Center for Independent Experts (CIE) independent peer review report for the Gulf of Mexico Menhaden Stock Assessment Review

Anders Nielsen

December 2018

Executive Summary

The assessment team clearly presented a thorough description of the data sources, data collection, assessment model configuration, model validation, and model derived quantities for Gulf of Mexico Menhaden. The data sources available for the assessment are collected in a consistent manner and especially the catch seems to be closely monitored (reduction fleet data weekly constitutes almost all catches). The decisions to include/exclude certain data sources appear consistent. The primary base assessment model is the Beaufort Assessment Model (BAM), which is a standard, well tested, and documented assessment model. The configuration of the model for Gulf of Mexico Menhaden does include fixing (not estimating) a lot of model parameters, but the model fitted the observations well. The assessment team was very comfortable adjusting different parts of the model on request, which further strengthens the confidence in the configuration for Gulf of Mexico Menhaden. The results of the assessment showed no indication of Gulf of Mexico Menhaden being overfished or undergoing overfishing. The estimated stock-recruitment relationship is not informative for defining reference points. The reviewed science is the best currently available for the Gulf of Mexico Menhaden. This report further suggests possible improvements to model validation, uncertainty quantifications, and reference points, but overall this assessment is evaluated to be reliable.

Background

The review workshop of the Gulf of Mexico Menhaden assessment is part of the Southeast Data, Assessment, and Review (SEDAR 63) cooperative process for assessments conducted in NMFS' Southeast region. The meeting was convened in New Orleans, LA, November 6-7. The relevant documents (see Appendix 1) were made available prior to the meeting (and a pre-meeting call was conducted). At the meeting the assessment team presented the details of the assessment and responded to all questions from the review panel (see appendix 3 for members of the assessment team and review panel). Everything was well organized and very efficient. The goal of the review meeting was to strengthen confidence that the assessment was scientifically sound and that that results were reliable. This report documents the independent review of Anders Nielsen (see appendix 2 for the Statement of Work).

Description of the reviewer's role

This reviewer has independently read the assessment report, its appendices and all supplementary documents deemed necessary in preparation for this review, participated in a pre-meeting online meeting, traveled and participated actively in the review meeting, identified key issues in the assessment, suggested guidance, and independently authored this review report.

Findings for each term of reference

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

1) Evaluate the data used in the assessment, addressing the following:

a) Are data decisions made by the Data and Assessment Workshop sound and robust?
b) Are data uncertainties acknowledged, reported, and within normal or expected levels?
c) Are data applied properly within the assessment model?
d) Are input data series reliable and sufficient to support the assessment approach and findings?

The main data sources are seine recruitment index, gill net adult index with corresponding length compositions, and catch data from the reduction fleet with corresponding age composition data. The time-line for the different data sources can be seen in figure 1 (which was requested at the review meeting).

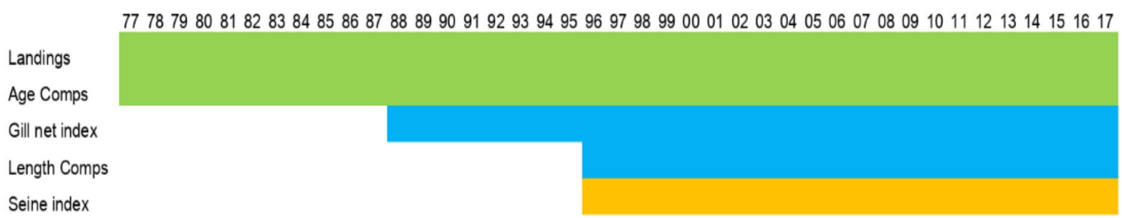

Figure 1: The time-line for the different data sources entering the model.

The catch data from the reduction fleet, which constitutes almost all catches, appear to be well reported and sampled by a standardized procedure. The fishery is conducted in a similar way throughout the assessment period (even including the use of spotter planes). The catch data is reported weekly but aggregated to yearly observations. The fishing season runs ca. from April to late October but has varied a bit with respect to both start- and end-time and when the intensity peaks. An effort time series is available (Vessel-ton-week) and appears to correlate well with the landings. It is not used in the current assessment model. The aging compositions are based on scale-readings and the observation noise and bias has been extensively explored by both within reader and across reader double-reading studies. The reduction landings are trusted to be very accurate (assigned CV of 0.04) and the age reading error is well explored.

