical information that should be explored. The assumption that stock assessment M + F = Z = P/B is a reasonable starting point, but potentially builds in resilience of stock productivity to F that may not exist in reality. This is a particular concern if the model was parameterized at a time when a stock was overfished.
- The reduction from NWACS-FULL to NWACS-MICE results in simplification of food web network topology; while general results from the two models were similar, impacts of this simplification on dynamic model behavior could be evaluated in more detail as management for additional interacting species is considered.
- Uncertainty in input parameters can be characterized in an EwE data pedigree which ranks quality and relevance of B, P/B, Q/B, and Diet Composition (DC) information sources. Including this information on all static parameter sources (for NWACS-FULL and NWACS-MICE) will be useful in future reviews.

For the dynamic (Ecosim) parameterization, the panel expressed concern about the inherent software constraints in EwE that limits the ability of the modeller to fully explore interactions between the model and input data, or to conduct sensitivities on combinations of dynamic parameters that must be altered by hand in a GUI framework that limits reproducibility. However, several approaches were taken to compensate for inflexibility in the software, which gave the panel more confidence in the results. First, estimation of vulnerability parameters was done in stages to ensure that the model had fully converged. Second, a constraint was placed on estimated vulnerability parameters to limit predation mortality by a single predator to 75-100% of total natural mortality for the prey. This constraint was applied manually after the EwE automated fitting procedure. Third, multiple parameterizations were presented with combinations of other manually set dynamic parameters (prey switching, foraging time adjustment, etc.) to bound the behavior of the model. Future work is recommended to either formally incorporate more flexibility in the software or to move the NWACS models into a modifiable framework. The panel also suggested that future work could investigate using an ensemble of dynamic model parameterizations to provide advice, rather than selecting a single model. The panel also found the parameterizations and specification of supporting models to be reasonable, and made only brief comments for consideration in the future.

<u>Surplus production models, VADER, NWACS-FULL</u>

The panel agreed with the use of the surplus production model with time varying r, with the RCPUE time series as the primary input, as a supporting model rather than as a primary model for ERP advice. This simple model could be used to evaluate correlations between r and predator metrics, and could potentially contribute to a powerful predictive analysis to evaluate time series of menhaden r and effects on striped bass. For exploring model sensitivities in the future, evaluating impacts of starting the model in specific years (rather than at the endpoints in the full time series) would be useful. The panel found the simulation testing approach and results useful, and suggests that all ERP models should attempt similar testing.

The panel agreed with the use of the Steele-Henderson surplus production model as a supporting model rather than as a primary model for ERP advice. One suggestion was to consider consumption by all predators together as a functional group rather than just striped bass to see if explanatory power improved. The panel also suggests that other wide-area search algorithms could be applied to free the model from its current implementation in proprietary software.

The panel agreed with the use of the VADER statistical catch at age model as a supporting model rather than as a primary model for ERP advice at this time. There was interest from the panel in exploring the assumption of constant ecosystem carrying capacity but variable other food versus the more common assumption of fixed other food. Further, the interaction of other food with assumptions about unexplained mortality (M0) for each species should be explored to determine which factors influence the intensity of species interactions within the model. Conversions from length to age to weight for parameterizing the consumption equations may lose information content in the data relative to re-parameterizing some equations to be length-based. The panel noted that this is a promising approach that has advantages over EwE in its ability to estimate F and other quantities directly from the data that would be relevant to management. Using both models together in the future would give managers information that incorporates structural uncertainty. For future work, the panel recommends conducting a Jitter analysis similar to that applied to BAM to ensure that optimization is working as expected, exploring more ecologically direct length-based predator prey dynamics, and conducting simulation testing.

The panel agreed with the use of the NWACS-FULL model as a supporting model rather than as a primary model for ERP advice. The panel agreed that updating this model is more time-consuming than NWACS-MICE. However, it would be useful in future iterations to apply the same parameter estimation techniques as used for NWACS-MICE (see above), which alleviate some concerns that arise from EwE software constraints. In addition, alternative specifications that fit NWACS-FULL to index time series instead of assessments would be useful for comparison with NWACS-MICE. Further exploration of incorporating habitat drivers into NWACS-FULL would also be useful to address the management objectives to minimize risks due to shifting environmental drivers. General comparisons to earlier models of the system or parts thereof would have been useful. For example, Christensen et al. (2009) developed an EwE model to represent the Chesapeake Bay system in 1950 with detailed descriptions of the data used for model parameterization and calibration.

ToR 4. Evaluate the methods used to estimate reference points and total allowable catch.

The methods used to estimate ecosystem reference points were reviewed. The panel agreed that the methods presented were sound. Compared to SEDAR 40 in 2014, extraordinary progress has been made, and the value of the EOM in providing ecosystem objectives was recognized. The models presented provide a transparent approach that allows the trade-offs between menhaden and selected predators to be evaluated within a multispecies context. For example, NWACS-MICE and BAM could be used to develop a scientific management framework to set single species TACs and evaluate their impact on predator species. The panel concluded that the approach illustrated in the ERP report seems appropriate and is ready for presentation to managers to initiate discussions about trade-offs among potentially competing fishery objectives.

