SEDAR

Southeast Data, Assessment, and Review

SEDAR 32

South Atlantic Blueline Tilefish

SECTION V: Review Workshop Report

September 2013

SEDAR
4055 Faber Place Drive, Suite 201
North Charleston, SC 29405

Table of Contents

1. Introduction 3
1.1 Workshop Time and Place 3
1.2 Terms of Reference 3
1.3 List of Participants 4
1.4 List of Data Workshop Working Papers 5
2. Review Panel Report 8
Executive Summary 8
2.1 Statements Addressing Each ToR 8
2.2 Summary Results of Analytical Requests 22

1. Introduction

1.1 Workshop Time and Place

The SEDAR 32 Review Workshop for South Atlantic blueline tilefish (*Caulolatilus microps*) was held August 27-30 in Morehead City, NC. It was held in conjunction with the Review Workshop for SEDAR 32A for Gulf of Mexico menhaden (*Brevortia patronus*).

1.2 Terms of Reference

1. Evaluate the data used in the assessment, addressing the following:
   a) Are data decisions made by the DW and AW 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 reliable? If not, are there other indicators that may be used to inform managers about stock trends and conditions?
4. Evaluate the stock projections, addressing the following:
   a) Are the methods consistent with accepted practices and available data?
   b) Are the methods appropriate for the assessment model and outputs?
   c) Are the results informative and robust, and useful to support inferences of probable future conditions?
   d) Are key uncertainties acknowledged, discussed, and reflected in the projection results?

5. Consider how uncertainties in the assessment, and their potential consequences, are addressed.
   - 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.
   - Ensure that the implications of uncertainty in technical conclusions are clearly stated.
6. Consider the research recommendations provided by the Data and Assessment workshops and make any additional recommendations or prioritizations warranted.
   - Clearly denote research and monitoring that could improve the reliability of, and information provided by, future assessments.
   - Provide recommendations on possible ways to improve the SEDAR process.
7. Provide guidance on key improvements in data or modeling approaches which should be considered when scheduling the next assessment.
8. 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.

1.3 List of Participants

Review Workshop Panelists

Steve Cadrin	Review Panel Chair	SAFMC SSC
Churchill Grimes	Reviewer	SAFMC SSC
Will Patterson	Reviewer	GSMFC Appointee
Gary Melvin	Reviewer	CIE
Stephen Smith	Reviewer	CIE
Kevin Stokes	Reviewer	CIE

Analytical Team

Kevin Craig	Lead analyst, SA BLT	NMFS Beaufort
Amy Scheuller	Lead analyst, GoM menhaden	NMFS Beaufort
Kyle Shertzer	Assessment Team	NMFS Beaufort
Erik Williams	Assessment Team	NMFS Beaufort
Katie Andrew	Assessment Team	NMFS Beaufort
Rob Cheshire	Assessment Team	NMFS Beaufort
Robert Leaf	Assessment Team	USM

Observers

Dewey Hemilright	Fishing Industry	Commercial, NC
Robert Johnson	Fishing Industry	Charter/Headboat, FL

GSMFC Menhaden Advisory Committee

John Mareska, ADCNR-MRD
Behzad Mahmoudi, FL FWC
Jerry Mambretti, TPWD
Borden Wallace, Daybrook Fisheries
Ron Lukens, Omega Protein, Inc.
Matt Hill, MDMR
Harry Blanchet, LDWF

Council Representative

Michelle Duval	Council Member	SAFMC

Council and Agency Staff

Julia Byrd	SEDAR Coordinator	SEDAR
Julie O'Dell	Administration	SEDAR/SAFMC
Michael Errigo	Fishery Biologist	SAFMC Staff
Steve VanderKooy	IJF Program Coordinator	GSMFC
Jessica Stephen	Fishery Biologist	SERO
Brian Langseth	Observer	SEFSC Beaufort
Joe Smith	Observer	NOAA

Data workshop observers

Tony Austin
Doug Vaughan
Mike Prager
Robert O'Boyle

1.4 List of Data Workshop Working Papers

South Atlantic blueline tilefish and gray triggerfish reference workshop document list.

Document #	Title	Authors
Documents Prepared for the Review Workshop		
SEDAR32-RW01	The Beaufort Assessment Model (BAM) with application to blueline tilefish: mathematical description, implementation details, and computer code	NMFS-SFB 2013
SEDAR32-RW02	Catch Curves for blueline tilefish from the commercial handline and longline fleets	NMFS-SFB 2013
Reference Documents		
SEDAR32-RD01	List of documents and working papers for SEDAR 4 (Caribbean – Atlantic Deepwater Snapper Grouper) – all documents available on the SEDAR website.	SEDAR 4
SEDAR32-RD02	Comparison of Reef Fish Catch per Unit Effort	Rudershausen et al.

	and Total Mortality between the 1970s and 2005–2006 in Onslow Bay, North Carolina	2008
SEDAR32-RD03	Source document for the snapper-grouper fishery of the South Atlantic region.	SAFMC 1983
SEDAR32-RD04	FMP, regulatory impact review, and final environmental impact statement for the SG fishery of the South Atlantic region	SAFMC 1983
SEDAR32-RD05	Age, growth and reproductive biology of blueline tilefish along the southeastern coast of the United States, 1982-99	Harris et al. 2004
SEDAR32-RD06	List of documents and working papers for SEDAR 9 (Gulf of Mexico Gray Triggerfish, Greater Amberjack, and Vermillion Snapper)	SEDAR 9
SEDAR32-RD07	Estimated Conversion Factors for Adjusting MRFSS Gulf of Mexico Red Snapper Catch Estimates and Variances in 1981-2003 to MRIP Estimates and Variances	Rios et al. 2012
SEDAR32-RD08	Estimates of Historic Recreational Landings of Spanish Mackerel in the South Atlantic Using the FHWAR Census Method	Brennan and Fitzpatrick 2012
SEDAR32-RD09	Excerpt from ASMFC Atlantic Croaker Stock Assessment & Peer Review Reports 2003 – Information on Jacquard Index	ASMFC 2003
SEDAR32-RD10	Survival estimates for demersal reef fishes released by anglers	Collins 1994
SEDAR32-RD11	Indirect estimation of red snapper (Lutjanus campechanus) and gray triggerfish (Balistes capriscus) release mortality	Patterson et al. 2002
SEDAR32-RD12	Estimating discard mortality of black sea bass (Centropristis striata) and other reef fish in North Carolina using a tag-return approach	Rudershausen et al. 2010
SEDAR32-RD13	Commercial catch composition with discard and immediate release mortality proportions off the southeastern coast of the United States	Stephen and Harris 2010
SEDAR32-RD14	Migration and Standing Stock of Fishes Associated with Artificial and Natural Reefs on Georgia's Outer Continental Shelf	Ansley & Harris 1981
SEDAR32-RD15	Age, Growth, and Reproductive Biology of the Gray Triggerfish (*Balistes capriscus*) from the Southeastern United States, 1992-1997	Moore 2001

SEDAR32-RD16	Size, growth, temperature, and the natural mortality of marine fish	Gislason et al. 2010
SEDAR32-RD17	Evolutionary assembly rules for fish life histories	Charnov et al. 2012
SEDAR32-RD18	A Review for Estimating Natural Mortality in Fish Populations	Siegfried & Sansó

2. Review Panel Report

Executive Summary

The stock assessment presented by the SEDAR 32 Assessment Workshop (AW) provided the Review Panel with outputs and results from two statistical assessment models and a catch curve analysis. The primary model was the Beaufort Assessment Model (BAM), while a secondary, surplus-production model (ASPIC), provided a comparison of model results. The Review Panel endorses the AW recommendation to determine stock status using the BAM base configuration. Fishing mortality in 2011 is estimated as 0.39, which is greater than the estimate of Fmsy (0.302), so overfishing is estimated to be occurring. Spawning biomass in 2011 is estimated as 445,000 lb, which is 91% of the estimate of Minimum Stock Size Threshold (489 ,000 lb), so the stock is estimated to be overfished.

2.1. Response to Terms of Reference

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

- Stock area

The management area was defined such that landings from Rhode Island to Florida were used for this stock assessment. There are no genetics or tagging data available for this species to define biological stocks or the management area, but many species exhibit a stock boundary along the US east coast at Cape Hatteras. Blueline tilefish are pelagic spawners and as a consequence, it was suggested that larvae would be wide-ranging. However, previous work on the confamilial golden tilefish indicate a stock break north and south of Cape Hatteras (Katz, et al.1983). There was concern expressed that the stock area may be too broad given that the fishery appears to be focused in a few small areas, and because this species is known to be highly residential, occupying scour depressions in carbonate substratum and burrows in soft bottom (Able, et al.1987). Such an aggregated species may be subject to local depletion.

Research Recommendation: Further research on stock structure would help align landings and the indices being used to monitor annual changes in stock size.