The gill net survey index adds important information about the old individuals. It is based on the Louisiana gill net data. Gill net data from other states (Texas, Mississippi, and Alabama) were excluded due to concerns regarding questionable length observations (species identification) and shorter time periods. The gill net index captures both larger and smaller individuals than the reduction fleet. It does not contain age compositions, but only length compositions. Spatially it covers the central part of the population's range.

The seine data from three states (Louisiana, Alabama, and Mississippi) are combined to produce a juvenile index. Seine data from Texas and Florida were excluded, because of concerns of problematic or no species identification and high degree of mixing. The seine gear used in this index is not designed to capture Menhaden, but captures mainly age 0, so the index is used as an age 0 only index with no additional age or length composition data.

Inshore trawl data also exists at state level. It is currently not used in the assessment because previous benchmarks and the current assessment panel have concluded that the gear is not suited to catch menhaden well, and because some states had problematic species separation. In the assessment report it is also listed, as a reason to exclude the trawl data, that the seine data provides a preferred index recruitment index. This is not a logical reason, because it is possible to have more than one index of the same thing in the model. In fact, that is preferable because it can demonstrate inconsistencies and provide valuable information about uncertainties.

The stock identity (one single stock) is supported by genetics, and the sampling of biological parameters follows very standardized procedures.

A main source of uncertainty in the assessment is the assignment of natural mortality. Many alternatives for setting natural mortality were considered (e.g., tagging studies, Lorenzen, Hoenig, Jensen, Alverson and Carney, "Rule of thumb", multi-species models). The values from these methods varied substantially. This led to the impression at the review meeting that natural mortality was particularly uncertainly determined for Gulf of Mexico Menhaden, which this reviewer thinks is a bit misleading. Large uncertainty about natural mortality is common in applied assessment models and natural mortalities are often assigned to arbitrary values. The large number of presented options for assigning natural mortality mainly demonstrated that the assessment team had explored a lot of options. All the explored options are not equally valid/plausible. The choice of using a Lorenzen curve scaled to the mean natural mortality from tagging study appears reasonable.

Overall, the data decisions made by the Data and Assessment Workshop are sound, consistent, and robust. The uncertainties are acknowledged, reported, and within normal or expected levels. The data are applied properly within the assessment model, but more on this point under TOR 2. The input data series appear reliable and sufficient to support the assessment approach and findings. The data could possibly support an even more fine time -scaled modelling approach.

2) Evaluate the methods used to assess the stock, taking into account the available data.

a) Are methods scientifically sound and robust?
b) Are assessment models configured properly and used consistent with standard practices?
c) Are the methods appropriate for the available data?

The model proposed for the assessment of Gulf Menhaden is scientifically sound and robust . The model is the Beaufort Assessment Model (BAM). BAM is a standard model which has been used for many already accepted SEDAR assessments (e.g. Atlantic Menhaden, Spanish Mackerel, and Red Grouper). All model details are documented (SEDAR63-RD13). The core model is a statistical catch-at-age model, but it also allows matching to length composition observations. Statistical catch-at-age models are commonly used to assess fish stocks worldwide (e.g., via models like SS3, SCAA, and ASAP).

The specific configuration for Gulf Menhaden is clearly documented in the assessment report. It was clear from the review meeting that the model is not a black box to the assessment scientists. They were able to answer any questions and modify the model beyond changing simple configurations (including modifying the code). This greatly strengthens confidence in the configuration. This reviewer is convinced the model is configured properly and consistent with standard practices.

Many model parameters are input as fixed values (partial selectivity, growth, steepness, recruitment deviation variance, natural mortality, and uncertainty parameters). The model is however able to give good predictions of all data sources (landings, indices, and age and length compositions), so the assigned values appear reasonable, at least for the model parameters influencing the model's mean value prediction. Strictly speaking, it cannot be seen directly from this if the uncertainty parameters have been assigned reasonable values, but at least it can be seen that no data source has been completely ignored due to improper implied weighting.

The uncertainty parameters are also not estimated directly and jointly within the model but input as fixed values. The values are fixed at levels that are either coming from sample sizes (with subsequent Francis adjustment) or from external evaluation of the uncertainties of the input data sources in isolation. This reviewer prefers estimating such parameters jointly within the model, but if that is not possible, then the approach taken here appears reasonable. It is certainly within standard practices.

The model is appropriate for describing the available data, but data for an even more fine time-scale resolution is available. It would be of interest to evaluate the effect of the timing of the fishing season, and the relative fishing pressure within the fishing season, relative to mortality from other sources (natural mortality). Catch data appears to be available weekly, so it should be possible. (More detailed description of this follows under TOR 5).

