Independent Peer Review Report on the

SouthEast Data, Assessment, and Review (SEDAR) 63
Gulf of Mexico Menhaden

Prepared for:
The Center for Independent Experts

Dr Kevin Stokes
Stokes.net.nz Ltd
59 Jubilee Rd
Khandallah
Wellington 6035
New Zealand
Ph: +64 (04) 973 7305
E-mail: kevin@stokes.net.nz

EXECUTIVE SUMMARY

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.

The South East Data, Assessment, and Review (SEDAR) 63 for Gulf of Mexico menhaden took place at the Sheraton Four Points, New Orleans, LA from 6-7th November 2018. The meeting was well organized and administered and was open and transparent. The Review Workshop (RW) Panel was able to reach consensus on all issues and the individual inputs to the RW Panel Report were provided according to the schedule for compilation by the Chair. This report is an individual report that largely reflects the Panel Report although with some additions and minor departures.

For the Gulf of Mexico menhaden, landings data are excellent and while there is some concern about the lack of age composition data for the key gillnet survey index, the availability of both juvenile and adult indices is good. Historic difficulties with, and lack of clear protocols for, ageing were not considered in detail but the attention to ageing moving forward is positive. The single species model has been well explored and appears to be reliable as a basis for informing decision making though the lack of continuity in defining accountability measures associated with F_{SPR} targets and limits is a problem. Against traditional single species standards, the fishery appears neither to be overfished nor experiencing overfishing. Rather than focusing on means to define status (overfished/overfishing), it is suggested the focus should be i) on pragmatic means of informing accountability measures, and ii) possibly use of management strategy evaluation to develop management procedures that meet defined performance criteria.

BACKGROUND

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.

Menhaden (GoM M)

Gulf of Mexico menhaden (Brevoortia patronus) is a clupeid species, distributed from southwest FL to the Gulf of Campeche, Mexico, but centred and ubiquitous in the northern Gulf of Mexico from western FL to eastern TX. There is no evidence of substructure within this central area of concentration. Menhaden are filter feeders with juveniles favouring phytoplankton and adults favouring zooplankton. Menhaden form dense schools near to the surface, particularly in warmer months, and are prey for many coastal predators. The lifespan of Gulf menhaden is possibly of the order of six or more years with very few fish at age 6 observed in the fishery. Previous assessments have assumed maturity at two years of age and with fecundity increasing with length/age; this assessment used more recent estimates of 80% maturity at age 1 and 100% and age 2.

Commercial fisheries for menhaden were developed after WWII when companies involved in Atlantic Menhaden fisheries moved in to the Gulf of Mexico. Operations increased rapidly between 1948 and the late 1950s, and by 1959 the annual commercial catch had increased to over 300,000 mt. The fishery continued to expand through the 1960s and 70s, reaching a peak in the 1980s with catches approaching 1,000,000 mt. Since the early 1990s catches have fluctuated in the range 400,000-600,000 mt with catches in the 2000s averaging near 500,000 mt. In 2011, catches reached 613,000 mt while in 2016 and 2017 catches have been below 500, 000 mt. During the 1990s the number of operating companies, processing plants and vessels declined. The operational context has been stable for the past decade though the two operating companies have in recent years been bought by major overseas companies with interests in expansion.

The Gulf menhaden fishery has been managed under a regional Fishery Management Plan (FMP) since 1978. The fishery assessment in 2013 (SEDAR 32) estimated the stock to be not overfished and not subject to overfishing, though no agreed benchmarks were available for reporting against. The most recent revision of the FMP (SEDAR63-RD12) still does not include benchmarks although Gulf menhaden is considered a key, ecologically important species within the Gulf ecosystem. However, the Gulf States Marine Fisheries Commission (GSMFC) has adopted F_{SPR} reference points as a basis for accountability measures. Specifically, if two years consecutive harvest should exceed $F_{35\%}$ (estimated as 663,583 mt) or any single year should exceed $F_{30\%}$ (680,765), then an assessment update would be requested. Regardless, a stock assessment cycle of five years has been adopted. The accountability measures have not been triggered and SEDAR63 is the scheduled five yearly updated assessment, following SEDAR32.

REVIEW PROCESS

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.
d. Reviewers shall provide a critique of the NMFS review process, including suggestions for improvements of both process and products.