- Natural mortality

Natural mortality at age was estimated using the methods of Charnov et al. (2012) which are based on estimates of K and L_∞ from von Bertalanffy growth curves, and therefore highly dependent upon the quality of the age data. Considerable uncertainty in age determination for blueline tilefish was documented by Harris et al. (2004).

Scaling the mean rate over the older ages to 0.1 was reasonable given the Hoenig estimate based on maximum age. Values of 0.15 and 0.05 were used for sensitivity training based upon a CV of

54% from the Hoenig method. However, the lack of fish of age 15 years and older in the landings suggests that either M may be higher because the maximum age of 43 is questionable due to the uncertainty in ageing or Fishing mortality was much higher than assumed. This suggests that the higher M alternative should receive more attention in the sensitivity analysis than the lower M and perhaps M estimates higher than 0.15 might also be considered.

- Maturity at age

Maturity-at-age was based upon estimates for golden tilefish with 50% mature at age 3 and 100% mature at age 4. While these results indicated a relatively younger maturity than might be expected for such a long-lived fish, similar results have been reported for other long-lived species in the region. However, maturity studies of golden tilefish, a confamilial species, suggest that functional maturity may occur at ages older than histological maturity because of territoriality, dominance and mate choice (Grimes et al. 1988 and McBride et al. 2013). If this is also true for blueline tilefish, then the apparent truncation of age composition due to harvesting may result in a decline in the size of males that gain access to the females for spawning.

- Ageing

The von Bertalanffy growth curve indicated that 98% of total growth is completed by age 15 yrs., and therefore ages 15 yrs. and older were adopted as a plus group. The underlying growth data were obtained from sampling recent landings for fisheries that appeared to target a very narrow range of ages (3-5 yrs. for recreational and 5-8 yrs. for commercial fisheries). There were no age composition data for landings in the earlier part of the series when it was expected that larger/older fish should have been a higher proportion of the population given the assumption of maximum age of 43 yrs. As noted above, the reliability of the underlying assumptions of the initial age composition raises issues about the current estimates of M and F, as well the assumption of flat-topped selectivity. Industry comments during the meeting suggested that there may be differing spatial distribution by size/age class. The available age composition data do not appear to track year-classes, even though high recruitment was estimated to have occurred prior to the period that the bulk of these data were collected.

Age and growth information used in the assessment was extracted from Harris et al. 2004. This study did not rigorously validate putative ages and reported low aging precision, e.g., ~ 60% within 2 yrs.

An ageing error matrix was developed at NMFS Beaufort comparing the results of two agers. Due to the small sample sizes, ageing errors were assumed to follow normal distributions. A symmetric distribution of errors was questioned as experience suggests that older ages tend to be more likely to be underestimated as annuli tend to pack at the otolith margin as the fish approach the asymptotic length. However, uncertainty in age determination as measured by the ageing error matrix was considered to be relatively small in comparison to other sources of uncertainty that had been identified.

While the age compositions were fitted by the model, the length compositions were removed from the analysis due to preliminary results indicating lack of fit. In light of the uncertainties associated with the ageing data, it seemed strange that the length composition data would not be better fitted by the model.

- Quality of commercial and recreational landings data

The landing data were considered to be reliable since 1974 and discarding for the commercial fishery was assumed to be negligible and consistent with there being no regulatory reasons for discarding (e,g., size limits). The recreational catch was sporadic and low relative to the commercial catch until 2006. There was considerable discussion about the reliability of the recreational landings estimate for 2006 to 2008 including the very high discard estimates in 2007. Most of these landings appeared to have occurred in North Carolina waters and there was a suggestion that the development of a "deep-drop" fishery may have driven the increase with the decrease in 2011 due to the implementation of a deep water closure. Examination of the MRIP data indicated that CVs for 2006 to 2011 decreased relative to the period before and the number of sample intercepts increased, both indicative of increased fishing activity. However, magnitude of the landings relative to the commercial landings in those same years still seemed to be unprecedented and industry participants questioned the reliability of the recreational estimates.

- Abundance indices:

The commercial and recreational catch rate information was key data for both the BAM and ASPIC models. These were the only annual abundance indices available and were developed using standard approaches, i.e., fit delta-GLM models to filter out annual trends from other factors associated with these data. The recreational index represents the earlier period when the SSB was being fished down but this index actually represents very low levels of catch. There was no overlap between this index and the two commercial indices.

- Landings, catch at age and CPUE

Landings and catch-at-age were estimated for the entire geographic domain of the fishery, including those that came from north of 35N. However, CPUE was only computed for areas north of 28 N and south of 35N. When we examined nominal CPUE by latitude, regardless of fishery it was higher north of 35N than the standardized composite CPUE used as an abundance index in the assessment. Therefore, increased landings north of 35N are not being fully indexed. One implication of this is the BAM model fits this increase in landings as an increase in recruitment, thus the greatest positive recruitment deviations in the model (see assessment document Fig. 3.13). This clearly has implications for projected future stock productivity.

2. Evaluate the methods used to assess the stock, taking into account 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. Use of the BAM permitted the inclusion of all available types of data, including total annual removals from commercial and recreational fleets (landings and discards), age and length compositions, and indices of biomass abundance, with appropriate error distributions and use of priors on parameters. Decisions on a priori data inclusion and exclusion are considered at ToR 1.

The specified assessment model used standard approaches to predicting landings and modeling growth and recruitment BAM also allowed an exploration of catchability and selectivity options.

The base case model and rationale for modeling decisions are well described in the AW report (section 3) and were further explored during the Review Workshop. The base case run included commercial and recreational landings, age composition data and three indices of abundance (recreational head boats, commercial long line and hand line). There was some concern that the recreational and commercial indices do not overlap, but this was explored during the RW and the general patterns seem to be consistent. Length compositions were excluded by the AW due to concerns about inconsistent sampling and conflicts in fitting. The AW concluded that length composition data help to inform selectivity estimates but conflict with information in abundance indices, do not track year classes well, and add unnecessary noise. The RW panel was concerned at this exclusion and the issue was explored further during the RW by looking at shadow fits comparing the base case predicted (but not fit) length compositions with the data and by examining models fits to the length composition data. The RW concluded that the residual patterns in indices were not acceptable from the model that included length compositions, and the results could not be considered as a viable base case (or sensitivity run). The decision by the AW to exclude length composition data was therefore upheld. Natural mortality was assumed constant through time but age-specific based on the method of Charnov (2013) and scaled consistent with maximum observed age. Steepness was fixed at 0.84 based on meta-analyses (Myers et al., 2002; Shertzer and Conn, 2012). Selectivities and catchabilities were all estimated as constant for the full assessment period (1974-2011).

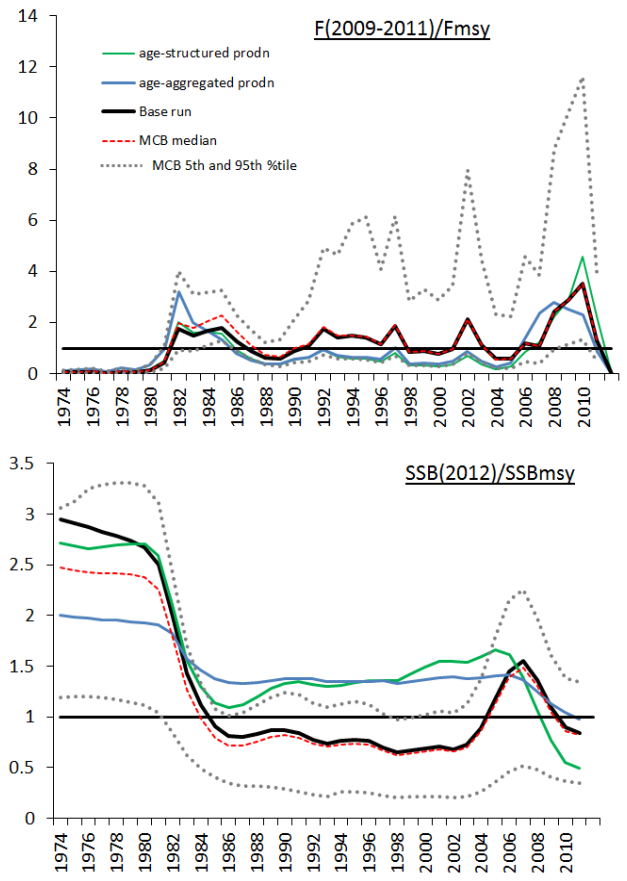
The model was fit to the data using appropriate methods, consistent with standard practice. Analysis included iterative reweighting using the method of Francis (2011) and exploration of a variety of data configurations and parameterizations. The modeling processes and decision making resulting in a proposed base case run and sensitivity testing are well described in the AW Report and AW WDs and were further elaborated during the SEDAR 32 Review Workshop where additional diagnostics (Likelihood components, weights, likelihood profiles) were made available. The modeling procedures adopted appear to be robust. Landings and discards were fit closely, and age composition data and abundance indices were fit to the degree that they are compatible and as indicated using the reweighting procedures. Landings and indices were fit

using lognormal likelihoods. Age composition data were fit using robust multinomial likelihoods.