There is some indication of conflict between the gill net index and the remaining data sources. The sensitivity runs indicated a somewhat different trend when the index was removed, and the retrospective pattern shows that recruitment events that are (e.g.) large in the last year (when they are only based on the recruitment index) are modified in subsequent years (when they have to match all data). Additional runs were performed to narrow down the cause and it appeared not to be caused by the standardization. It would be a big improvement if age compositions could be provided for the gill net index.

Additional support for the BAM's main results was provided by a biomass model run presented at the review meeting. After the biomass model had been suitably configured it showed the same overall trends as the BAM. This validation by a secondary model is important and could be added to the assessment report. It is important because it demonstrates that the overall conclusions are robust even in a model with a different structure.

The model validation could be improved. Residuals should be perfect (e.g., standard normal) if the model's assumptions are perfectly fulfilled. The presented residuals are not standardized or de-correlated, so it is unclear what constitutes good or bad residuals. When observations are assumed to be independently log-normal, then it would be natural to present standardized residuals at the logarithmic scale. When multivariate distributions are assumed, like it is for composition observations, then the residuals should be de-correlated. This is not simple, but it can be done. One way is to use the so-called one-observation-ahead residuals. These are computed corresponding to each individual observation by transforming the observation with the cumulative distribution function of that observation conditioned on all the previous observations to get =cdf(). If is a discrete integer-valued random variable, then is randomized uniformly between cdf() and cdf(). If the model is correct, then will follow a uniform distribution between 0 and 1. The standard normal quantile function is used to transform these into residuals , which are standard normal if the model is correct (Smith, 1985).

3) Evaluate the assessment findings with respect to the following:

a) Are abundance, exploitation, and biomass estimates reliable, consistent with input data and population biological characteristics, and useful to support status inferences?
b) Is the stock overfished? What information helps you reach this conclusion?
c) Is the stock undergoing overfishing? What information helps you reach this conclusion ?
d) Is there an informative stock recruitment relationship? Is the stock recruitment curve reliable and useful for evaluation of productivity and future stock conditions?
e) Are the quantitative estimates of the status determination criteria for this stock appropriate for management use? If not, are there other indicators that may be used to inform managers about stock trends and conditions?

As described under TOR 2 above, the assessment model provides a consistent description of the data and passes basic model diagnostics. Furthermore, a range of alternative formulations/configurations and data options were evaluated as sensitivities with acceptable results. Hence, the abundance, exploitation, and biomass estimates are reliable, consistent with the input data and population biological characteristics, and useful to support status inferences.

The stock is not overfished and overfishing is not occurring. This conclusion is based partly on the assessment team's suggested reference points (F < M, F < 0.75M, SSB > [0.25SSB at F =0], and SSB > [0.5SSB at F=0]) and partly on a common sense evaluation of the historic time series of F and SSB (SSB and recruitment has increased or at least been stable since about 1995 and the fishing mortality is currently at the low end of the range it has been in for that period).

The stock-recruitment relationship assumed by the assessment model is a Beverton-Holt model. The Beverton-Holt function is a monotone increasing function and this shape-constraint combined with a non-increasing actual estimated stock-recruitment points forces the steepness parameter towards its upper bound (to make the predicted curve horizontal in the observed range). To avoid numerical problems, the steepness parameter was fixed at 0.99 in the assessment model. The estimated Beverton-Holt curve is not informative about the stock-recruitment relationship and is not reliable or useful for evaluation of productivity and future stock conditions (Figure 2).