The South East Data, Assessment, and Review (SEDAR) 63 for Gulf of Mexico menhaden took place at the Sheraton Four Points, New Orleans, LA from 6th to 7th November 2018. Terms of Reference (ToR) for the stock reviews are given in Appendix 2, Annexes 2a and 2b.

Participants in the review are listed in Appendix 3. The SEDAR Panel comprised a Gulf States Marine Fisheries Council (GSMFC) SSC appointed Chair (Patterson), a (GSMFC) appointed reviewer (Powers) and three CIE reviewers (Cieri, Nielsen and Stokes). Notification of the meeting and dissemination of papers followed closely the schedule laid out in the CIE Statement of Work (see Appendix 2). Materials were provided in using Google Drive (see Appendix 1). Public comments were made using the SEDAR website. Overall, administration of the review was sound.

The SEDAR Panel was tasked with providing inputs to the Panel report as outlined by the chair during the opening session. The chair assumed overall responsibility for the Panel report with draft sections due from members by 26th November. Contributions for text on all ToR for both assessments were split between the three CIE reviewers and GSMFC-appointed reviewer.

The meeting followed the general outline of the draft agenda (Appendix 2, Annex 3) but with sufficient flexibility to allow necessary responses from the stock assessment team. In my opinion, the meeting was well run and Panelists, Analytical Team members, and the public were afforded proper opportunities for input and comment. I am not aware of any problems with notification of the meetings and interpret from the presence of stakeholder representatives, and lack of complaint, that notification was appropriate. All participants were able to participate throughout the meeting and opportunity was explicitly and regularly given by the chair for input. Many participants other than Panelists and Analytical Team members contributed usefully to discussion, and I believe that all were provided appropriate opportunity for involvement both during the Panel meeting and during extra-mural discussions. Enough ti me was provided to look in reasonable detail at data inputs and modeling decisions and to contemplate assessment outputs. Although in general I consider the time tight for the tasks at hand, I am confident that the SEDAR 63 resulted in informed and reasonable conclusions.

REVIEWERS ROLE IN THE REVIEW ACTIVITIES

The role of the reviewer is set out in the CIE Statement of Work, Attachment A, attached here in Appendix 2, Attachment A. CIE reviewers are tasked with producing an independent report to the CIE. As part of the stock specific ToR (ToR 7), the reviewers are additionally tasked with contributing to a Peer Review Summary Report for the review.

In addition to conducting necessary pre-review preparations, including the review of background material and papers provided in advance by the SEDAR project and a conference call, I (Stokes) participated in all discussions and contributed a brief section on data (ToR 1) to the draft Summary Report, provided to the Chair on 23rd November following agreement with other panelists. I also considered and responded to draft sections forwarded to me by the other CIE experts (on ToR2, 4, 5 and 6).

SUMMARY OF FINDINGS BY STOCK

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.

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.
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.

ToR 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 assumed stock structure is appropriate, and uncertainty was acknowledged with respect to possible differences in dynamics and trends between the eastern and western portions. There was good discussion on the issues and, as at SEDAR 32, I see no cause for concern with the

single stock structure assumed for the assessment.

The assessment report and review workshop (RW) presentations describe all biological information used. I note the update to the maturity schedule due to Brown-Peterson et al. (2017). The schedule is quite different to that used previously, with 80% maturity at age 1 compared to zero as previously used. This should have implications for YPR and SPR-type estimates, though not necessarily of importance when ratios are used to define status. The use of the updated schedule and continued use of fecundity (often still notated as SSB) is reasonable, though I note considerable inconsistency in labelling as SSB and fecundity, which can be confusing.

The methods used to estimate the pattern and scale of time invariant M are well described. The choice of the scaled Lorenzen estimate of M-at-age is well justified and continues the practice adopted at SEDAR32. The assessment report comprehensively details the methods considered and rationale for use of the Lorenzen form. During the RW, all M methods/estimates were considered. My only concern about M is that it is highly unlikely to be time invariant given the role of menhaden in the multispecies fish complex of the Gulf. My concern is that there could be high inter-annual variability in M at age which would ideally be incorporated into uncertainty characterisation. Absent a multispecies model, it is difficult to see any way of incorporating this but note the assessment group is clearly aware of and has considered these issues and possible approaches. Discussions in the RW on this issue included consideration of higher M estimates based on Atlantic multispecies modelling and recent work using mark-recapture models for Atlantic menhaden resulting in estimates of M three times higher than previous estimates (from a University of Maryland masters thesis by Liljestrand, 2017). It is unclear how to incorporate this information in to the Gulf menhaden assessment but is noteworthy that at least the low M variant used for sensitivity testing of the base case scaled age-profiled M could be highly pessimistic while the high variant could be more meaningful.