The example trade-off analysis, presented in Figure 144 of the ERP report, illustrated why estimation of reference points for menhaden needs to be integrated with discussion of reference points for predator stocks. Striped bass are currently above their F target and the future status of this population will influence the outcome for a range of menhaden reference points. More generally, ecosystem models such as NWACS-MICE provide a valuable tool for exploring scenarios corresponding to alternative stock levels and exploitation rates for the species included in the models. It will be important to explore a variety of scenarios and to communicate that ERPs are inter-dependent since changes in one stock will affect the levels of other species.

The panel believes that the models are ready to be used to provide management advice. The models are able to provide a quantitative representation of system and predator-prey dynamics. The next step is to start a dialogue with managers and to evaluate trade-offs between management objectives. This will support the selection of targets, limits and thresholds to balance overall management objectives for the resource.

ToR 5. Evaluate the diagnostic analyses performed as appropriate to each model, including but not limited to:

a. Sensitivity analyses to determine model stability and potential consequences of major model assumptions

The different classes of models evaluated here have different structures, and therefore used different sensitivity tests. The review of each of the models is presented below.

For the surplus production models, a sensitivity to the choice of two potential CPUE indices (RCPUE and PRFC) was conducted, and indicated significant differences in trend between 1970 and 1990, although a much closer agreement since 1990. The rationale presented to use the RCPUE as the base case and the PRFC as the sensitivity was clearly presented, and the panel concurred with the conclusion. Sensitivity tests for the production models also included a brief analysis of the impact of the start date and this is addressed under the retrospective analysis below.

For the VADER model, the sensitivity run comparing the model with and without trophic interactions produced counter-intuitive results. This could point at problems with the proportion of total mortality (Z) allocated to predation, and a research recommendation is to investigate this through a more detailed sensitivity analysis. The modelled sensitivity to alternate tuning indices and prey composition was also presented. The review panel recommended a sensitivity test of the choice to fix overall food biomass (other food plus modelled prey) against the alternate hypothesis of fixing other food and allowing total biomass to vary.

The panel concluded that the suite of sensitivity tests performed on the VADER model and the two surplus production models was adequate at this time, given that these are not being currently proposed for direct use in management. Should the VADER model be used to inform management in the future, the additional sensitivity tests noted above would be recommended.

For the NWACS-MICE and NWACS-FULL model a suite of sensitivity runs were conducted with alternative dynamic (Ecosim) parameterizations using iterative vulnerability estimation as described above under ToR 3. NWACS-FULL sensitivities explored model behavior with and without vulnerability caps, with and without manual adjustments to selected parameters, and with observed and increased diet proportions of menhaden for predators in the static Ecopath model. The NWACS-MICE sensitivities explored similar parameterizations to NWACS-FULL as well as the effect of EwE-estimated primary production anomalies. A final sensitivity examined impacts of fitting to recruitment deviations as well as increasing the prey-switching exponent to 1.5.

The panel found the range of sensitivity runs useful and informative. Exploration of sensitivity to Ecosim dynamic parameters is especially valuable because there are so few examples in the literature, and model results tend to be highly sensitive to these parameter settings. In general, sensitivity runs for NWACS-FULL suggested that manual tuning of parameters was necessary to balance model fits to biomass with reasonable stock-recruitment dynamics. For NWACS-MICE (fitting to indices rather than stock assessment outputs), sensitivity runs demonstrated that vulnerability caps reduced or eliminated model instabilities in projections, which is desirable.

One additional sensitivity test was performed for the NWACS-MICE model during the meeting, at the request of the review panel. This investigated the sensitivity of the results to increases in predation mortality for menhaden. The ecotrophic efficiency (EE) parameter represents the fraction of species production that is used within the ecosystem and for forage species. Food web models usually account for a substantial proportion of production as predation mortality, with EE often approaching 1. The panel noted that predation mortality on menhaden estimated by the model was quite low, and the proportion caused by any given predator was even lower (for 0 group menhaden around 4% of overall mortality came from striped bass, while for age 1+ menhaden predation the value was around 1%). The single sensitivity run indicated that increasing the EE to a higher (but reasonable) value by increasing predator diet proportions of menhaden increased the slope of the curve relating B/Btarget for striped bass to F in menhaden (Figure 148 in the ERP report). This resulted in little change in the results for small changes around the current menhaden F. However, the distance between Btarget and Bthreshold decreased as the slope increased, indicating that the results from larger changes in menhaden F could be sensitive to the choice of EE parameters.