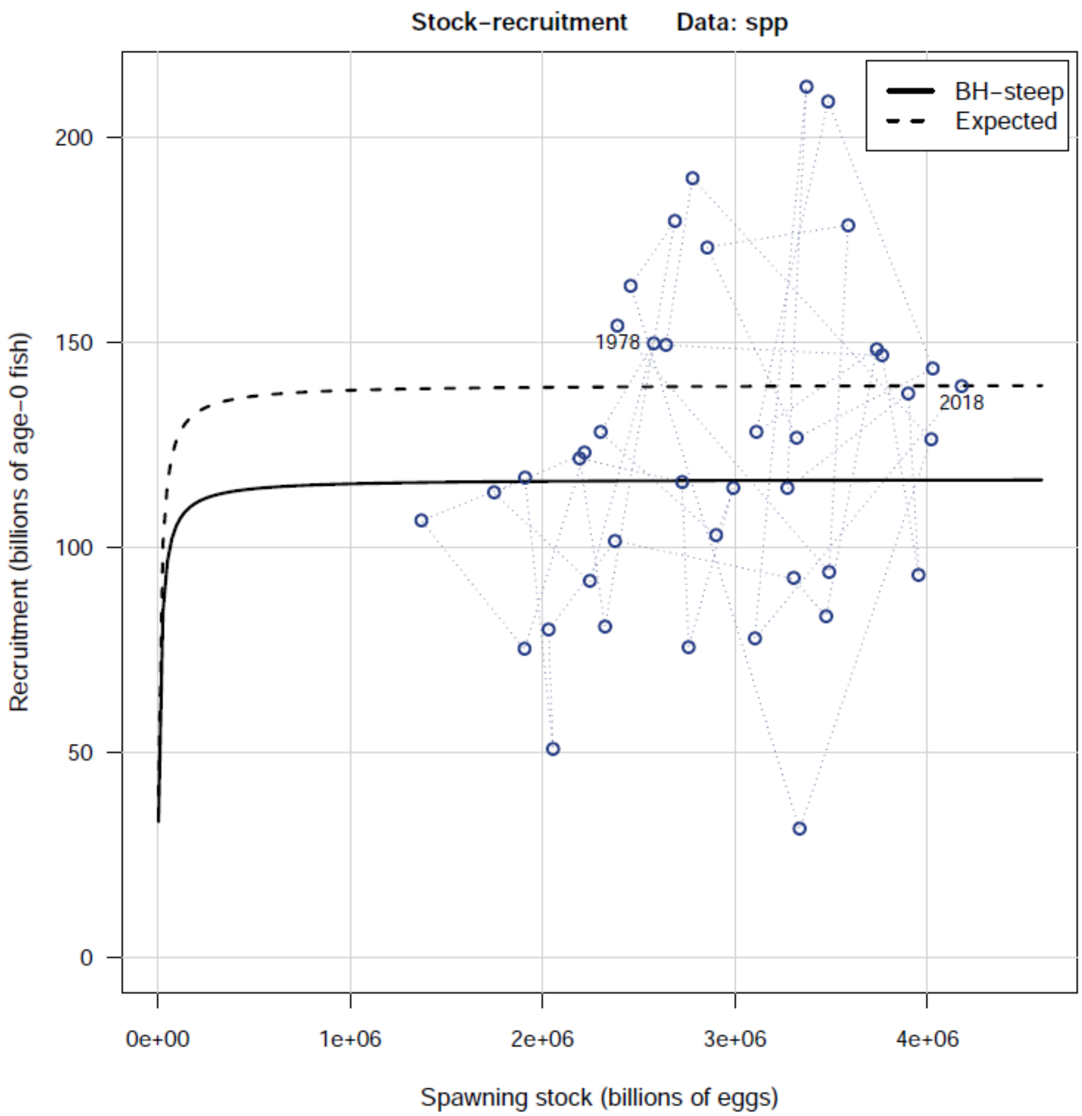

Figure 2. The estimated Beverton-Holt stock-recruitment relationship.

At the review meeting it was requested to fit a Ricker stock-recruitment curve (figure 3) instead of the Beverton-Holt curve (figure 2). The estimated Ricker stock-recruitment relationship is displayed as median (solid blue) and bias-corrected expectation (dashed) in both figures. The Ricker curve is a better fit to the actual estimated stock-recruitment points, and it does not have an artificially fixed steepness parameter. Hence the estimated Ricker curve appears potentially to be more informative about the stock-recruitment relationship and to be more reliable and useful for evaluation of productivity and future stock conditions. However, the range of SSB values observed only covers values of SSB where the recruitment values vary around similar levels. This is not uncommon, but it means that the actual shape of the estimated stock-recruitment curve is uncertain. The important interval of the stock-recruitment curve (from SSB=0 and up to SSB values giving the current level of recruitment) is unobserved.