Ageing of menhaden using scales has been described in SEDAR 32, since a change in personnel and equipment has recently been made after an extended period of a single, continuous reader and ageing equipment. Despite difficulties, it is clear that ageing is reliable especially at ages 0-2 which are the key ages in the assessment. With changes in personnel and equipment from 2019 onward, however, reading methodologies will change and comparison readings are being made. My understanding is that preliminary results suggest consistency with older readings. There is no mention of this issue in the assessment report section on research recommendations and I presume it is not considered a major issue.

The assessment team, however, has reported on estimates of ageing estimation error made using Punt et al.s (2008) Agemat method using i) a large sample size (n=5275) of scales reread by the same reader from a period spanning four decades, and ii) a small sample (n=78) of scales read by two newer readers. While there are differences in the estimates of error between the two sets, they are mostly at ages 3 and 4 and the differences at younger ages are small. It is not clear if the differences are due to the reader(s)/methods used or primarily a consequence of sample size. Both ageing error matrices have been used for sensitivity testing (ToR 4) while

spreading ages must and does impact on fecundity, recruitment, and F estimation it does not appear to impact status determination, at least as considered from a non-mandated single species perspective. It would, however, potentially, create additional uncertainty in F_{SPR} estimates used to trigger accountability measures. Noting the target and limit F_{SPR} used are already very close in absolute terms, this could lead to some sensitivity which is not considered.

Removals data are good for Gulf menhaden due to lack of multiple fleets and sectors (more than 99% of removals are commercial and from two companies) and a long-term, high quality logbook system and port sampling. The reduction fishery is well-sampled and the lack of composition data for the very small bait and recreational catches is not of concern. The reduction fishery sampling system was well reported in the assessment report and at the RW. I see no areas of concern. It is good to see a fishery with such sound fundamental data keeping underpinning assessment and management. Fishery-dependent data considerations and decisions by the Data and Assessment Workshop (DW and AW) teams include exploration of alternative measures of nominal fishing effort. The continued use of vessel-ton-weeks (VTW) was well explained during the Review process. No fishery CPUE is used in the stock assessment; given difficulties associated with interpretation of CPUE for pelagic purse seine fisheries, this is reasonable a priori, but it is good to see continued consideration of the issue as part of the data and assessment processes.

As at SEDAR 32, the DW and AW teams thoroughly considered the large number of potential abundance indices, though only the Louisiana (Mississippi and some Alabama) seine and Louisiana-only gillnet indices were ultimately used in the assessment. All other surveys were excluded a priori for a variety of reasons which have been fully considered and explained previously and in the recent assessment report. While it would have been possible to examine indices and exclude them during the modelling process, using e.g., likelihood profiling, I am comfortable with the approach taken by the DW and AW teams and think it is appropriate to judge the utility of indices on a priori considerations. The teams did a good job in this respect. Considering all potential indices at the modelling stage would have been time consuming and likely unproductive.

The gill net index in particular is influential in the assessment (intentionally) but seemingly conflicts somewhat with other fishery and fishery-independent data. Uncertainty was acknowledged by the DW and AW and was considered during review (see ToR 4) by considering a run with a non-standardised index and one truncated to the period for which length composition data are available. While there are length composition data for the index, there are no corresponding age compositions, though samples exist and could be read. Given the weight put on the index and apparent inconsistency with other data, it is unfortunate that ageing has not been done for the Louisiana gillnet survey (see ToR 5).

Overall, considering ToR 1 (a-d), I am confident that the DW and AW made reasonable, sound and robust decisions about data (ToR 2), acknowledged uncertainties (ToR 4), and applied data correctly. The data used reasonably support the assessment and findings (ToR 3).

ToR 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?

Overall, the assessment models used are appropriate, configured properly and used consistent with standard practices. The methods used are appropriate for the available data.

The Beaufort Assessment Model (BAM) was used as the principal assessment tool. The BAM, implemented in AD Model Builder software (Fournier et al., 2012), is structured to allow implementation of forward projecting, statistical catch-at-age assessment models, also allowing matching to length composition observations through a growth function. Use of the BAM permitted the inclusion of all available types of data. Decisions on a priori data inclusion and exclusion were carefully considered and explained by the DW and AW and are considered at ToR 1. The specified assessment model used standard approaches to predicting landings and modelling recruitment, and the BAM allowed a thorough exploration of catchability and selectivity options.