The treatment of the data and the relative importance given to the various components were well explored by the AW and at the RW and appear appropriate. The model structure is adequate to capture the main patterns in the data.

In addition to the catch-at-age primary assessment, two biomass dynamics stock assessments were carried out using the ASPIC software, one fully age-aggregated and the other age structured. The biomass dynamics models were considered as complementary rather than alternative analyses, because the catch-at-age model makes fuller use of composition data and represents a more detailed investigation of population dynamics. The biomass dynamics models provide a useful comparison with the catch-at-age model results (see Figure below), which they broadly support, showing the similar status of the stock in relation to MSY benchmarks (ToR 3). The biomass dynamics models and methods used are well known and were appropriately configured and implemented.

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 the age composition data. The Monte Carlo component entails drawing values of M and steepness from specified pdf's. 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 vase case run. No reweighting procedures are used for individual realizations.



Trajectories of status benchmarks for the catch-at-age base case model, two biomass dynamics model runs, and the MCB analysis. Refer to key for explanation.

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 (MSY, BMSY, FMSY) from the MCB runs do not equal estimates from the base run. 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. Of course, these differences will not be apparent when looking only at ratio benchmarks as in the figure above. It is important to note that for consistency, if MCB is used for projections, the MCB estimates of the management quantities should also be used for evaluating stock status.

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?

      The review panel examined the consistency of input data and population biological characteristics with abundance, exploitation and biomass estimates. The Review Panel agreed with the AW that the base run provided the best representation of stock status, and the MCB should be used for projection estimates. The base rune outputs are generally consistent the inputs, given assumptions and weighting choices.

   b) Is the stock overfished? What information helps you reach this conclusion?

      The RW Panel endorses the AW recommendation to determine stock status using the BAM base configuration. Based on the base run estimates of SSB, the South Atlantic Blueline tilefish is overfished. Spawning biomass in 2011 is estimated as 445 thousand lb, which is 91 per cent of the estimate of Minimum Stock Size Threshold (489 thousand lb), so the stock is overfished. SSB has been below the MSST for the past two years (2010-2011). The majority of viable sensitivity runs indicate that the SSB2011 was < SSBmsy. The only exception is if M is higher, in which case SSB may be estimated greater than SSBmsy. The RP initially had some concerns about the assumed M value and suggested that a higher value might be credible. However, likelihood profiles presented during the meeting supported the use of the assumed value (0.1).Production model outputs of population status generally agree with the BAM base run and indicate a B/Bmsy of less than 1 in 2011.

   c) Is the stock undergoing overfishing? What information helps you reach this conclusion?

      Based on the BAM base run fishing mortality (F) estimates, overfishing is occurring for the South Atlantic blueline tilefish. The ratio of the geometric mean F over the past 3 years to Fmsy was greater (2.37) than 1.0 and has been for the past several years. The decrease in $F_{(2011)}$ was primarily the result of a fishery closure, which no longer exists. Production model outputs all indicate an average F/Fmsy well in excess of 1.0.

4. Evaluate the stock projections, addressing the following:

   a) Are the methods consistent with accepted practices and available data?

      The methods used by the AW are consistent with accepted practices in the region and elsewhere, and the available data. Initially the review panel had several concerns regarding the use of Monte Carlo and bootstrap (MCB) approach as a measure of precision and to compute uncertainty. The MCB analysis is considered an

approximation of uncertainty for an individual run. Unconverged and unrealistic runs were removed (3200 reduced to 3043) from the analysis, however, there was still the possibility of including nonsense variable inputs that individually could occur within the established parameter bounds, but combined (biologically)could not, resulting in unrealistic outputs of R0 and Fmsy. All unfiltered runs were given equal weight and were included in the estimate of uncertainty. These limitations were identified in the assessment report. In addition, there was the mixing of deterministic and stochastic parameters, the latter introducing process error. The Review Panel concluded that although the MCB approach is a common approach used in SEDAR assessments to estimate uncertainty, the results may be different if a true Bayesian approach was applied.

The panel questioned if the assumed F in 2012 and 2013 was overestimated because of changes in regulations and closures. However, examination of the preliminary 2012 landings showed a substantial increase from 2011, thereby justifying the assumed F. The Panel recommends that projections of future catch should be based on direct estimates of past catch when available rather than assumed F.

b) Are the methods appropriate for the assessment model and outputs?

Five-year projections were made using the MCB model to capture uncertainty in data and parameter inputs. The assumed error structures on data are as used for fitting the BAM base run. The pdf on M is effectively uniform from 0.05 to 0.15, consistent with the sensitivity tests using the BAM and covering the central assumption. The pdf for h has a mean of 0.84, consistent with the BAM base run and is based on a published meta-analysis (Shertzer and Conn, 2012). Numbers in 2012 are based on 2011 estimates for ages 2 to 15+, discounted by estimated Z. Initial recruits are computed from the spawning-recruit model with h drawn from the pdf at each realization. Consistent with the F used to determine status, F2012 is calculated as F2009-2011. A total of 10,000 projected time series were made in the MCB and four alternative F scenarios were investigated (F0, Frebuild, Fmsy, and Fcurrent).

The method used for projections are appropriate but the RP noted that because the estimates of Fmsy, Bmsy and Msy are different between the MCB and BAM (due to inclusion, and dependent on the degree, of process error in the BAM) then it would make sense also to use the MCB to determine stock status. This needs further consideration generally.

c) Are the results informative and robust, and useful to support inferences of probable future conditions?

Projection results are informative and robust within the range of observations and inputs from the MCB. Currently F is estimated as the mean of the 3 previous years,

one of which (2011) was subject to lower F due to a closure. Given the observed rapid changes in F and the preliminary landings estimates for 2012 and 2013, consideration might be given to using actual landings for future projections or drop the 2011 from estimate of F for 2013 and 14.

d) Are key uncertainties acknowledged, discussed, and reflected in the projection results?

The assessment report identified and evaluated uncertainties associated with the assessment through the MCB. The report identifies the degree of uncertainty associated with M, ageing error, steepness, model component weights, indices and recruitment deviations. Some concern was expressed by the Review Panel on the appropriateness of using the mean F (high relative to the time series) for the previous 3 years given the high F's of 2009 and 2010 and the low value for 2011 for projections. However, examination of the preliminary landings for 2012 and 2013 support the use of a large F.

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

Uncertainty was explored in the assessment modeling using extensive sensitivity runs and likelihood profiling, retrospective analyses and Monte Carlo Bootstrapping (MCB). All of 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, for example, processes such as fishery or environmentally induced geographic changes in distribution of the stock or fishery induced local depletion. There was much discussion at the RW on these issues and on data inclusion or exclusion in indices to represent stock abundance. Ultimately, the stock assessment assumes a single dynamic pool of fish and there are insufficient data at this time to support investigating alternative hypotheses. With the exception of this structural uncertainty, the other uncertainties in the assessment and its outputs have been appropriately and comprehensively considered.

Issues considered in sensitivity runs include variations in M and steepness, alternative maturity vector, adjustment of model weights and exclusion of each series of indices, allowing catchability to vary, inclusion of ageing error, and allowing recreational selectivity to be dome shaped. Issues of uncertainty not covered explicitly in sensitivity tests include the quantum of landings assigned to recreational landings and especially discards in 2007-9 (see ToR 1).

The MCB is alluded to at ToR 2. A total of 3200 realizations were made using M and h values drawn from specified pdf's and with the landings, indices and age composition data bootstrapped. Each realization 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 3200 realizations). However, it should be noted that reweighting can have major implications for fitting and parameter estimation and that each realization may not be feasible. The degree to which this may or may not matter is model and data specific. As all realizations are afforded equal weight in determining distributions of outputs there is in general need for care in interpreting MCB results. For blueline tilefish, the SDNRs for all sensitivity tests are surprisingly good when runs are made using the base case weights. This is encouraging, however, is no guarantee that for specific M and h combinations drawn from the pdfs, which may be incompatible, the base case weights would in any way be appropriate.

Notwithstanding, the RW was comfortable that the AW had fully explored uncertainty to the extent possible and that the characterization of benchmark trajectories (Figure above) and hence stock status (ToR 3) and projections (ToR 4) are suitable for informing management decisions.

6. Consider the research recommendations provided by the Data and Assessment workshops and make any additional recommendations or prioritizations warranted.

Research recommendations for blueline tilefish were provided in the data and assessment working group documents (see below). The Panel noted that many of these recommendations reflected concerns across a range of deep-water species and therefore confined their attention to those specific to the stock assessment of blueline tilefish.

While the panel supports work on stock structure, we recommend starting with the available information on describing the differences in demographics/life history characteristics over the range of the management area. Additionally, the available information on habitat in the areas listed should be evaluated before initiating any new studies.