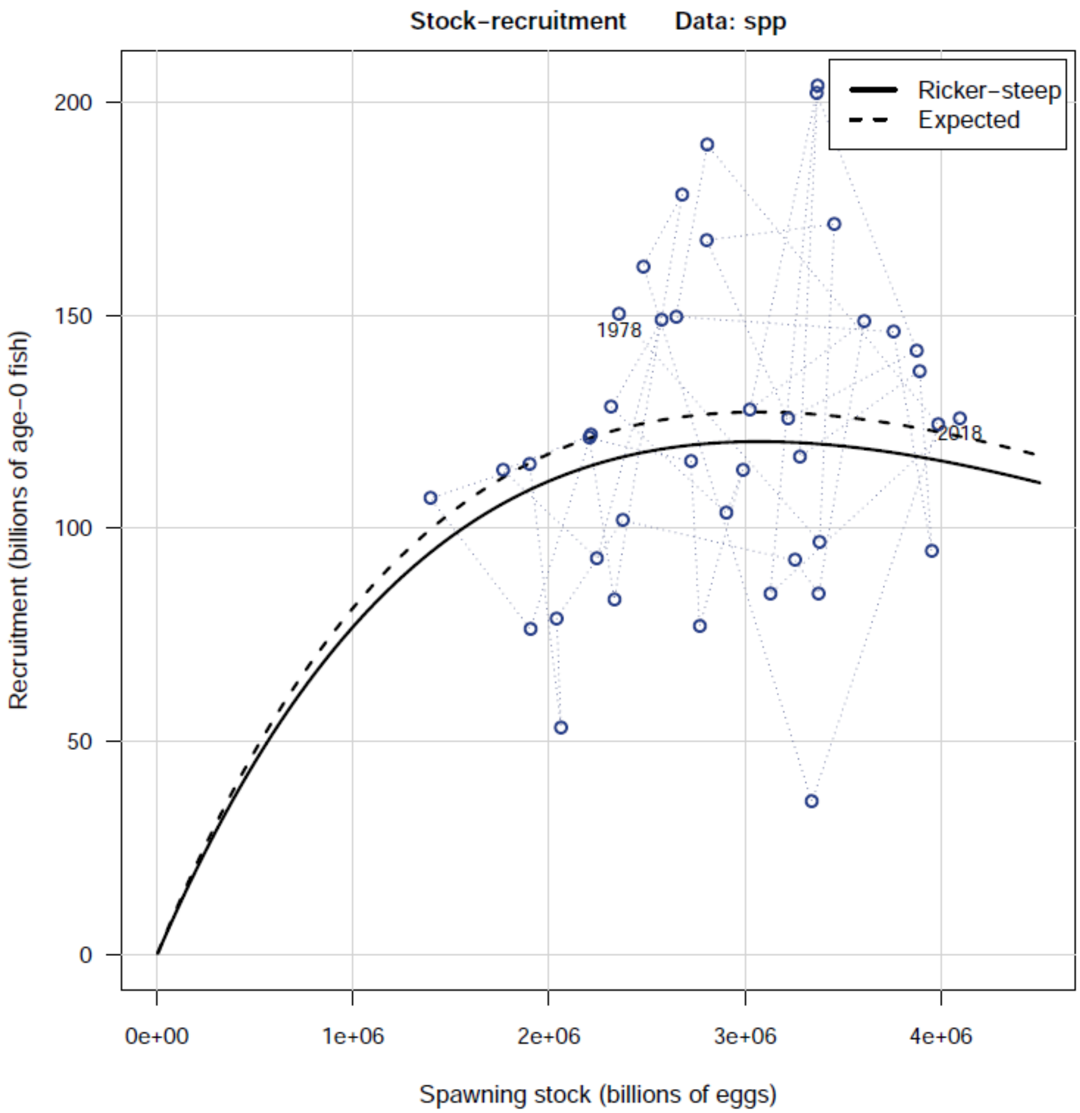

Figure 3. The estimated Ricker stock-recruitment relationship.

The MSY-based reference points estimated via the fixed steepness for Beverton-Holt stock-recruitment, or via the yield per recruit proxies were not useful. (Fishing mortality reference points were above 10. Further comparing figures 7.52 and 7.53 shows that fishing at only around 10% of F30% resulted in SSB to decrease below SSB30% in 1992, which is not logical and hence supports the conclusion that these reference points are not useful).

The presented reference points (F < M, F < 0.75M, SSB > [0.25SSB at F=0], and SSB > [0.5SSB at F=0]) are crude measures and not suggested for management. Reference points are often the `soft underbelly' of fishery management. MSY-based reference points do require some knowledge about the relationship between stock size and recruitment and often the important part of this relationship is unobserved. This means that the calculated reference points heavily rely on the parametric shape (e.g., Ricker, Beverton-Holt, or Hockey-stick) of the assumed relationship. In this common situation there is no solution that will lead to certain, well-established, and objective reference points, and it is important to communicate this uncertainty.