This sensitivity test was welcomed by the panel, and indicated that the overall NWACS-MICE result was robust to both reasonable increases in predator consumption of menhaden from those currently observed in food habits data and to small changes from current management. The panel therefore recommends a further suite of sensitivity tests to examine how robust the results are for greater deviations from current management. The results of the sensitivity tests on all of the key outputs for management (ERP Report, Figures 144-148) should be investigated. These tests should cover:

- A more thorough investigation of reasonable bounds on predation mortality to evaluate the effect of low observed predation mortality on low EE

- Runs with menhaden B and P/B at different values (using bounds from BAM sensitivity runs), to evaluate the effect of high production on low EE

- Runs including a range of values for other predators in the ecosystem (current runs looked at only status quo F, while one could use target, threshold, or specified F based on catch limits expected for future years)

In addition, the panel recommends testing of key ERP results to the static (Ecopath) model input parameters (B, P/B, Q/B) for predators of menhaden and other key groups.

The panel concluded that the NWACS-MICE model is suitable for use in exploring trade-offs in a management context.

b. Retrospective analysis

A retrospective analysis, involving a three-year peel, was presented for the VADER model. This was limited by the three-year block used for averaging the prey preferences. This would not prevent a longer peel, but one would expect a discontinuity every three years as the peel extended to a different three-year block of diet preferences. Within the three-year period, the model was stable.

Retrospective analyses were also conducted for the production models by removing up to four years of data from the end of the time series. Since there is minimal contrast in the CPUE data at the end of the time series, this had little effect on model performance. In contrast, the outputs of the surplus production models were strongly influenced by the start time of the model. This is not a surprising result given that the greatest data contrast occurs in the early years. A research recommendation to conduct a retrospective-style analysis at the start of the surplus production models to identify which years had the greatest impact on model performance was offered.

ToR 6. Evaluate the methods used to characterize uncertainty in estimated parameters. Ensure that the implications of uncertainty in technical conclusions are clearly stated.

For the ERP models, less formal attention was given to characterizing uncertainty in estimated parameters. In a sense, the consideration of multiple models constitutes an approach to accounting for model uncertainty. The model comparisons presented in the ERP report generally suggested qualitative alignment among comparable simulations, particularly when the models were adjusted for scaling differences in relevant parameters, which is not surprising given the common datasets used to inform the various models. When alternative time series (e.g., PRFC index) were used to inform the models, the outcomes were quite different, but the panel interpreted this as evidence of the unsuitability of this local index.

Overall, the panel felt the level of uncertainty analysis was appropriate for this stage of ERP model development and application. However, two recommendations were offered for future consideration: i) if development of the VADER model continues and is considered informative for management advice, examination of the uncertainty/sensitivity related to the magnitude of M0 and its potential interaction with assumptions about the “other food” biomass pool (fixed versus variable, relative to size of explicit prey pools) would be desirable, and ii) examination of the influence of uncertainty about the distributions of base parameters for the NWACS MICE model would increase confidence in the results of trade-off analyses among species reference points and management strategies.

Tor 7. Minority report

There was no minority report.

ToR 8. Recommend best estimates of stock biomass, abundance, exploitation, and stock status of Atlantic menhaden from the assessment for use in management, if possible, or specify alternative estimation methods.

The panel recommended using the stock biomass, abundance, exploitation and stock status estimates from the base run of the BAM model for use in management, as per ToR 6 for the single species review.

ToR 9. Review the research, data collection, and assessment methodology recommendations provided by the TC and make any additional recommendations warranted. Clearly prioritize the activities needed to inform and maintain the current assessment, and provide recommendations to improve the reliability of future assessments.

The report included a number of recommendations for future research, data collection, modelling and management, for both in the short and long term. These included expanding collection of diet and condition data, to include non-finfish predators and data-poor prey species, to conduct management-strategy evaluation (MSE) to identify harvest strategies that will meet ecosystem management objectives, and to continue the development of the NWACS-MICE, NWACS-FULL and VADER models.

The panel fully supported these recommendations and a number of additional specific recommendations for research on the ERP models and assessment methods presented earlier in this report in the context of other the ToRs.

The panel recognized the potential strategic importance of conducting an MSE and noted the benefit of having already completed an Ecosystem Management Objectives (EMO) Workshop in 2015 to identify fundamental ecosystem management objectives for Atlantic menhaden. MSE could be used to examine alternative scenarios to ensure the management advice is robust and to fully explore trade-offs among alternative management strategies. However, as noted under ToR 8 for the single species review, it will be important to undertake careful planning prior to the initiation of an MSE.

6. Conclusion

The ASMFC Atlantic Menhaden Stock Assessment Subcommittee (SAS) and associated working groups (ERP working group) are to be commended for an extremely comprehensive and illuminating treatment of each of the ToRs with the conduct of the meeting held in a highly professional, well-organized manner. The team approach worked exceedingly well in dealing with such a large number of complicated objectives and the dedication and enthusiasm to completing the task at hand by all participants was greatly appreciated. The SEDAR staff were equally professional and effective both in advance of the meeting and throughout its conduct. The other members of the review panel provided excellent reviews and displayed an expert knowledge base closely aligned with meeting objectives. Finally, the review panel chair is to be commended for keeping everyone focused and engaged throughout the course of the week-long meeting. This was done in an effective and cordial manner. I have no negative comments about any aspect of the review process.