Given that this is an age-based assessment, the comparison and calibration studies for the age determination should receive high priority along with the marginal increment analysis to determine if the opaque zone is formed annually. Many species would probably benefit from expanding the MRIP program to include age sampling.

The collection of information to better describe spawning season and spawning periodicity could probably start with fishery-dependent sources, but will need data from fishery-independent programs to cover the range of the species. The latter program would probably have to be tailored to provide samples across the deep-water snapper/grouper complex.

Studies of discard mortality should be low priority given the current negligible discard rate in the commercial fishery. The collection of additional information on discards and catch (e.g, lengths,

ageing material) is important especially for the areas north of Hatteras, but would likely require an observer program developed for all fisheries focusing on the deep-water snapper/grouper complex.

The BAM model is reliant on historical information and any data on size compositions, maximum size, etc., that can be obtained from historical recreational fishing photos could be quite useful. One of the main issues raised about the recreational fishery concerned the high landings in the mid-late 2000s, especially the high landing and discard estimates for 2007. Closer scrutiny of these estimates requires data at higher resolution than was apparently available for this stock assessment.

With respect to developing a fishery-independent survey, sampling of deep-water habitats may elucidate habitat characteristics, and spatial distributions of blueline tilefish and other deep-water reef fishes. If a sufficient time series is developed, then a fishery-independent index may be developed.

Research Recommendations from the Data and Assessment Working Groups

7.1 Life History

- Stock Structure
    - Blueline tilefish stock definition needs to be investigated further. Genetic study or some other form of stock identification study needs to be undertaken with samples (muscle, fin clips, etc.) collected from several locations within the Gulf of Mexico and the northwestern Atlantic.
    - Habitat studies of deep water sites in the mid-Atlantic, specifically Norfolk Canyon, Baltimore Canyon, and Hudson Canyon need to be undertaken. Temperature data from research conducted in the 1970s in Norfolk Canyon can be used for comparison purposes.
- Age Data
    - Age readings of blueline tilefish need to be validated. Within and between lab variability in readings is large and needs to be addressed. The potential bias in age readings between laboratories also needs to be addressed with another age workshop and exchange of calibration sets of samples.
    - Marginal increment analysis needs to be undertaken in order to convert increment counts to calendar ages. Samples processed and read in older studies will need to be re-examined and margin codes recorded for each.
    - More recreational fishery age samples need to be collected.
- Reproductive Biology Data
    - Overall, more reproductive samples need to be collected. Because small, young fish were lacking from the biological collections, specimens under 18 inches will be needed to address age and size at maturity. Whole gonads will need to be

collected for a fecundity study. Specimens collected from throughout the species range and covering all months of the year are needed to better describe spawning season and spawning periodicity.

- Ad-hoc Discard Mortality Sub-group
    - Future research is needed to examine discard mortality rates for this species, as well as factors that affect survival (e.g., gear type, temperature, depth).

7.2 Commercial Fishery Statistics

- Discard
    - Investigate the validity and magnitude of "no discard" trips. This may include fisher interviews throughout the region.
    - Examine potential impacts on "no discard" trips, including:
        - Trip length
        - Trip dates in relation to fishery regulations
        - Trip targeting
        - Trip area fished
    - Improve discard logbook data collections via program expansion or more detailed reporting (e.g. more detailed logbook, electronic reporting)
    - Develop an observer program that is representative of the fishery in the South Atlantic.
- Biosampling
    - Standardize TIP sampling protocol to get representative samples at the species level.
    - Develop an observer program that is representative of the fishery in the South Atlantic.
    - Increase untargeted sampling in NE and Mid-Atlantic observer programs.
    - Increase untargeted dockside sampling in NE and Mid-Atlantic.

7.3 Recreational Fishery Statistics

- Continued research efforts to incorporate/require logbook reporting from recreational anglers.
- Quantify historical fishing photos for use in future SEDARs.
- Fund research efforts to collect discard length and age data from the private sector.
- Improve metadata collection in the recreational fishery.
- Pre-stratify MRIP Keys, N-S Canaveral, N – S Hatteras.
- Research possibility of implementing private recreational reef fish stamp to determine universe and reporting strategies.

- At-sea observers collect surface and bottom temperature.
- At-sea observer protocols should include all fields currently used in FL i.e., condition and depth of released fish.

7.4 Indices

- Evaluate various sub-setting methods to identify effective effort. Methods that have been applied or considered include in this and previous SEDAR assessments include the Jaccard statistic, Stephens and MacCall approach, variations of Stephens and MacCall approach (e.g., using amount of catch rather than presence-absence), and other multivariate statistical approaches (e.g., cluster analysis).
- Evaluate various standardization methods to handle zeros in the catch, e.g., delta-GLM, zero-inflated Poisson, zero-inflated negative binomial, hurdle models, etc.
- Evaluate possible effects of circle hooks on catchability of reef fishes.
- Need fishery independent sampling of deep-water species, including blueline tilefish. Need funding to support these efforts.

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

The Atlantic blueline tilefish assessment relies upon fishery dependent indexes of abundance to inform theBAM. No fishery independent indices are available for this stock. As such the geographic distribution, seasonal movement, spawning, and consistency of the fishery over time all have an impact on the indices and contribute to the uncertainty associated with the assessment. Whether or not the stock is truly a single spawning population distributed throughout the stock range or a series of multiple spawning components is unknown given its broad spatial occurrence along the Atlantic coast. Changes in the state proportional contribution to total landings and catches from the commercial handline and longline fisheries implies a divergence from a more southerly dominated (Florida and South Carolina) fishery during the 1980's to a northern (North Carolina, especially above Cape Hatteras) focused fishery in more recent years. The reason(s) for these observed changes in landings are unknown. The changes in catch and subsequent catch rates used as indices of abundance may be a function of population dynamics, serial depletion, or a northerly migration in response to environmental variability. Further investigation of this issue should be undertaken before the next assessment to insure the current commercial indices represent changes in abundance and not the adaption of the fishing fleets to availability. Development of a fishery independent index of abundance would help to resolve some of these issues.

While the size of this fishery may not by itself warrant the cost of implementing such a survey, there may be broader advantages in designing a survey for the complex of deep-water species.

During the initial review and presentation of the stock assessment it was unclear that the commercial CPUE indices were being truncated or trimmed at Cape Hatteras, thereby excluding the catch and effort data north of this area. The landings data used in the assessment model included all reported catches taken throughout the entire range of the stock. Given a large portion of recent landings are being reported north of Cape Hatteras are not included in the commercial CPUE indices the effects on the abundance indices are unknown. The review panel suggests the increased catches be addressed and that this apparent inconsistency between the indices and the fishery be resolved before the next assessment.

The blueline tilefish assessment uses 3 CPUE indices based on information from the headboat (1980-1992), handline (1993-2010) and longline (1993-2004), with no data for 2011 due to a commercial and recreational closure. The headboat time series was terminated due to the low number of trips/catches. No overlapping years between the headboat and the other two indices were used in the assessment suggesting uncertainty in the scaling of the indices. Limited information was available for the headboat over the entire time series. During the review the panel requested additional analysis on the headboat time series to investigate if there were consistencies in CPUE patterns. When the headboat data were binned into 3 year averages the data generally tracked the ups and downs of the other indices. The headboat data should be investigated further to see if the times series can be extended, especially given the recent increases in headboat catches since 2008.

References

Able, K.W., D.C. Twichell, C.B. Grimes and R.S. Jones. 1987. Tilefishes of the genus Caulolatilus construct burrows in the sea floor. Bull. Mar. Sci. 40(1):1-10.

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 254:65–75,

Charnov, E.L., H. Gislason and J.G. Pope. 2012. Evolutionary assembly rules for fish life histories. Fish and Fisheries.

Fournier, D. A., Skaug, H. J., Ancheta, J, Ianelli, J., Magnusson, A., Maunder, M.N., Nielsen, A., and Sibert, J. 2012. AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. Optimization Methods and Software 27:233–249.

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

Grimes, C. B., C. F. Idelberger, K. W. Able, and S. C. Turner. 1988. The reproductive biology of

tilefish, *Lopholatilus chamaeleonticeps* Goode and Bean, from the United States Mid-Atlantic Bight, and the effects of fishing on the breeding system. Fish. Bull. 86:745–762.

Harris, P.J., D.M. Wyanski, and P.T. Powers Mikell. 2004. Age, growth and reproduction of blueline tilefish along the southeastern coast of the United States, 1982-1999. Transactions of the American Fisheries Society 133:1190-1204.

Katz, S.J., C.B. Grimes and K.W. Able. 1983. Delineation of tilefish, *Lopholatilus chamaeleonticeps*, stocks along the U.S. east coast and in the Gulf of Mexico. Fish. Bull. 81(1):41-50.