The base case run included commercial and recreational landings, age and length composition data and two indices of abundance, one each representing age 1 and age 2 fish (from seine and gill net surveys respectively). Natural mortality was assumed constant through time but age-specific based on the method of Lorenzen (1996) and scaled based on tagging studies. Steepness was fixed at 0.99 (cf 0.75 at SEDAR32). Selectivities and catchabilities were estimated as constant through time for two commercial selectivity blocks and a single gillnet survey block. Commercial selectivities-at-age were fixed except for age-1 which was estimated. There is little difference in the estimated age 1 selectivity in the two commercial blocks. The assessment team explored free selectivity parameter estimation but there is little or no information in the data to allow this.

The model was fit to the data using appropriate methods, consistent with standard practice. Analyses included iterative reweighting using the method of Francis (SEDAR63-RD03) and exploration of a variety of data configurations and parameterisations. The modelling processes and decision making resulting in a proposed base case run and sensitivity testing are well described in the assessment report, which includes information on Likelihood components, weighting, standard deviations of the normalized residuals (SDNRs) by data component and weight, likelihood profiles, etc. Further diagnostics were made available and elaborated during the SEDAR 63 RW. The modelling procedures adopted appear to be robust. Landings were fit

closely, as were age composition data. Landings and indices were fit using lognormal likelihoods. Age composition data were fit using robust multinomial likelihoods.

Abundance indices were reasonably well fit though there is some conflict between the gill net index and other data sources. The reported sensitivity runs indicated a somewhat different trend when this index was removed, and the retrospective pattern shows that recruitment events that are large in the terminal year (when they are only based on the recruitment index) are adjusted downwards in subsequent years (when they have to match all data). Two runs were requested during the RW to investigate this conflict: i) a run with no gill net index, and ii) a run with the index truncated from 1996 (as prior years had no associated length compositions). However, any results depend critically on weighting and there was little time to explore this issue in detail. There is little obvious benefit to excluding the early index points. Visual examination of RW outputs against the base case is difficult and given the lack of concomitant re-weighting, over-interpretation is not advisable. Improving model fit in detail is always possible but the key issues are i) is the age 2 gillnet index credible and, if so, ii) how much weight to put on it relative to the age 1 seine index and composition data? It is not just a statistical/fitting issue but also one of expectations about representativeness. My view is that i) the gillnet standardization appears sound and the index should be used, and ii) given the amount of noise in recruitment and survival to age 1 and 2 (high M but not accounting for annual variability), the age 2 index is key to the assessment of SSB (expressed as fecundity) and should be relatively highly weighted.

As discussed extensively during the RW and considered under research recommendations, reading of existing age samples for the gill net survey to provide age compositions could provide the largest single information boost to the assessment.

Monte Carlo Bootstrapping (MCB) was used to portray uncertainty around model outputs, including status estimates. MCB combines parametric bootstrapping to landings and indices data and resampling from composition data. The Monte Carlo component entails drawing values of M and steepness from externally (post model) specified pdfs. Outputs provided are the quantiles of the distribution resulting from application of the MCB simulations. Each simulation applies a single BAM model using the weights developed for the base case run. No reweighting procedures are used for individual assessment realisations.

The MCB generates a stochastic version of the BAM model by introducing process error to the model components of natural mortality and steepness. Means of management quantities (e.g., MSY, BMSY, FMSY) from the MCB runs do not equal estimates from the base run. As noted at SEDAR 32, the direction of the differences observed between the MCB based estimates and those of the base run are in the direction predicted by Bousquet et al. (2008). FMSY from the MCB runs will be less than the deterministic estimates from the BAM base run, estimates of MSY will be slightly higher and those for BMSY slightly lower. The size of the differences will be

a function of the amount of stochastic error in the model. These differences, however, will not be apparent when looking only at ratio benchmarks.

Additional to the BAM model, an age-aggregated biomass dynamics stock assessment was carried out using the ASPIC software. The biomass dynamics models is considered important to the assessment process as a complementary rather than an alternative analysis, because the catch-at-age model makes fuller use of composition data and represents a more detailed investigation of population dynamics and is hence able to capture higher frequency changes in indices better. The value of the simpler model is in confirmation of the BAM results (or otherwise) at a broad level. The RW noted some potential difficulties with the presented ASPIC runs (e.g., both biomass and F being below those estimated by BAM for the same catches) and alternative runs were provided on request, including a gillnet only run and runs with different production model and starting states. The rerun biomass dynamics model provided a useful comparison with the catch-at-age model that broadly supports trends in biomass and fishing mortality estimated using the BAM.