7. References

Alheit, J. and Peck, M.A. 2019. Drivers of dynamics of small pelagic fish resources: biology, management and human factors. Marine Ecology Progress Series 617-618:1-6.

Buchheister, A., Miller, T. J., Houde, E. D., Secor, D. H., and Latour, R. J. 2016. Spatial and temporal dynamics of Atlantic menhaden (*Brevoortia tyrannus*) recruitment in the Northwest Atlantic Ocean. – ICES Journal of Marine Science, 73: 1147–1159.

Christensen, Villy, and Alasdair Beattie, Claire Buchanan, Hongguang Ma, Steven J. D. Martell, Robert J. Latour, Dave Preikshot, Madeline B. Sigrist, James H. Uphoff, Carl J. Walters, Robert J. Wood, and Howard Townsend. 2009. Fisheries Ecosystem Model of the Chesapeake Bay: Methodology, Parameterization, and Model Explanation. U.S. Dep. Commerce, NOAA Tech. Memo. NMFS-F/SPO-106, 146 p.

Conn, P.B. 2008. Hierarchical analysis of multiple noisy abundance indices. Can. J. Fish. Aquat. Sci. 67: 108–120.

Frank, K.T., Petrie, B., Leggett, WC, and Boyce, D. G. 2016. Large scale, synchronous variability of marine fish populations driven by commercial exploitation. Proc. Natl. Acad. Sci. US 113(29): 8248-8253.

Frank, K. T., Petrie, B., Leggett, W. C., and Boyce, D. G. 2018. Towards a more balanced assessment of the dynamics of North Atlantic ecosystems—a comment on Drinkwater and Kristiansen. ICES Journal of Marine Science, doi:10.1093/icesjms/fsz102).

Lewis, R.M., Ahrenholz, D.W. and Epperly, S.P. 1987. Fecundity of Atlantic menhaden, *Brevoortia tyrannus*. Estuaries 10(4): 347-350.

Liljestrand, E.M., M.J. Wilberg, and A.M. Schueller. 2019a. Estimation of movement and mortality of Atlantic menhaden during 1966-1969 using a Bayesian multi-state mark recapture model. Fisheries Research 210: 204-213.

Liljestrand, E.M., Wilberg, M.J., Schueller, and A.M., 2019b. Multi-state dead recovery mark-recovery model performance for estimating movement and mortality rates. Fisheries Research 210: 214-233.

Lorenzen, K. 1996. The relationship between body weight and natural mortality in juvenile and adult fish: A comparison of natural ecosystems and aquaculture. Journal of Fish Biology 49: 627-647.

Simpson, C.A., Wilberg, M.J., Bi, H., Schueller, A.M., Nesslage, G.M. and Walsh, H.J. 2016. Trends in relative abundance and early life survival of Atlantic menhaden during 1977–2013 from long-term ichthyoplankton programs. Transactions of the American Fisheries Society, 145:5, 1139-1151.

Walters, C.J., Hilborn, R. and Christensen, V., 2008. Surplus production dynamics in declining and recovering fish populations. Canadian Journal of Fisheries and Aquatic Sciences, 65(11), pp.2536-2551.

Appendix 1: Bibliography of materials provided for review

SEDAR 69
Atlantic Menhaden
Document List

Document #	Title	Author
SEDAR 69 – SAR1	Assessment of Atlantic Menhaden Single Species Benchmark Report	To be prepared by SEDAR 69
SEDAR 69 – SAR2	Assessment of Atlantic Menhaden Ecological Reference Point Report	To be prepared by SEDAR 69
	ASMFC Instructions for Reviewers	
Supplementary Materials		
SEDAR 69 – RD01	SEDAR 40 Stock Assessment Report Atlantic Menhaden	SEDAR 2015
SEDAR69 – RD02	Hierarchical analysis of multiple noisy abundance Indices	P. Conn 2010
SEDAR 69 – RD03	Estimation of movement and mortality of Atlantic menhaden during 1966–1969 using a Bayesian multi-state mark-recovery model	Liljestrand et.al. 2019
SEDAR 69 – RD04	Trends in Relative Abundance and Early Life Survival of Atlantic Menhaden during 1977–2013 from Long-Term Ichthyoplankton Programs	Simpson et.al. 2016
SEDAR 69 – RD05	Multi-state dead recovery mark-recovery model performance for estimating movement and mortality rates	Liljestrand et. al. 2019
SEDAR 69 – RD06	A MULTISPECIES STATISTICAL CATCH-ATAGE (MSSCAA) MODEL FOR A MIDATLANTIC SPECIES COMPLEX	McNamee, 2018
SEDAR 69 – RD07	Evaluating the performance of a multispecies statistical catch-at-age model	Curti, 2013
SEDAR 69 – RD08	Parameter estimation in Stock Assessment Modelling: Caveats with Gradient-based algorithms	Subbey, 2018
SEDAR 69 – RD09	Reconciling single-species TACs in the North Sea demersal fisheries using the Fcube mixed-fisheries advice framework	Ulrich et.al. 2011
SEDAR 69 – RD10	Working Group on Mixed Fisheries Advice (WGMIXFISH-ADVICE)	ICES Advisory Committee, 2016