McBride, R.S., T.E. Vidal and S.X. Cadrin. 2013. Changes in size and age at maturity of the northern stock of Tilefish (*Lopholatilus chamaeleonticeps*) after a period of overfishing. Fish. Bull. 111:161–174.

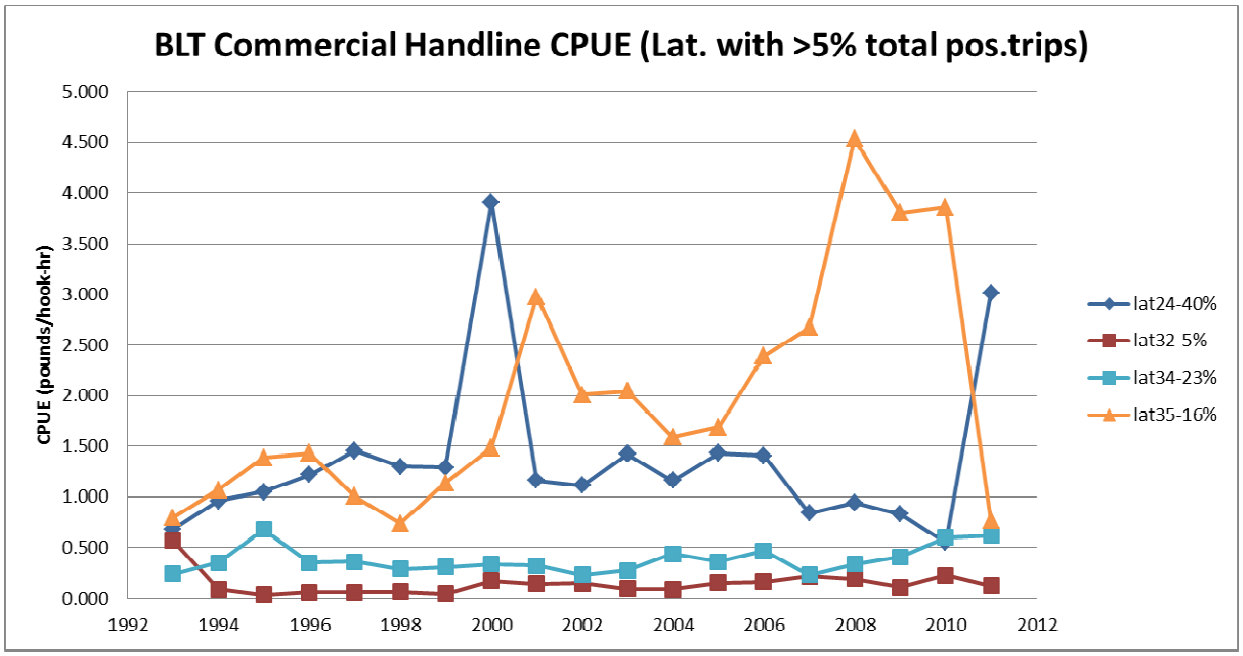
Myers, R., Barrowman, N., Hilborn, R., and Kehler, D. 2002. Inferring Bayesian priors with limited direct data: applications to risk analysis. North American Journal of Fisheries Management. 22: 351–364.

Shertzer, K.W., and Conn, P.B. 2012. Spawner–recruit relationships of demersal marine fishes: prior distribution of steepness. Bulletin of Marine Science. 88: 39–50.

2.2 Summary Results of Analytical Requests

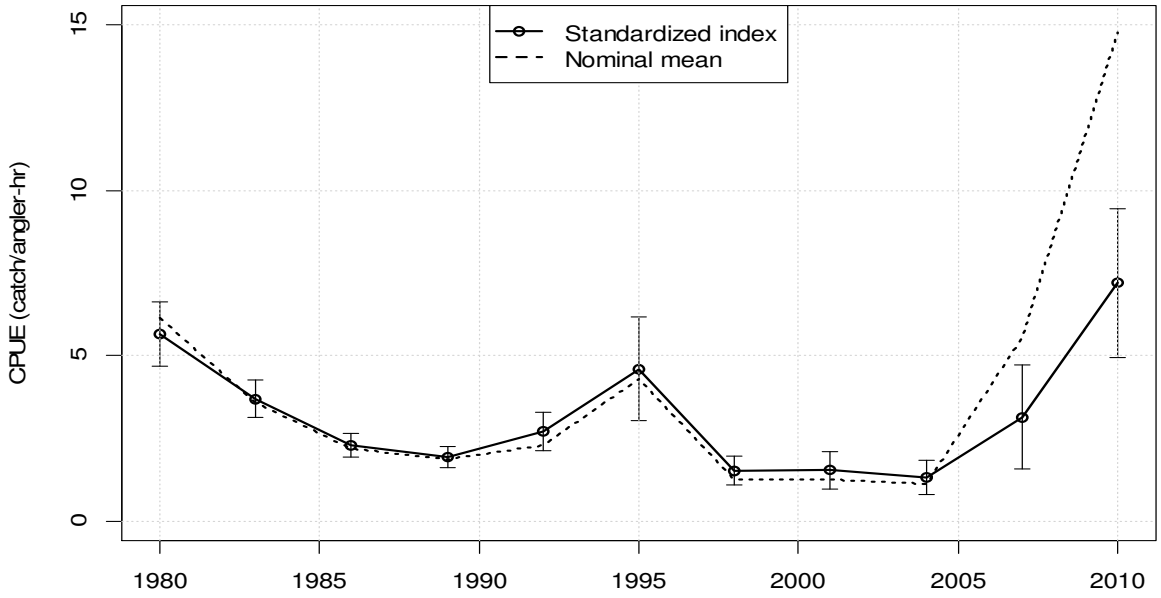
- The review panel requested geographic plots of the fishery to evaluate the extent of the spatial distribution of the fishery.

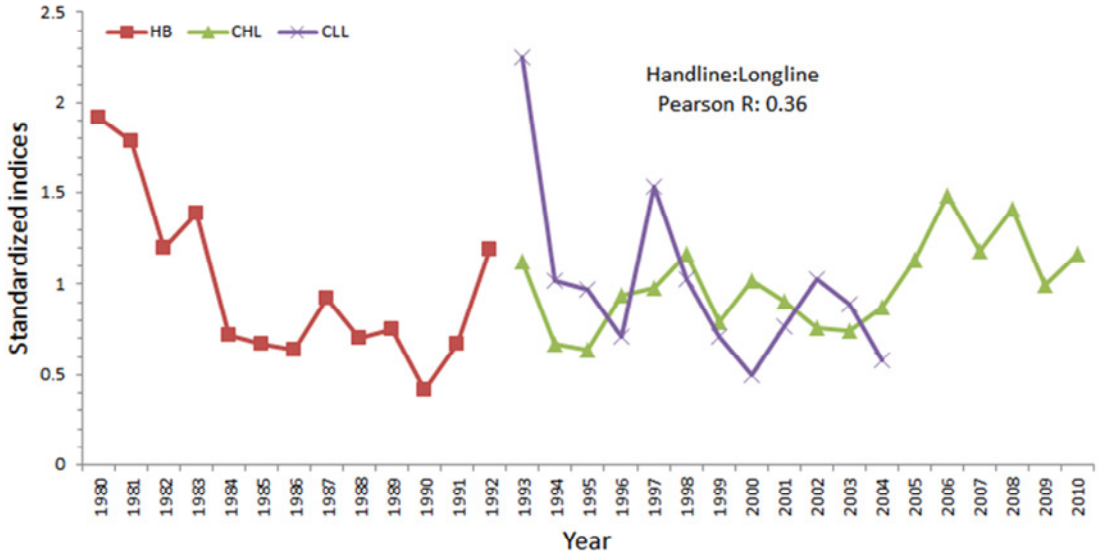
Landings and catch-at-age were estimated including those that came from north of 35N. However, CPUE was only computed for areas north of 28N and south of 35N. When we examined nominal CPUE by latitude, regardless of fishery it was higher north of 35N than the standardized composite CPUE used as an index in the assessment. Therefore, resource trends associated with increased landings north of 35 are not being indexed fully. One implication of this is the BAM model fits this increase in landings as an increase in recruitment, thus the greatest positive recruitment deviations (assessment document Fig. 3.13). This clearly has implications for projected future stock productivity.



- The Panel requested the results of the model fit to the length compositions from the base-model. The results illustrate the data conflicts and support the AW decision to exclude length compositions from the objective function.

- The review panel requested further exploration of the data to examine any period of potential overlap between the recreational and commercial indices to detect similar or dissimilar trends. When the headboat data were binned into 3 year averages (top panel in following figure) the data generally tracked the ups and downs of the other indices (commercial handline and longline, bottom panel in following figure).