The procedure used in ICES (ICES 2017) could be used as inspiration to derive reference points for the Gulf of Mexico Menhaden. The ICES procedure involves estimating (or assigning) a Blim as the SSB point below which recruitment is impaired and then simulate the stock forward using a stock-recruitment relationship with an appropriately wide uncertainty envelope. Then optimize the sustainable yield with respect to fishing level. The reference points derived from this procedure are used to inform managers for many managed ICES stocks. Using this procedure would accommodate many of the valid suggestions in the public comment by Doug S Butterworth.

4) Consider how uncertainties in the assessment, and their potential consequences, are addressed.

a) Comment on the degree to which methods used to evaluate uncertainty reflect and capture the significant sources of uncertainty in the population, data sources, and assessment methods
b) Ensure that the implications of uncertainty in technical conclusions are clearly stated.

In the ideal scenario, a stock assessment model is formulated such that all important model parameters are estimated from the observations. Then the observation noise is propagated into correct uncertainty estimates of the estimated model parameters and derived quantities of interest (e.g., SSB and F). This is uncommon for fish stock assessment models, because it is most often necessary to fix some model parameters (e.g., natural mortality).

The assessment panel correctly identified that simply supplying the standard uncertainties from the model estimation procedure (derived from the inverse Hessian matrix of the objective function at its minimum) is not realistically representing the uncertainty. The assessment panel went on to derive a Monte Carlo bootstrap procedure, where observations were simulated along with values of the fixed model parameters. Then for each new set of observations the model provided new estimates and thereby the distribution of these new estimates provided a new quantification of the uncertainties (e.g., confidence intervals).

This derived procedure does capture more uncertainty, but the interpretation of this new uncertainty estimate is not straightforward. First there are a few technical caveats:

a) The observations are simulated assuming the subjectively assigned level of uncertainty from the base run, so the procedure still is conditioned on those arbitrary assigned values.

b) The natural mortality and selection parameters are simulated from a joint uniform distribution, which basically assumes that all values are equally likely.

c) Non-converging runs are removed, but it seems unlikely that convergence problems should occur at random.

d) The range of uncertainty for natural mortality and selection parameters are taken from the profile likelihood in the base run, which means that the ranges are consequences of the observation noise, so the bootstrap procedure is strictly speaking including the same uncertainty twice.

e) Some fixed parameters (for growth and likelihood weighting) are not bootstrapped over, so the results are still conditioned on these values.

Some of the issues a)-e) will lead to larger uncertainties and some to smaller uncertainties, hence the interpretation is problematic. Furthermore, there is a possibly more theoretic objection, which is that, if a frequentist interpretation of the uncertainty statements is intended, then all uncertainty should originate from hypothetical repeated experiments, so we should not simulate the quantities which would be constant under repeated experiments (e.g., natural mortality).

The most positive interpretation of this bootstrap procedure is that it is an ad-hoc combination of sensitivity runs and uncertainty. This reviewer would prefer to keep uncertainty estimation and interpretation separate from the sensitivity runs. Sensitivity runs are often not equally likely runs, and hence should not be combined and weighted equally.

In addition to the bootstrap the assessment team presented a range of different sensitivity runs and supplemented them with additional reviewer requested runs, which clearly illustrated the range of uncertainty, and which parts of the model are sensitive to which inputs. The relative results appeared to be fairly robust to the different fixed inputs, and the absolute values changed as could be expected with respect to direction and range. The one exception was the gill net index, which appears to conflict with other data sources. Uncertainty with respect to age-readings was explored by double-reading studies (both within and among readers) and the ranges of the age reading errors were used to define sensitivity runs. This gave a clear illustration of the effect of age reading on the main results of the model.

The assessment team has put a lot of effort into exploring the uncertainties. They have clearly stated what has been done and have illustrated the effect on main model results.

5) Consider the research recommendations provided by the Data and Assessment workshop and make any additional recommendations or prioritizations warranted.

a) Clearly denote research and monitoring that could improve the reliability of, and information provided by, future assessments.
b) Provide recommendations on possible ways to improve the SEDAR process.

This reviewer supports the research recommendations from the data workshop and from the assessment panel collected in section 10.0 of the assessment report and the joint recommendations in the review panel's summary report.

In terms of data collection, high priority should be given to working with state agencies to collect age data from the gill net surveys. It is possible that getting the age data will resolve the conflict between the gill net survey and the other data sources and decrease the overall uncertainty of key outputs (e.g., fishing mortality and SSB).