ToR 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?

All BAM model base case estimates of selectivity, recruitment deviances and unfished abundance are consistent with the data inputs (catches, indices, size and age compositions), given the model structure (single species/area), assumptions (growth, fecundity/maturity, M), and weighting. The RW panel accepted, and I agree, that the BAM outputs are an appropriate basis to support status determination. My main concern is that while the M profile and scaling are reasonably justified, evidence from the Atlantic suggests the scaling may be low and a priori reasoning suggests the inter-annual variability could be higher than captured through sensitivity runs and the pdfs of M used for the MCB. While there was some concern at the RW that the gillnet index may not be consistent with other information, I tend to put higher natural weight on the index as the primary source of information on mature abundance and am comfortable with the weighting used, though recognise the development of associated age composition data would be helpful.

There are no standards set for Gulf menhaden to determine whether it is overfished or experiencing overfishing though there are F_{SPR}-based accountability measures in place. The AW provided estimates and discussion of a variety of standard single species status determinants, all of which suggested the stock was not overfished or experiencing overfishing, though may (depending on definition) have experienced overfishing in the 1980s/90s. Accountability measures currently rely on SPR-based estimates of yield. On a single species basis using accepted SPR-based standards, the stock is clearly neither overfished nor experiencing overfishing. However, given the FMP does not require status determination per se, there is arguably little value in dwelling on the matter. More important for the FMP, given the latest assessment does not provide clear estimates of yield associated with $F_{30\%}$ and $F_{35\%}$, is how should the accountability measures be set? This was not considered by the AW or RW but unlike at SEDAR32, the values are undefined. Absent those estimates, but accepting the new stock assessment, it is unclear how the accountability measures should be estimated/reset. With hindsight, perhaps a continuity run would be useful and a comparison of the SEDAR 32 and 63 model assumptions and fits.

It is difficult to compare the assessments from SEDAR32 and SEDAR 63, but with the exception of new, accumulated data, the main differences are the assumed steepness (which does not appear important in sensitivity analyses), the new fecundity/maturity schedule, and the selectivity at ages 3 and 4. SEDAR63 sensitivity analyses do not show any major patterns, but it is striking that the SEDAR32 base case with terminal year 2011 and the SEDAR63 retrospective with terminal year 2012 (no 2011 is shown) show quite different estimated trends in F and biomass. For example, the SEDAR32 base case (SEDAR 32 assessment Fig 7.37) has a flat F from 1995 onwards and a peak three times as great in the early 1980s, plus a fast drop between 1990 and 1995. The SEDAR 63 retrospective (SEDAR 63 Fig 7.26) shows a more gradual drop from a later peak which is double that of the later period. Conversely, the SEDAR32 base case biomass estimates (SEDAR 32 Fig. 7.29) are fairly stable over the time series except for a small decline in the late 1980s/early 1990s compared to the SEDAR 63 retrospective (Fig 7.29) which has a consistent increase from 1990. The SEDAR32 assessment also shows substantial retrospective patterning compared to the SEDAR63 base case. More comparison of the two base cases may have been useful at the RW to understand current model fits and to unravel issues to do with MSY proxy estimation.

The AW suggests M-based reference points for status determination and could be used for accountability measures. This is pragmatic and with some precedent, but the greatest uncertainty in the assessment is arguably in M and as indications are that M may be underestimated, M-based reference points could be highly conservative.

Perhaps more important, given the ecological role of Gulf menhaden and the clear interest in Marine Stewardship Council (MSC) certification exposed at the RW and through public comment, it would be useful to consider standards as appropriate for low trophic level stocks

and whether the stock assessment can inform these. This is something that might be explored directly or even more usefully using management strategy evaluation (MSE) and development of management procedures. This is a major topic and goes beyond the remit of this report.