SEDAR 69 – RD11	Evaluation of Current and Alternative Harvest Control Rules for Blue Whiting Management using Hindcasting	Kell and Levontin, 2019
SEDAR 69 – RD12	Public comment Forum Submissions	SEDAR, 2019
SEDAR 69 – RD13	Cookbook for Using Model Diagnostics in Integrated Stock Assessments	Carvalho, 2019

Appendix 2: Performance Work Statement

Performance Work Statement (PWS) for Kenneth T. Frank

Center for Independent Experts (CIE) Program

External Independent Peer Review

SEDAR 69 Atlantic Menhaden Assessment Review

Background

The National Marine Fisheries Service (NMFS) is mandated by the Magnuson-Stevens Fishery Conservation and Management Act, Endangered Species Act, and Marine Mammal Protection Act to conserve, protect, and manage our nation's marine living resources based upon the best scientific information available (BSIA). NMFS science products, including scientific advice, are often controversial and may require timely scientific peer reviews that are strictly independent of all outside influences. A formal external process for independent expert reviews of the agency's scientific products and programs ensures their credibility. Therefore, external scientific peer reviews have been and continue to be essential to strengthening scientific quality assurance for fishery conservation and management actions.

Scientific peer review is defined as the organized review process where one or more qualified experts review scientific information to ensure quality and credibility. These expert(s) must conduct their peer review impartially, objectively, and without conflicts of interest. Each reviewer must also be independent from the development of the science, without influence from any position that the agency or constituent groups may have. Furthermore, the Office of Management and Budget (OMB), authorized by the Information Quality Act, requires all federal agencies to conduct peer reviews of highly influential and controversial science before dissemination, and that peer reviewers must be deemed qualified based on the OMB Peer Review Bulletin standards.
http://www.cio.noaa.gov/services_programs/pdfs/OMB_Peer_Review_Bulletin_m05-03.pdf).
Further information on the CIE program may be obtained from www.ciereviews.org.

Scope

The SouthEast Data, Assessment, and Review (SEDAR) is the cooperative process by which stock assessment projects are conducted in NMFS' Southeast Region. SEDAR was initiated to improve planning and coordination of stock assessment activities and to improve the quality and reliability of assessments.

SEDAR 69 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 69 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 specified format and contents of the individual peer review reports are found in Annex 1. The Terms of Reference (TORs) of the peer review are listed in Annex 2. Lastly, the tentative agenda of the panel review meeting is attached in Annex 3.

Requirements

NMFS requires three (3) reviewers to conduct an impartial and independent peer review in accordance with the Performance Work Statement (PWS), OMB guidelines, and the TORs below. The reviewers shall have a working knowledge 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 fisheries stock assessment. It would be preferable for CIE reviewers to have expertise in forage fish population dynamics, Statistical Catch-at-Age modeling, Multispecies/Ecosystem Models with a focus on Multispecies Statistical Catch-at-Age models and Ecopath with Ecosim models, menhaden/forage fish life history and ecology, and/or management strategy evaluations/decisional frameworks.

Tasks for Reviewers

1) Two weeks before the peer review, the NMFS Project Contacts 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 Contacts will consult with the contractor 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 PWS scheduled deadlines specified herein. The CIE reviewers shall read all documents in preparation for the peer review.

2) Attend and participate in the panel review meeting. The meeting will consist of presentations by NOAA and other scientists, stock assessment authors and others to facilitate the review, to answer any questions from the reviewers, and to provide any additional information required by the reviewers.

3) After the review meeting, reviewers shall conduct an independent peer review report in accordance with the requirements specified in this PWS, OMB guidelines, and TORs, in adherence with the required formatting and content guidelines; reviewers are not required to reach a consensus.

4) Each reviewer should assist the Chair of the meeting with contributions to the summary report.

5) Deliver their reports to the Government according to the specified milestones dates.

Foreign National Security Clearance

When 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 reviewers who are non-US citizens. For this reason, the 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/ and http://deemedexports.noaa.gov/compliance_access_control_procedures/noaa-foreign-nationalregistration-system.html. The contractor is required to use all appropriate methods to safeguard Personally Identifiable Information (PII).

Place of Performance

The place of performance shall be at the contractor's facilities, and in Charleston, SC.

Period of Performance

The period of performance shall be from the time of award through January 2020. Each CIE reviewer's duties shall not exceed 14 days to complete all required tasks.

Schedule of Milestones and Deliverables: The contractor shall complete the tasks and deliverables in accordance with the following schedule.

Within two weeks of award	Contractor selects and confirms reviewers
2 weeks prior to the panel review	Contractor provides the pre-review documents to the reviewers
November 4-8, 2019	Panel review meeting
Approximately 3 week later	Contractor receives draft reports
Within 2 weeks of receiving draft reports	Contractor submits final reports to the Government

Applicable Performance Standards

The acceptance of the contract deliverables shall be based on three performance standards: (1) The reports shall be completed in accordance with the required formatting and content; (2) The reports shall address each TOR as specified; and (3) The reports shall be delivered as specified in the schedule of milestones and deliverables.