Standardized index
Nominal mean
CPUE (catch/angler-hr)
0
5
10
15
1980
1985
1990
1995
2000
2005
2010


HB
CHL
CLL
Handline:Longline
Pearson R: 0.36
Standardized indices
Year

Center for Independent Experts (CIE) Reviewer's Report on the SEDAR 32/32A South Atlantic Blueline Tilefish and Gulf of Mexico Menhaden Review Workshop Morehead City, NC August 27-30, 2013

Gary D. Melvin[1]

Prepared for:

Center for Independent Experts (CIE)

[1]285 Water Street, St. Andrews, New Brunswick, Canada E5B1B8. Phone: (506)529-5874. E-mail: Gary.Lynn.Melvin@gmail.com.

Table of Contents

Executive summary 4

1.0 Background 6

1.1 Project Description 8

2.0 Review activities 10

3.0 Summary of Findings 12

3.1 South Atlantic blueline tilefish 12

3.1.1 Data Evaluation 12

3.1.2 Methods Evaluation 16

3.1.3 Evaluation of Assessment Findings 19

3.1.4 Evaluation of Stock Projections 21

3.1.5 Consideration of Uncertainties 23

3.1.6 Research Recommendations 24

3.1.7 Guidance on Improvements 26

3.2 Gulf of Mexico Menhaden 27

3.2.1 Data Evaluation 27

3.2.2 Methods Evaluation 31

3.2.3 Evaluation of Assessment Findings 34

3.2.4 Consideration of Uncertainties 35

3.2.5 Research Recommendations 37

3.2.6 Guidance on Improvements 39

4.0 Conclusions and Recommendations 40

4.1 South Atlantic Blueline Tilefish 40

4.2 Gulf of Mexico Menhaden 42

4.3 SEDAR Process 44

5.0 References 46

Appendix I: Bibliography of materials provided for review 47

Appendix II: Statement of Work 50

Appendix III: Review Panel Members and Participants 61

EXECUTIVE SUMMARY

The SEDAR 32 Review Workshop for South Atlantic Blueline Tilefish (BLT) and the Gulf of Mexico menhaden stock assessments was held was at the at the Crystal Cove Civic Center, Morehead City, NC from August 27th to 30th, 2013. The main objectives of the meeting were to provide an independent review of the assessment input parameters, methods, models, analytical approaches, sensitivity analysis, uncertainties, outputs and stock status.

All travel arrangements for the CIE reviewers were organized by the CIE, while the local venue and the meeting room was the responsibility of the SEDAR coordinator from the South Atlantic Fishery Management Council. No logistic problems were encountered with travel or the meeting facility. Background material was available almost two weeks in advance, allowing plenty of time to prepare for the meeting. The review workshop adhered closely to the agenda provided prior to the meeting, although some deviations did occur in the discussion and questions of ongoing issues. Much of the success of the Review was due to the preparation and presentations of the assessment teams, who did an excellent job of providing overviews, and their willingness to respond to the Panel's requests for clarification and additional information.

South Atlantic Blueline Tilefish: After reviewing the input data, methods, analysis and results from the South Atlantic blueline tilefish (BLT) assessment the Review Panel concluded that the decisions made by the Data and Assessment Workshops were appropriate, generally sound, robust and made use of the best available data. The models and analytical approaches used for the assessment are commonly employed to evaluate stock status of fisheries and the sensitivity runs undertaken were sufficient to estimate uncertainties in the input parameters. All were within expected levels.

The primary model used for the BLT assessment was the Beaufort Assessment Model (BAM). A surplus-production model (ASPIC) a provided comparison of model results and was complementary to the primary model. The assessment results clearly show that this stock is overfished and that overfishing is occurring. With the exception of one sensitivity run (M=0.15), all scenarios estimated $SSB_{2011} < SSB_{msy}$ and $F_{(2009\text{-}2011)} > F_{msy}$. This evaluation of stock status is supported by both the BAM and ASPIC models. The BAM base configuration, as recommended by the Assessment Workshop (AW), was used to determine stock

status. Fishing mortality in 2011 was estimated as 0.39, which is greater than the estimate of Fmsy (0.302), Spawning biomass in 2011 is estimated as 445 thousand pounds, which is less than the estimate of Minimum Stock Size Threshold (489 thousand pounds).

Projections indicate that with F=0 the stock should build to above MSY by 2014 and with F<Fmsy in 1 to 2 years. Fishing at F=MSY and F=recovery the stock should gradually increase over 5-6 year. However, fishing at the current rate will only lead to a continuing decline in SSB. Research recommendations from the Data and Assessment Workshops were reviewed and prioritized. Guidance was provided for consideration on key improvements to data and modeling approaches that should be implemented before the next assessment.

Gulf of Mexico Menhaden: After a thorough review of the input data, modeling, sensitivity runs and results the Review Panel agreed that the data decisions made by the Data and Assessment Workshops for Gulf of Mexico menhaden were appropriate, generally sound, robust and properly applied. Uncertainties in the data inputs were also appropriately acknowledged. The models used for this assessment are commonly employed to evaluate stock status of fisheries and the sensitivity runs undertaken were sufficient to estimate uncertainties in the input parameters. All were within expected levels.

The primary model used for the Gulf of Mexico menhaden assessment was the Beaufort Assessment Model (BAM), a highly flexible, integrated analysis, statistical catch-at-age model. Surplus-production model (ASPIC) results were also provided for comparison and were complementary to the primary BAM model. Numerous sensitivity analyses and exploration of alternative scenarios were presented during the Assessment Workshop, and additional model exploration and sensitivity runs were requested during the Review Workshop. Fecundity is used as a proxy for SSB.

The Review Workshop (RW) Panel examined the consistency of the input data and population biological characteristics with the abundance estimates, exploitation, and biomass estimates. Panelists felt the base BAM parameterization chosen by the AW provided the best representation of stock status.

Currently there are no formal benchmarks established for Gulf of Mexico menhaden to evaluate if the stock is overfished or if over fishing is occurring.

Benchmarks for Gulf menhaden are currently being discussed and developed by the Gulf States Marine Fisheries Commission. However, the assessment team presented a suite of potential options commonly used to evaluate stock status for other fisheries in the region. The results suggest that the stock is not over fished and over fishing is not occurring. A surplus production model confirmed the evaluations. The Review Panel agreed with the AW conclusion on stock status.

No projections were undertaken for the Gulf of Mexico menhaden. Research recommendations from the Data and Assessment Workshops were reviewed and prioritized. Guidance was provided for consideration on key improvements to the data and modeling approaches that should be implemented before the next assessment.

1.0 BACKGROUND

South Atlantic Blueline Tilefish (*Caulolatilus microps*) and the Gulf of Mexico Menhaden (Brevoortia patronus) are assessed under the Southeast Data, Assessment, and Review (SEDAR) process. SEDAR is a cooperative Fishery Management Council process to improve the quality and reliability of fishery stock assessments in the South Atlantic, Gulf of Mexico, and US Caribbean. SEDAR is managed by the Caribbean, Gulf of Mexico, and South Atlantic Regional Fishery Management Councils in coordination with NOAA Fisheries and the Atlantic and Gulf States Marine Fisheries Commissions. The process involves constituents and stakeholders and includes field personnel, biologists, fishermen, database managers, stock assessment biologists, Council members and staff throughout each stage of the process. SEDAR is a publicly open approach designed to improve the quality of stock assessment through a series of workshops for the compiling, evaluating and reporting on the assessments. There are three workshops in the SEDAR process: A data workshop to review all the available data, to determine what data are appropriate for the assessment, and to identify data and research needs; a stock assessment workshop to formulate the stock assessment, to interpret information, and to identify how uncertainty is to be incorporated into the assessment; and, a peer review workshop to provide a rigorous and independent scientific review of the completed stock assessments. At the latter workshop the Review Panel provides a consensus report on the strengths and weaknesses in the assessment and makes recommendations to fishery managers for future data and research requirements.

In the USA these independent peer reviews are coordinated and managed by the National Marine Fisheries Service's (NMFS) Office of Science and Technology through the Center for Independent Experts (CIE). CIE reviewers/experts are selected by the CIE Steering Committee to conduct an impartial and independent peer review of scientific activities without conflicts of interest. Under the terms of the contract each reviewer is required to address predetermined Terms of Reference (Appendix 2). For the SEDAR 32/32A South Atlantic Blueline Tilefish and Gulf of Mexico Menhaden Review Workshop the Review Panel consisted of:

Steve Cadrin - Review Panel Chair, SAFMC SSC
Churchill Grimes - Reviewer, SAFMC SSC
Will Patterson – Reviewer, GSMFC Appointee
Gary Melvin - CIE Reviewer, Center for Independent Experts

Stephen Smith - CIE Reviewer, Center for Independent Experts
Kevin Stokes - CIE Reviewer, Center for Independent Experts

A complete list of participants, including the analytical team, observers, and advisory committee representatives, is provided in Appendix III for both assessments.

The specific tasks to be undertaken by the CIE reviewers for the independent external Panel review were to:

1) Conduct necessary pre-review preparations, including the review of background material and reports provided by the NMFS Project Contact in advance of the peer review.