The AW considered MSY estimation and the RW discussed at some length alternative stock-recruit approaches that would enable estimation of MSY-related reference points and status determination. The RW Panel report provides detail but I am unconvinced about the utility of this. It may well be possible to use an alternative stock-recruit form to derive status conclusions . However, it is clear that the differences in dynamics between any functional forms that would lead to purportedly useful estimates are in fact all below any observations of abundance/biomass. In my view, it would be far more useful to consider i) pragmatic approaches to setting accountability measures in the FMP based possibly on the suggestions of the AW at Table 7.10; and/or ii) using MSE to develop management procedures for the fishery, noting that in the absence of mandated status criteria, it is performance criteria that are paramount.

ToR 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.

Uncertainty was explored in the assessment modelling using extensive sensitivity runs and likelihood profiling, retrospective analyses and Monte Carlo Bootstrapping (MCB). All the methods used are standard and much used. The AW reported widely on the various analyses and more materials were provided and used in discussion at the RW. The application of methods appears to be comprehensive and appropriately focused. Sensitivity runs as variants of the base case run are numerous and good information was provided on the impacts on fits (through detailed likelihood components and also weighting diagnostics, SDNRs, likelihood profiles, etc). Such runs can only look at what the model structure accommodates and cannot consider structural uncertainties such as alternative stock structures. No such structural uncertainties were identified for menhaden, and the assessment and its outputs have been appropriately and comprehensively considered. Additional sensitivity runs using alternative gill net indices (normative, i.e. non-standardised, and truncated) were made during the RW to investigate apparent conflict between the gillnet index and other data sources (see also ToR 1).

Issues considered in sensitivity runs by the AW include scaling of M, S-R steepness, adjustment of index weighting and exclusion, alternative selectivity assumptions for the commercial reduction fishery, start year, and alternative growth specification.

The MCB is alluded to at ToR 2. A total of 5,000 realisations were made using parameter values drawn from specified pdfs and with the landings, indices and composition data bootstrapped.

However, not all parameters are bootstrapped over (e.g., growth). Realisations were censored due to non-convergence but also by the AW for some runs due to high parameter estimates. A total of 2,557 realisations (51%) were used to compile the final MCB quantile plots. The process for discarding realisations was not discussed in detail and it is unclear precisely what criteria were used beyond non-convergence. Each realisation of the BAM model was run using the iteratively reweighted weights from the base case (it would have been impossible to automate this process for each of the 5,000 realisations). It should be noted that reweighting can have major implications for fitting and parameter estimation and that each realisation may not be feasible, possibly explaining why some realisations did not converge and why some were additionally censored by the AW. The degree to which this may or may not matter is model and data specific, but as all realisations are afforded equal weight in determining distributions of outputs, there is in general need for care in interpreting MCB results.

The RW was comfortable, as am I, that the AW had explored uncertainty to the extent possible and that the characterisation of benchmark trajectories and hence stock status (ToR 3) are suitable for informing management decisions.

ToR 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.

The DW and AW made a number of research recommendations, as did the RW. I am comfortable with the recommendations of the RW. The following represents my opinion as to the most useful research avenues.

I note a key recommendation at SEDAR32 in 2013 related to the Louisiana seine net fishery. Specifically, given the importance of the survey index to the assessment, the panel recommended that the survey return to the former sampling frequency (monthly as opposed to quarterly from 2010) and geographic coverage. It is pleasing to see that the survey reverted to monthly sampling in 2014.

Also, at SEDAR32, it was noted that difficulties with species identification in the TX, FL and AL gillnet surveys precluded their use in abundance indices and that cost-effective methods to improve species identification, including simple genetic approaches, could be usefully developed. At SEDAR63 the DW and AW recommendations included a related item as medium/high priority. The RW assigned this a medium priority with a number of higher priority items. I agree with this priority rating and note that research to improve ageing, M estimation and potential management approaches are a higher priority than expanding the geographic

range of indices.

The SEDAR 63 RW suggested high priority for tagging analysis, stock status benchmarks, examination of alternative stock assessment models, exploration of finer time scales in the stock assessment, further work on ageing error and bias, and ageing of fish sampled in surveys.

The AW suggested high priority be afforded to a new tagging study to allow better estimation of natural mortality, growth, etc. It also suggested a low priority for evaluation of existing, historic tag data. The RW, in contrast, suggested a high priority for evaluation of the historic data and a medium priority to a new (expensive) study. I agree with the RW suggestion and see value in exploring full use of existing data before attempting new studies. I agree fully with the RW also on exploring alternative, simpler models but also of finer-scale models. Simpler models are needed for ground truthing and support of more complex models while finer-scaled models might allow for better use of the unusually fine-scaled landings data while accommodating the fast growth and mortality of Gulf menhaden. Ageing error and bias are key and need exploring, but most importantly, ageing of the gillnet survey age samples is required to help resolve differences in signals between the index and associated size compositions with other data.