Travel

All travel expenses shall be reimbursable in accordance with Federal Travel Regulations (http://www.gsa.gov/portal/content/104790). International travel is authorized for this contract. Travel is not to exceed $10,000.

Restricted or Limited Use of Data

The contractors may be required to sign and adhere to a non-disclosure agreement.

Project Contacts:

Larry Massey – NMFS Project Contact
150 Du Rhu Drive, Mobile, AL 36608
(386) 561-7080
larry.massey@noaa.gov

Kathleen Howington - SEDAR Coordinator
Science and Statistics Program
South Atlantic Fishery Management Council
4055 Faber Place Drive, Suite 201
North Charleston, SC 29405
Kathleen.howington@safmc.net

Annex 1: Peer Review Report Requirements

1. The report must 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 report must contain a background section, description of the individual reviewers' roles 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 must describe in their own words the review activities completed during the panel review meeting, including 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, but especially where there were divergent views.

c. Reviewers should elaborate on any points raised in the summary report that they believe 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 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 report shall represent the peer review of each TOR, and shall not simply repeat the contents of the summary report.

3. The report shall include the following appendices:

Appendix 1: Bibliography of materials provided for review

Appendix 2: A copy of this Performance Work Statement

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

Annex 2. Terms of reference.

TERMS OF REFERENCE

For the 2019 ASMFC Atlantic Menhaden Single-Species Benchmark Peer Review and
2019 ASFMC Atlantic Menhaden Ecological Reference Points Benchmark Peer Review

Terms of Reference for the Atlantic Menhaden Single-Species Peer Review

1. Evaluate the thoroughness of data collection and the presentation and treatment of fishery-dependent and fishery-independent data in the assessment, including the following but not limited to:
    a. Presentation of data source variance (e.g., standard errors).
    b. Justification for inclusion or elimination of available data sources,
    c. Consideration of data strengths and weaknesses (e.g., temporal and spatial scale, gear selectivities, aging accuracy, sample size),
    d. Calculation and/or standardization of abundance indices.

2. Evaluate the methods and models used to estimate population parameters (e.g., F, biomass, abundance) and biological reference points, including but not limited to:
    a. Evaluate the choice and justification of the preferred model(s). Was the most appropriate model (or model averaging approach) chosen given available data and life history of the species?
    b. If multiple models were considered, evaluate the analysts' explanation of any differences in results.
    c. Evaluate model parameterization and specification (e.g., choice of CVs, effective sample sizes, likelihood weighting schemes, calculation/specification of M, stock-recruitment relationship, choice of time-varying parameters, plus group treatment).

3. Evaluate the diagnostic analyses performed, including but not limited to:
    a. Sensitivity analyses to determine model stability and potential consequences of major model assumptions
    b. Retrospective analysis

4. Evaluate the methods used to characterize uncertainty in estimated parameters. Ensure that the implications of uncertainty in technical conclusions are clearly stated.

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. Recommend best estimates of stock biomass, abundance, and exploitation from the assessment for use in management, if possible, or specify alternative estimation methods.

7. Evaluate the choice of reference points and the methods used to estimate them. Recommend stock status determination from the assessment, or, if appropriate, specify alternative methods/measures.

8. Review the research, data collection, and assessment methodology recommendations provided by the TC and make any additional recommendations warranted. Clearly prioritize the activities needed to inform and maintain the current assessment, and provide recommendations to improve the reliability of future assessments.

9. Recommend timing of the next benchmark assessment and updates, if necessary, relative to the life history and current management of the species.

10. Prepare a peer review panel terms of reference and 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.

Terms of Reference for Atlantic Menhaden Ecological Reference Points Peer Review

1. Evaluate the justification for the inclusion, elimination, or modification of data from the Atlantic menhaden single-species benchmark assessment.

2. Evaluate the thoroughness of data collection and the presentation and treatment of additional fishery-dependent and fishery-independent data sets in the assessment, including but not limited to:
    a. Presentation of data source variance (e.g., standard errors).
    b. Justification for inclusion or elimination of available data sources,
    c. Consideration of data strengths and weaknesses (e.g., temporal and spatial scale, gear selectivities, aging accuracy, sample size),
    d. Calculation and/or standardization of abundance indices.

3. Evaluate the methods and models used to estimate Atlantic menhaden population parameters (e.g., F, biomass, abundance) that take into account Atlantic menhaden's role as a forage fish, including but not limited to:
    a. Evaluate the choice and justification of the recommended model(s). Was the most appropriate model (or model averaging approach) chosen given available data and life history of the species?

b. If multiple models were considered, evaluate the analysts' explanation of any differences in results.
c. Evaluate model parameterization and specification as appropriate for each model (e.g., choice of CVs, effective sample sizes, likelihood weighting schemes, calculation/specification of M, stock-recruitment relationship, choice of time-varying parameters, choice of ecological factors).