2) Participate during the panel review meeting in Morehead City, North Carolina, from 27-30 August 2013.

3) Conduct an independent peer review in accordance with the ToRs (Appendix II - Annex 2a and 2b).

4) Individually submit an independent peer review report addressed to the "Center for Independent Experts," no later than September 13, 2013. Each CIE report shall be written using the format and content requirements specified in Appendix II - Annex 1, and address each ToR in Annex 2.

.

1.1 Project Description

SEDAR 32 is a compilation of data, an assessment of the stock, and an assessment review conducted for South Atlantic blueline tilefish (BLT) and Gulf of Mexico menhaden. The CIE peer review is essentially responsible for ensuring that the best possible assessment has been provided through the SEDAR process. The South Atlantic BLT stock falls within the jurisdiction of the South Atlantic Fisheries Management Council and the state waters of North Carolina, South Carolina, Georgia, and Florida. The Gulf of Mexico menhaden stock falls within the jurisdiction of the Gulf States Marine Fisheries Commission and the state waters of Texas, Louisiana, Mississippi, Alabama, and Florida.

The tasks and timing associated with the Review Workshop begins with a pre-review of background documents: Approximately two weeks before the Review Workshop, the NMFS Project Contact sent (by electronic mail or made available at an FTP site) to the contract officer's representative (COR) the necessary background information and reports (i.e., working papers) for the reviewers to conduct the peer review, and COR then forwarded the documents to the contractor. Reviewers were responsible only for the pre-review documents that were delivered to the contractor in accordance to the Statement of Work (SoW) scheduled deadlines specified. The reviewers were responsible for reading all documents deemed as necessary in preparation for the peer review.

At the Review Workshop each reviewer shall conduct the independent peer review in accordance with the SoW and stock assessment ToRs, and shall not serve in any other role unless specified. Each reviewer shall actively participate in a professional and respectful manner as a member of the meeting review panel, and their peer review tasks shall be focused on the stock assessment ToRs. The NMFS Project Contact will be responsible for any facility arrangements (e.g., conference room for panel review meetings or teleconference arrangements). The NMFS Project Contact will also be responsible for ensuring that the Chair understands the contractual role of the reviewers. The contractor can contact the COR and NMFS Project Contact to confirm any peer review arrangements, including the meeting facility arrangements.

After the panel review meeting, each reviewer is required to prepare an independent peer review report in the forma described in SoW. This report should explain whether each stock assessment ToR was or was not completed successfully during the SEDAR meeting. If any existing BRP or their proxies are considered inappropriate, each independent report shall include recommendations and justification for suitable alternatives. If such alternatives cannot be identified, then the report shall indicate that the existing BRPs are the best available at this time. Additional questions and pertinent information related to the assessment review addressed during the meetings that were not in the ToR's may be included in a separate section at the end of an independent peer review report.

The following report to the CIE reflects my independent opinions and views on the issues and questions identified in the terms of reference, statement of work, and the above goals and objectives. The report is, however, generally consistent

with the recommendations and conclusions of the other panel reviewers. Panel members met on the final day of the meeting to review their observations, conclusions, and recommendations. Overall there was agreement amongst the panel members regarding their conclusions and recommendations. This summary report also meets the requirements for south Atlantic blueline tilefish ToR # 8 and the Gulf of Mexico menhaden ToR #7.

2.0 REVIEW ACTIVITIES

The initial phase of the review process began with the provision of background material from the Data and Assessment Workshops and other research activities/results that contributed to the decision making process for the assessments. This included the final reports from both workshops as well as a large number of reference documents associated with data inputs and assessment methodology. In addition, a conference call was organized for August 21, 2013 to review the agenda, discuss initial perception of the assessment documents and to determine if any major problems were encountered that might be corrected prior to the meeting. Only a few minor editorial inconsistencies were identified.

The Review Workshop (RW) was held at the Crystal Cove Civic Center, Morehead City, NC from August 27th to 30th, 2013. Chaired by Steve Cadrin, the Panel consisted of six members and was supported by the stock assessment teams (mostly from the NMFS Southeast Fisheries Science Centre, Beaufort Lab). Lead analyst for the south Atlantic BLT assessment and the Gulf of Mexico menhaden assessment were Kevin Craig and Amy Scheuller, respectively. Both were responsible for the majority of the presentations, addressing questions, and providing additional information requests to the Review Panel. A complete list of participants for both assessments is provided in Appendix 3.

The RW began with introductions and a general welcome by the Chair. This was followed by a few housekeeping necessities, a review of the agenda and the Terms of reference for each stock. In general, the Review adhered to the agenda provided prior to the meeting to allow participants for the different assessments to attend the presentations and discussions of interest. The first two days, one day each, was allocated to each stock. Given the relatively small number of participants, the chair was flexible with input and questions during the presentation, however, the majority of questions and discussion was reserved for

after the presentations. Priority was given to the Review Panel members, followed by the fishing industry and other observers. All participants were provided an opportunity at the end each assessment presentation to ask questions or make comments relative to the fishery and the assessment.

Once the initial formalities associated with the RW were complete, the Chair moved quickly on the Terms of Reference for each stock. In this review each member of the panel was assigned specific TOR's to provide a summary and text for inclusion in the Panel report. The CIE reviewers were requested to prepare text for two or three of the TOR's from each of the stock assessments. The two general reviewers were assigned the task of merging the input from the CIE reviewers into coherent sections for the final report. Although this was not part of the ToR for the CIE reviewers, it was discussed and agreed to, based on the necessity for each CIE reviewer to address the same ToR in their report. The meeting then proceeded with the presentations and review of each stock assessment beginning with South Atlantic BLT on day 1 followed by the Gulf of Mexico menhaden on day 2. All day Thursday August 29 and Friday morning was reserved for review, and discussion of additional analysis and sensitivity runs requested by the Review Panel, Panel discussions of the overall assessment outcome related to analysis, and the development of consensus recommendations and comments.

The RW was conducted in a professional and timely manner with the Chair providing ample opportunity for clarification and discussion of issues among the participants. Throughout the meeting all CIE reviewers played an active role in the questioning, discussion, and request for additional information upon which to base the Panel's conclusions and make recommendations. Each CIE Reviewer also contributed to the specific subset of ToR's they were assigned at the beginning of the workshop, which were subsequently used in the Review Panel Consensus Report. The main output from the Workshop/review was to conduct and summarize an independent peer review of each stock in accordance with the ToR's. The ToR's and their associated recommendations/conclusions are discussed in the section that follows.

3.0 SUMMARY OF FINDINGS

Under the terms of the contract, the CIE review report shall include an independent peer review of ToRs for each stock assessed. In this case TOR's were developed specifically for both stocks reviewed at the assessment under the SEDAR process. In this summary report the ToR's for southern Atlantic BLT will be addressed first followed by the TOR's for the Gulf of Mexico Menhaden. Each ToR and sub-term will be discussed in the context of the best available information. Readers will likely discover relative consistency and overlap in the text for the other CIE reviewers and the Review Workshop report as there was general agreement in the summary, concerns and recommendations among the panel. In addition, unlike other reviews, the CIE reviewers made a significant contribution to the text contained within the Review Workshop report. Each CIE reviewer was assigned several TOR's for each stock to summarize for the Panel Report. Consequently, and in the absence of a finalized Panel Report (to be submitted after the CIE report due date), this summary report contains some of the same material submitted to the panel chair regarding the ToR's for inclusion in the Panel Report.

3.1 South Atlantic blueline tilefish assessment review

3.1.1 Evaluate the data used in the assessment, addressing the following:

a) <u>Are data decisions made by the Data Workshop and Assessment Workshop sound and robust?</u>

Overall the decisions made by the Data and Assessment Workshops for South Atlantic BLT were generally sound and robust. The data summary presented by Kevin Craig touched on a number of the key model input data sources. During the review the panel expressed concern about several subjects including the broad geographical distribution of the stock, natural mortality, maturity-at-age, ageing/growth, quality of the landings data, and the abundance indices. Each of these issues is described below and all were discussed and resolved to the best of the Panel's ability with the available information. In some instances additional information was requested by the Review Panel and was provided by the assessment team.

The stock/management area for this stock assessment extends from Rhode Island to Florida with all BLT landings used as input to the assessment model.

Unfortunately, there are no genetic studies or tagging data available for this species to define the management area; but many species exhibit a stock boundary along the US east coast at Cape Hatteras. BLT are pelagic spawners and as a consequence, it was suggested that larvae would be expected to be wide-ranging. However, previous work on golden tilefish indicates a stock break north and south of Cape Hatteras (Katz et al.1983). The Panel expressed concern that the stock area may be too broad given that the fishery appears to be focused in a few small areas. In addition, this species is known to be highly residential, occupying scour depressions in carbonate substratum and burrows in soft bottom (Able et al.1987). Such an aggregated species may be subject to local fisheries and depletion.