On the issue of stock status benchmarks, I do not fully agree with the RW as a whole. As noted at ToR 3, I am unconvinced that fitting alternative stock-recruit relationships in order to force estimation of MSY-related reference points and status estimates is worthwhile. I do agree fully that MSE is worthwhile but see its value not so much as to examine single species reference points as such, but rather as to underpin management procedure development within a structured process to achieve agreed and articulated performance criteria. Unlike federally-managed fisheries with specific mandates requiring MSY- or SPR-type reference points, Gulf menhaden might be managed on wider and alternative performance measures.

The ToR asks for specific recommendations on how to improve the SEDAR process. As stated previously, I consider the process to be well organised and administered and open and transparent. My only minor concern in this specific process is that the ToR appear to be generic for the AW and RW it is not clear, for example, that ToR on overfished and overfishing status are directly relevant for Gulf menhaden, nor indeed on the stock-recruit curve. ToR articulated to relate directly to the FMP and exploration of alternative management approaches may have been more useful.

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

The next assessment is scheduled in five years or sooner should accountability measures be triggered. Critically, given the new assessment, it is unclear if/how the existing measures will be updated. Notwithstanding, from a traditional stock assessment perspective, key areas of

uncertainty/sensitivity are the scaling and variability in M, apparent inconsistency between the gillnet index and other data, and ageing error/bias. Any work to refine M estimates would be a huge step forward. Further consideration of estimates from the Atlantic menhaden would be useful but the evaluation of existing tagging data and possibly new tagging studies would be most informative. Ageing of existing and future samples from the gillnet survey could provide greater confidence in the model but is unlikely to change results fundamentally unless the index were down-weighted something I would not advise.

Improvements to data can certainly be made and uncertainty in model outputs can probably be reduced. However, given the stock history, it may still not be possible to estimate traditional MSY-related reference points or avoid the problems of determining and choosing benchmarks. Improving data might result simply in being more precisely wrong while an alternative objective , perhaps using an MSE approach and management procedures could be to be roughly right.

Further exploration of simpler (e.g., ASPIC) models would be useful but is in any case part of the assessment process. I am not sure that it requires special mention. Further exploration of finer-scaled models (in time, not space) may provide insight and reduce uncertainty within an assessment, but whether used for assessment purposes or not could be useful in operating model development to be used with MSE I would contemplate a fine-scaled operating model, possibly with multispecies/ecosystem components, and management procedures using simple data-based rules to set annual and/or real time constraints. Robustness testing of management procedures in an MSE framework could also help to provide insight as to the value of information and to focus research.

GENERAL CONCLUSIONS AND RECOMMENDATIONS

The Gulf of Mexico menhaden stock assessment was carefully considered and conducted and well reported. The Analytical Team is clearly well on top of the many issues and provided excellent materials and presentations, as well as responding well to requests during the RW.

The methods used are standard in the region, but it is unclear why BAM and MCB are used in preference to other widely used and more integrated approaches. There is some risk in the BAM plus MCB approach, and it would be useful to see a clear explanation as to why it is used and what advantages or disadvantages may result. Putting this issue aside, the implementation of the BAM and MCB was robust and careful.

For Gulf of Mexico menhaden, landings data are excellent and while there is some concern about the lack of age composition data for the key gillnet survey index, the availability of both juvenile and adult indices is good. Historic difficulties with, and lack of clear protocols for, ageing were not considered in detail but the attention to ageing moving forward is positive. The

single species model has been well explored and appears to be reliable as a basis for informing decision making though the lack of continuity in defining accountability measures associated with F_{SPR} targets and limits is a problem. Against traditional single species standards, the fishery appears neither to be overfished nor experiencing overfishing. Rather than focusing on means to define status (overfished/overfishing), it is suggested the focus should be i) on pragmatic means of informing accountability measures, and ii) possibly use of management strategy evaluation to develop management procedures that meet defined performance criteria.

APPENDIX 1

BIBLIOGRAPHY

Prior to the Workshop, referenced materials were provided via Google Drive. The complete document list is included in the table below.