In terms of modelling efforts, it should be investigated if the timing of the fishing season influences the estimates, especially with respect to the changes in fishing intensity throughout the year. Current model assumes that the fishing mortality is constant over the year, which it is not. This is possibly not important if the fishing pressure is applied at the same time every year, but if the fishing season changes from year or if the allocation of fishing pressure within fishing season changes (with respect to timing), then a different fraction of the natural mortality is applied before fishing, and that can have some effect with respect to the predicted catch for a given fishing mortality. A rough quantification of this effect could be obtained by applying the observed variation in fishing pressure to a fine time-scale simulation of catch under realistic parameters (a short R-script). The script could for instance be set up such that initial N, yearly catch, and M are fixed, then search for fishing mortality F that corresponds to the yearly catch if a) applied uniformly over the year , b) as the earliest observed fishing pattern, or c) as the latest observed fishing pattern. Then the possible gains of moving to a fine time-scale assessment model would be illustrated. If the differences are sufficiently large, then a finer time-scale assessment model should be considered. A finer time-scale model should be possible, because of the fine time-scale availability of the data on catch.

Any effort to improve estimates of natural mortality should also be supported. One such effort would be the re-analysis of the tag based natural mortality estimates from Ahrenholz.

Two efforts could be combined. At the meeting there were discussions about the assumed selectivity pattern. The doming was explained by the use of spotter planes and that they were more likely to spot and report larger schools, which would likely not be of the oldest ages. It was also discussed if it was realistic to assume constant selectivity over time. This could be investigated by trying models with more flexible selection. One option could be state-space assessment models (Nielsen and Berg 2014), but spline approaches are also possible. Exploring such a model would also give a second validation of the results seen, which is currently only provided by the biomass model (ASPIC).

It is clear from the discussions at the review meeting that some effort should be given to derive relevant and useful reference points. This is not an easy task, and often the outcome is more uncertain than one had hoped for, but this reviewer feels that some guidance could possibly be drawn from the ICES guidelines (ICES 2017). There is a detailed description with corresponding software (EQsim) and it should not be too time consuming to try out.

The SEDAR process for this meeting was - as far as this reviewer is concerned - close to optimal. The meeting was very efficient. The meeting was only two days, which was sufficient time to present and discuss this assessment. Compared to other longer review meetings, the main downside to a short meeting like this was that there were fewer evenings/mornings available for the reviewers to produce text paragraphs for the joint summary report. This means more work on the joint report after returning. During the meeting the assessment panel was able to quickly answer questions and produce new runs and changes in the model. The reviewers were sufficiently diverse in experience to cover most subjects. The location of the meeting was convenient, and the support staff was excellent.

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

The assessment is in an acceptable state (in this reviewer's evaluation), so no absolute requirements should be made here. The following could however be considered:

The effect of the timing of fishing could be considered (as detailed above).

It could be investigated if the detailed ICES guidelines for MSY based reference points could provide plausible reference points.

The model validation provided in the report could be improved. Residuals should be provided at the logarithmic scale where independent normal distributions are assumed. It is more difficult, but possible, to provide useful residuals when a multivariate distribution is assumed (like is done for the compositions). Useful residuals should be perfect (e.g., standard normally distributed) if the model assumptions are perfectly fulfilled. See guidance under TOR 2.

7) Prepare a Peer Review Summary summarizing the Panel's evaluation of the stock assessment and addressing each Term of Reference. Develop a list of tasks to be completed following the workshop. Complete and submit the Peer Review Summary Report in accordance with the project guidelines.

This report meets the above term of reference.

References

ICES 2017. fisheries management reference points for category 1 and 2 stocks (DOI: 10. 17895/ices.pub.3036). http://ices. dk/sites/pub/Publication%20Reports/Advice/2017/2017/12.04.03.01_Reference_points_for_category_1_and_2.pdf

Nielsen A and Berg CW 2014. Estimation of time-varying selectivity in stock assessments using state-space models. Fisheries Research 158, 96-101.

Smith J 1985. Diagnostic checks of non-standard time series models. J Forecast 4(3) :283–291

Appendix 1: List of documents.

Gulf Menhaden document list.