4. Evaluate the methods used to estimate reference points and total allowable catch.

5. Evaluate the diagnostic analyses performed as appropriate to each model, including but not limited to:
   d. Sensitivity analyses to determine model stability and potential consequences of major model assumptions
   e. Retrospective analysis

6. Evaluate the methods used to characterize uncertainty in estimated parameters. Ensure that the implications of uncertainty in technical conclusions are clearly stated.
7. 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.

8. Recommend best estimates of stock biomass, abundance, exploitation, and stock status of Atlantic menhaden from the assessment for use in management, if possible, or specify alternative estimation methods.

9. Review the research, data collection, and assessment methodology recommendations provided by the TC and make any additional recommendations warranted. Clearly prioritize the activities needed to inform and maintain the current assessment, and provide recommendations to improve the reliability of future assessments.

10. Recommend timing of the next benchmark assessment and updates, if necessary, relative to the life history and current management of the species.

11. Prepare a peer review panel terms of reference and 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.

Annex 3. Agenda

Tentative Agenda (Draft 08.02.19)

SEDAR 69 Atlantic Menhaden & Ecological Reference Points

Review Workshop

Charleston, South Carolina
November 4-8, 2019

Monday

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. – 11:00 a.m.	**Assessment Presentations: Atlantic menhaden**	
	- Assessment History	*Amy Schueller*
	- Life History	*Amy Schueller*
	- Regulatory History	*Max Appelman*
	- Commercial Reduction Fishery	*Ray Mroch*
	- Commercial Bait Fishery	*Kristen Anstead*
	- Indices of Abundance	*Kristen Anstead*
11:00 a.m. – 11:15 a.m.	**Break**	
11:15 a.m. – 12:15 p.m.	**Continue Assessment Presentations**	
	- Assessment Model and Results	*Amy Schueller*
12:15 p.m. – 1:30 p.m.	**Lunch Break**	
1:30 p.m. – 3:30 p.m.	**Continue Assessment Presentations**	
	- Reference Points and Stock Status	*Amy Schueller*
	- Projection Methodology	*Amy Schueller*
	- Research and Modeling Recommendations	*Kristen Anstead*
3:30 p.m. – 3:45 p.m.	**Break**	
3:45 p.m. – 4:45 p.m.	**Panel Discussion**	**Chair**
	- Begin discussion with SAS	
	- Identify additional analyses, sensitivities, corrections	
4:45 p.m. – 5:15 p.m.	**Panel Comments**	**Chair**

	- Initial panel comments on assessment	
5:15 p.m. – 5:45 p.m.	**Day 1 Summary & assignments to analytical team**	**Chair**
5:45 p.m. – 6:00 p.m.	**Public Comment**	

Monday Goals**:** Initial single-species assessment presentations completed, sensitivity and base model discussion begun, additional analyses requested

Tuesday

8:30 a.m. – 9:00 a.m.	**Review additional single-species analyses**	*Amy Schueller*
9:00 a.m. – 10:30 a.m.	**Ecological Reference Points Assessment**	
	-Ecological Modeling Objectives	*Matt Cieri*
	-Modeling History	
	-Predator & Prey Choices	
	-*Multispecies Data*	*Katie Drew*
10:30 a.m. – 10:45 a.m.	**Break**	
10:45 a.m. – 11:45 a.m.	**Ecosystem Modeling Presentations**	
	Multispecies Surplus Production Models	*Katie Drew*
11:45 a.m. – 12:15 p.m.	**Panel Discussion**	**Chair**
	- *Discussion on surplus production models*	
	- Identify additional analyses to be requested	
12:15 p.m. – 1:30 p.m.	**Lunch Break**	
1:30 p.m. – 2:30 p.m.	**Ecosystem Modeling Presentations Continued**	
	Multispecies Statistical Catch-at-Age Model	*Jason McNamee*
2:30 p.m. – 3:15 p.m.	**Panel Discussion**	**Chair**
	-Discussion of MSSCAA model	
	-Identify additional analyses to be requested	
3:15 p.m. – 3:30 p.m.	**Break**	
3:30 p.m. – 4:30 p.m.	**Ecosystem Modeling Presentations Continued**	
	Ecopath with Ecosim Models	*Dave Chagaris*

4:30 p.m. – 5:15 p.m.	**Panel Discussion**	**Chair**
	-Discussion of EwE models	
	-Identify additional analyses to be requested	
5:15 p.m. – 5:45 p.m.	**Day 2 Summary & assignments to analytical team**	**Chair**
5:45 p.m. – 6:00 p.m.	**Public Comment**	