During the workshop multiple presentations were given, and additional materials were provided on request, including further background documents and presentations as well as responses to Panel requests. All files were made available using Google Drive via an open Wi-Fi connection throughout the meeting. Wi-Fi access was generally adequate.

REFERENCES

The supplied reference list is shown in the table below. Additional references are:

- Bousquet, N., T. Duchesne, and L. Rivest. 2008. Redefining the maximum sustainable yield for the Schaefer population model including multiplicative environmental noise. Journal of Theoretical Biology 254:6575.
- Brown-Petersen, N., R.T. Leaf, A.M. Schueller, and M.J Andres (2017) Fishery Bulletin 115 (3):284-299.
- Fournier et al. (2012) see: https://tandfonline.com/doi/abs/10.1080/10556788.2011.597854
- Lorenzen (1996) see: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1095-8649.1996.tb00060.x
- Punt et al. (2008) see: https://www.researchgate.net/publication/233684145_Quantifying_age-reading_error_for_use_in_fisheries_stock_assessments_with_application_to_species_in_Australia's_southern_and_eastern_scalefish_and_shark_fishery

SEDAR 63
Gulf Menhaden
Workshop 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
SEDAR63-RD14	Fishery Models	Shertzer et al. 2014

SEDAR63-RD15	Gulf menhaden (*Brevoortia patronus*) fishery-independent catch-rate trends for Louisiana	West and Zhang 2018

APPENDIX 2

Attachment A: Performance Work Statement for Dr. Kevin Stokes

External Independent Peer Review by the Center for Independent Experts

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 nations 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 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

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

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.

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.

Annex 2: Terms of Reference for the Peer Review

SEDAR 63 Gulf of Mexico Menhaden Review Workshop Terms of Reference

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?

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?

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?

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.

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.

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

7) Prepare a Peer Review Summary summarizing the Panels 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.

The panel shall ensure that corrected estimates are provided by addenda to the assessment report in the event corrections are made in the assessment, alternative model configurations are recommended, or additional analyses are prepared as a result of review panel findings regarding the TORs above.

Annex 3: Tentative Agenda - SEDAR 63 Gulf of Mexico Menhaden Review

New Orleans, LA

November 6-7, 2018

Tuesday		
8:30 a.m.	Convene	
8:30 a.m. 9:00 a.m	Introductions and Opening Remarks	
Coordinator/Chair		
	- Agenda Review, TOR, Task Assignments	
9:00 a.m. 12:00 p.m.	Assessment Presentations	TBD
12:00 p.m. 1:00 p.m.	Lunch Break	
1:00 p.m. 3:30 p.m.	Continue Presentations / Panel Discussion	Chair
	- Assessment Data & Methods	
	- Identify additional analyses, sensitivities, corrections	
3:30 p.m. 4:00 p.m.	Break	
4:00 p.m. 5:30 p.m.	Continue Discussion	Chair
5:30 p.m. 6:00 p.m.	Public Comment	

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

Wednesday		
8:30 a.m. 12:00 p.m.	Panel Discussion	Chair
	- Discuss initial findings	
	- Continue deliberations	

	- Review additional analyses	
12:00 a.m. 1:00 p.m.	Lunch Break	
1:00 p.m. 3:00 p.m.	Panel Discussion	Chair
	-Recommendations and comments	
	-Final sensitivities and projections reviewed	
	-Review draft report sections	
3:00 p.m. 3:30 p.m.	Public Comment	
3:30 p.m.	ADJOURN	

Wednesday Goals: Additional analyses reviewed, preferred models selected and projection approaches approved, assessment work and discussions completed, draft reports reviewed

APPENDIX 3
PERTINENT INFORMATION FROM THE REVIEW

1) Participants List*

Review Workshop Panelists		
Will Patterson	Review Panel Chair	GSMFC Appointee
Joe Powers	Reviewer	GSMFC Appointee
Matt Cieri	Reviewer	CIE
Anders Nielsen	Reviewer	CIE
Kevin Stokes	Reviewer	CIE
Analytical Team		
Amy Scheuller	Lead analyst	SEFSC Beaufort
Robert Leaf	Assessment Team	GCRL
Ray Mroch	Assessment Team	SEFSC Beaufort
Council and Commission Staff		
Julia Byrd	Coordinator	SEDAR
Steve VanderKooy	IJF /Aquaculture Coordinator	GSMFC
Kimberley Cole	Admin	SEDAR/SAFMC

*Industry observers/advisors also participated but at time of report submission to CIE, the list has not been advised.