Document #	Title	Authors
Final Assessment Report		
SEDAR63-SAR1	Assessment of Gulf Menhaden	To be prepared by SEDAR 63
Reference Documents		
SEDAR63-RD01	Genetic Population structure of the Gulf Menhaden (Brevoortia patronus) Presentation from SFFMC Menhaden Advisory Committee & GSMFC Spring Meeting	Anderson 2016
SEDAR63-RD02	The Selection and Role of Limit Reference Points for Pacific Herring (Clupea pallasii) in British Columbia, Canada	Canadian Science Advisory Secretariat 2017
SEDAR63-RD03	Data weighting in statistical fisheries stock assessment models	Francis 2011
SEDAR63-RD04	A Review of Biological Reference Points in the Context of the Precautionary Approach	Gabriel and Mace 1999
SEDAR63-RD05	A new role for MSY in single-species and ecosystem approaches to fisheries stock assessment and management	Mace 2001
SEDAR63-RD06	NPFMC Groundfish Species Profiles 2015	NPFMC 2015
SEDAR63-RD07	Fisheries for small pelagic species: an empirical approach to management targets	Patterson 1992
SEDAR63-RD08	Status of the Pacific Coast Groundfish Fishery: Stock Assessment and Fishery Evaluation	PFMC 2016
SEDAR63-RD09	A spatial model for fishery age-selection at the population level	Sampson & Scott 2011
SEDAR63-RD10	GDAR 02: Gulf Menhaden Stock Assessment - 2016 Update	Schueller 2016
SEDAR63-RD11	Model-based estimates of effective sample size in stock assessment models using the Dirichlet-multinomial distribution	Thorson et al. 2017
SEDAR63-RD12	The Gulf Menhaden Fishery of the Gulf of Mexico: A Regional Management Plan, 2015 Revision	VanderKooy and Smith 2015
SEDAR63-RD13	Technical documentation of the Beaufort Assessment Model (BAM)	Williams and Shertzer 2015

Appendix 2: Statement of work for Anders Nielsen.

Performance Work Statement (PWS)
National Oceanic and Atmospheric Administration (NOAA)
National Marine Fisheries Service (NMFS)
Center for Independent Experts (CIE) Program
External Independent Peer Review

SEDAR 63 Gulf of Mexico Menhaden 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[1]. Further information on the Center for Independent Experts (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

1 http://www.cio.noaa.gov/services_programs/pdfs/OMB_Peer_Review_Bulletin_m05-03.pdf

and reliability of assessments. SEDAR 63 will be a CIE assessment review conducted for GSMFC Gulf of Mexico Menhaden. The Review Workshop will provide an independent review of the Gulf of Mexico Menhaden stock assessment. 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 stock assessed through SEDAR 63 is within the jurisdictions of the Gulf States Marine Fisheries Commission and the states of Florida, Alabama, Mississippi, Louisiana, and Texas.

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 PWS, OMB guidelines, and the TORs below. The reviewers shall have a working knowledge in the application of fisheries stock assessment processes and results, 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.

Tasks for Reviewers

1) Review the following background materials and reports prior to the review meeting:

SEDAR 63 Workshop Reports and background documents will be available on the SEDAR website at the links below.

http://sedarweb.org/sedar-63
http://sedarweb.org/sedar-63-review-workshop-

2) Attend and participate in the panel review meeting. The meeting will consist of presentations by NOAA scientists, other members of the analytical team 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-national-registration- 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 New Orleans, LA.

Period of Performance

The period of performance shall be from the time of award through January 2019. The CIE reviewers' 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
No later than Oct. 23, 2018	Contractor provides the pre-review documents to the reviewers
November 6 - 7, 2018	Panel review meeting
Approximately 4 weeks 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.

Article I. 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

Julia Byrd - SEDAR Coordinator
4055 Faber Place Drive, Suite 201
North Charleston, SC 29405
(843) 571-4366
julia.byrd@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 Statement of Work
Appendix 3: Panel membership or other pertinent information from the panel review meeting.

Appendix 3: List of participants.

REVIEW PANEL		
Will Patterson	Review Panel Chair	GSMFC Appointee
Joe Powers	Reviewer	GSMFC Appointee
Matt Cieri	CIE Reviewer	CIE
Anders Nielsen	CIE Reviewer	CIE
Kevin Stokes	CIE Reviewer	CIE
ANALYTICAL REPRESENTATIVES		
Amy Schueller	Lead analyst	SEFSC Beaufort
Robert Leaf	Assessment Team	GCRL
Ray Mroch	Assessment team	SEFSC Beaufort
COUNCIL AND AGENCY STAFF		
Julia Byrd	Coordinator	SEDAR
Kimberly Cole	Admin	SEDAR/SAFMC
Steve VanderKooy	GSFMC	GSMFC