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

Wednesday

8:30 a.m. – 10:30 a.m.	**Ecological Reference Points Presentation**	
	- Review & Synthesis of Result	*Matt Cieri &*
	- Management & reference points recommendations	*Dave Chagaris*
10:30 a.m. – 11:00 a.m.	**Break**	
11:00 a.m. – 12:00 p.m.	**Panel Discussion**	**Chair**
	- *Ecological reference points & management*	
	- Identify additional analyses to be requested	
12:00 p.m. – 1:30 p.m.	**Lunch Break**	
1:30 p.m. – 3:30 p.m.	**Continue Panel Discussion**	**Chair**
	- *Ecological reference points & management*	
	- Identify additional analyses to be requested	
3:30 p.m. – 4:00 p.m.	**Break**	
4:00 p.m. – 5:00 p.m.	**Review additional ecosystem modeling analyses**	*TBD*
5:00 p.m. – 5:45 pm.	**Day 3 Summary & assignments to analytical team**	**Chair**
5:45 p.m. – 6:00 p.m.	**Public Comment**	

Wednesday Goals**:** Initial review and discussion of reference points and management recommendations

Thursday

8:30 a.m. – 10:30 a.m.	**Panel Discussion**	**Chair**
	- Final menhaden analyses & projections reviewed	
10:30 a.m. – 11:00 a.m.	**Break**	

11:00 a.m. – 12:00 p.m.	**Panel Discussion**	**Chair**
	-Single-species discussions continues	
12:00 p.m. – 1:30 p.m.	**Lunch Break**	
1:30 p.m. – 3:30 p.m.	**Panel Discussion**	**Chair**
	- Final ecosystem analyses reviewed	
3:30 p.m. – 4:00 p.m.	**Break**	
4:00 p.m. – 5:45 p.m.	**Panel Discussion**	**Chair**
	- Ecological reference points assessment	
5:45 p.m. – 6:00 p.m.	**Public Comment**	

Friday

8:30 a.m. – 10:30 a.m.	**Panel Discussion/Panel Work Session**	**Chair**
	- Continue deliberations	
	- Recommendations and comments	
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**	

Appendix 3 Panel Membership

The review panel consisted of Dr. Michael Jones (Chair), and Council of Independent Expert reviewers Dr. Kenneth T. Frank (author of this report), Dr. Laurence Kell, and Dr. Daniel Howell. In addition, Dr. Sarah Gaichas was a member of the review panel, although not a CIE reviewer. Dr. Michael Jones is Professor Emeritus at the Quantitative Fisheries Center at Michigan State university. Dr. Kenneth Frank is Research Scientist at Fisheries and Oceans Canada. Dr. Laurence Kell is Visiting Professor in Fisheries and Management at Imperial College London. Dr Daniel Howell is Research Professor at IMR, Norway. Dr. Sarah Gaichas is Research Fisheries Biologist at NOAA.

Workshop Participants

SEDAR 69
Atlantic Menhaden
Single Species & Ecological Reference Points Review Workshop Participants

APPOINTEE	FUNCTION	AFFILIATION/LOCATION
Review Panel		
Mike Jones	Review Panel Chair	ASFMC Appointee
Sarah Gaiches	Reviewer	ASMFC Appointee
Kenneth Frank	Reviewer	CIE
Daniel Howell	Reviewer	CIE
Laurence Kell	Reviewer	CIE
Analytical Representatives		
Amy Schueller	Single Species Lead Analyst & Chair	SEFSC – Beaufort, NC
Jason McNamee	ERP Lead Analyst	RI DEM – Jamestown, RI
Matt Cieri	ERP Work Group Chair	ME DMR – Boothbay, ME
Katie Drew	Assessment Team	ASMFC – Arlington, VA
Kristen Anstead	Assessment Team	ASMFC – Arlington, VA
Dave Chagaris	ERP Work Group	UF – Gainesville, FL
Ray Mroch	Assessment team	SEFSC- Beaufort, NC
Staff		
Max Appelman	Atlantic Menhaden Coordinator/Rapporteur	ASMFC – Arlington, VA
Sarah Murray	ERP Coordinator/Rapporteur	ASMFC – Arlington, VA
Pat Campfield	ASMFC Contact	ASMFC – Arlington VA
Ciera Graham	Admin	SAFMC
Kathleen Howington	Coordinator	SEDAR
Observers		
Bob Beale	Observer	ASFMC
Julie Neer	Observer	SEDAR
Joseph Ballenger	Observer	SCDNR
Peter Himcheck	Observer	Omega Protien
Genny Nesslage	Observer	UMCES
Chris Dollar	Observer	TRCP
Howard Townsend	Observer	NOAA Fisheries
Jeff Kaelin	Observer	Lunds Fisheries

Acronyms

ASMFC – Atlantic States Marine Fisheries Commission
CIE – Center for Independent Experts
ERP – Ecological Reference Points
ME DMR – Maine Department of Marine Resources
RI DEM – Rhode Island Department of Environmental Management
SEDAR – Southeast Data, Assessment, and Review
SEFSC – Southeast Fisheries Science Center, NMFS
UF – University of Florida
SAFMC – South Atlantic Fishery Management Council
UMCES – University of Maryland Center for Environmental Science
TRCP – Theodore Roosevelt Conservation Partnership