Natural mortality at age for the BLT assessment was determined based on estimates of K and $L\infty$ from Von Bertalanffy growth curves using the methods of Charnov et al. (2012) and is therefore highly dependent upon the quality of the age data. Considerable uncertainty in age determination for blueline tilefish was documented by Harris et al. (2004) and in the ageing error matrix for this assessment. The Panel agreed that scaling the mean M over the older ages to 0.1 was considered reasonable given the Hoenig estimate based on maximum age. A maximum M of 0.15 and a minimum of 0.05 was used for sensitivity testing based upon a CV of 54%. However, the lack of fish of age 15 and older in the landings data suggest that either M may be higher because the maximum age of 43 is questionable due to the uncertainty in ageing or Fishing mortality (F) was much higher than assumed. This would imply that the higher M alternative should receive more attention in the sensitivity analysis than the lower M, and perhaps M estimates higher than 0.15 should be considered.

No direct estimates of maturity at age were available for BLT. For the assessment maturity-at-age was based upon estimates for golden tilefish where 50% are mature at age 3 and 100% mature at age 4. While these results indicated a relatively younger maturity than might be expected for such a long-lived fish, similar results have been reported for other long-lived species in the region. However, maturity studies of golden tilefish, also suggest that functional maturity may occur at ages older than histological maturity because of territoriality, dominance and mate choice (Grimes et al. 1988 and McBride et al. 2013). If true for BLT, then the apparent truncation of age composition may be due to harvesting.

The von Bertalanffy growth curve indicates that 98% of total growth has been completed by age 15, therefore fish aged 15 yrs and older were assigned to a

plus group. The underlying growth data were obtained from sampling recent (last couple of years) landings for fisheries that appeared to target a very narrow range of ages (3-5 yrs. for recreational and 5-8 yrs. for commercial fisheries). There were no age composition data for landings in the earlier part of the time series when it was expected that larger/older fish should have represented a higher proportion of the population given a maximum age of 43 yrs. The reliability of the underlying assumptions of the initial age composition raises issues about the current estimates of M and F, as well the assumption of flat-topped selectivity. Furthermore, age composition data do not appear to track year-classes, even though high recruitment was estimated to have occurred just prior to the period when the bulk of the data were collected. Industry suggested that there may be differences in the spatial distribution of size/age class.

Age and growth information used in the assessment was extracted from Harris et al. 2004. This study did not rigorously validate putative ages and reported a low aging precision of about 60% within two years.

A comparative ageing study showed the agreement between readers was relatively poor. The ageing error matrix assumed a normal distribution to compare the results from two BLT readers. The symmetric distribution of errors was questioned as experience suggests that older ages tend to be more likely to be underestimated due to annuli packing at the otolith margin as the fish approach the asymptotic length. However, uncertainty in age determination as measured by the ageing error matrix was considered to be relatively small in comparison to other sources of uncertainty that had been identified.

The age compositions were fitted by the assessment model, yet the length compositions were removed from the analysis due to preliminary results indicating lack of fit. In light of the uncertainties associated with the ageing data, it seemed unusual that the length composition data would not be better fitted by the model.

b) <u>Are data uncertainties acknowledged, reported, and within normal or expected levels?</u>

Based on the Data Workshop (DW), the Assessment Workshop (AW), and the information presented at the AW, the uncertainties associated with this assessment were acknowledged and reported. For almost all data, the uncertainties were within normal and expected levels, except possibly those

associated with the ageing. The percent agreement between readers decreased rapidly (5%/year) from ~50% at age 3. This poor level of agreement is unusual for the young of such long living fish, yet it represents the best available data.

c) Are data applied properly within the assessment model?

All things considered the data were applied properly within the assessment.

d) Are input data series reliable and sufficient to support the assessment approach and findings?

The commercial and recreational catch rates are key datasets for both the BAM and ASPIC models. These fishery dependent CPUE indices represent the only annual abundance indices available and were developed using the standard approaches, (i.e., fit delta-GLM models to filter out annual trends from other factors associated with the data). The recreational index characterizes the earlier period when SSB was being fished down but it actually represents a period of very low levels of catch. The index was truncated in 1992 due to the limited samples. Unfortunately, there was no overlap between when the recreational index was truncated and the two commercial indices began.

Landings and catch-at-age were estimated for the entire geographic domain of the fishery, including those that came from north of 35N. However, CPUE was only computed for areas north of 28N and south of 35N. When the Panel examined nominal CPUE by latitude, regardless of fishery, it was higher north of 35N than the standardized composite CPUE used as an abundance index in the assessment. Consequently, the increased landings north of 35N are not being fully indexed. One implication is that the BAM model fits this increase in landings as an increase in recruitment, thus producing the greatest positive recruitment deviations in the model (see assessment document Fig. 3.13). This clearly has implications for projections of future stock productivity.

Landings data were considered to be reliable since 1974 and discarding for the commercial fishery was assumed to be negligible and consistent as there are no regulatory reasons for discarding (e.g., size limits). The recreational catch has been sporadic and low relative to the commercial catch until 2006. There was considerable discussion about the reliability of the recreational landing estimates for 2006 to 2008 including the very high discard estimate in 2007. Most of these landings appeared to have occurred in North Carolina waters and there was a

suggestion that the development of a “deep-drop” fishery may have driven the increase. The decrease in 2011 was due to the implementation of a deep water closure. Examination of the Marine Recreational Information Program (MRIP) data indicated that CVs for 2006 to 2011 decreased relative to the period before and the number of sample intercepts increased, both are indicative of increased fishing activity. However, the magnitude of landings relative to the commercial landings in those same years still seemed to be unprecedented and industry participants questioned the reliability of the recreational estimates.

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

a) Are methods scientifically sound and robust?

The Beaufort Assessment Model (BAM) was used as the principal assessment tool for BLT. 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. Use of the BAM permitted the inclusion of all available types of data, including total annual removals from commercial and recreational fleets (landings and discards), age and length compositions, and indices of biomass abundance, with appropriate error distributions and use of priors on the parameters. Decisions on *a priori* data inclusion and exclusion are considered under ToR 1.

The specified assessment model used standard approaches to predicting landings, modeling growth and recruitment, and the BAM allowed an exploration of catchability and selectivity options.

The base case model and rationale for modeling decisions are well described in the AW report (section 3) and were further explored during the Review Workshop. The base case run included commercial and recreational landings, age composition data and three indices of abundance (recreational head boats, commercial long line and hand line). There was some concern that the recreational and commercial indices do not overlap but this was explored during the RW and the general patterns seem to be consistent. Length compositions were excluded by the AW due to concerns about inconstant sampling and conflicts in fitting. The AW concluded that length composition data help to inform selectivity estimates but conflicted with information in the abundance indices, did

not track year classes well, and added unnecessary noise. The Review Panel was concerned at this exclusion and the issue was explored further during the RW by looking at shadow fits comparing the base case predicted (but not fit) length compositions with the data and by examining model fits to the length composition data. The RW concluded that the residual patterns in the indices were not acceptable from the model that included length compositions, and the results could not be considered as a viable base case (or sensitivity run); the decision by the AW to exclude the length composition data was therefore upheld. Natural mortality was assumed constant through time but age-specific based on the method of Charnov (2013) and scaled consistent with maximum observed age. Steepness was fixed at 0.84 based on meta-analyses (Myers et al., 2002; Shertzer and Conn, 2012). Selectivities and catchabilities were all estimated as constant for the full assessment period (1974-2011).

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

The model was fit to the data using appropriate methods, consistent with standard practice. Analysis included iterative reweighting using the method of Francis (2011) and exploration of a variety of data configurations and parameterizations. The modeling processes and decision making resulting in a proposed base case run and sensitivity testing are well described in the AW Report and AW working documents and were further elaborated during the SEDAR 32 Review Workshop where additional diagnostics (likelihood components, weights, likelihood profiles) were made available. The modeling procedures adopted appear to be robust. Landings and discards were fit closely, and age composition data and abundance indices were fit to the degree that they are compatible and as indicated using the reweighting procedures. Landings and indices were fit using lognormal likelihoods. Age composition data were fit using robust multinomial likelihoods.

c) Are the methods appropriate for the available data?

The treatment of the data and the relative importance given to the various components were well explored by the AW and at the RW and appear appropriate. The model structure is adequate to capture the main patterns in the data.

In addition to the catch-at-age primary assessment, two biomass dynamics stock assessments were carried out using the ASPIC software, one fully age-aggregated and the other age structured. The biomass dynamics models were considered as confirmatory rather than alternative analyses, because the catch-at-age model makes fuller use of composition data and represents a more detailed investigation of population dynamics. The biomass dynamics models provide a useful comparison with the catch-at-age model results (Fig 1), which they broadly support, showing similar status of the stock in relation to MSY benchmarks (ToR 3). The biomass dynamics models are well known and the methods used were appropriately configured and implemented.

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

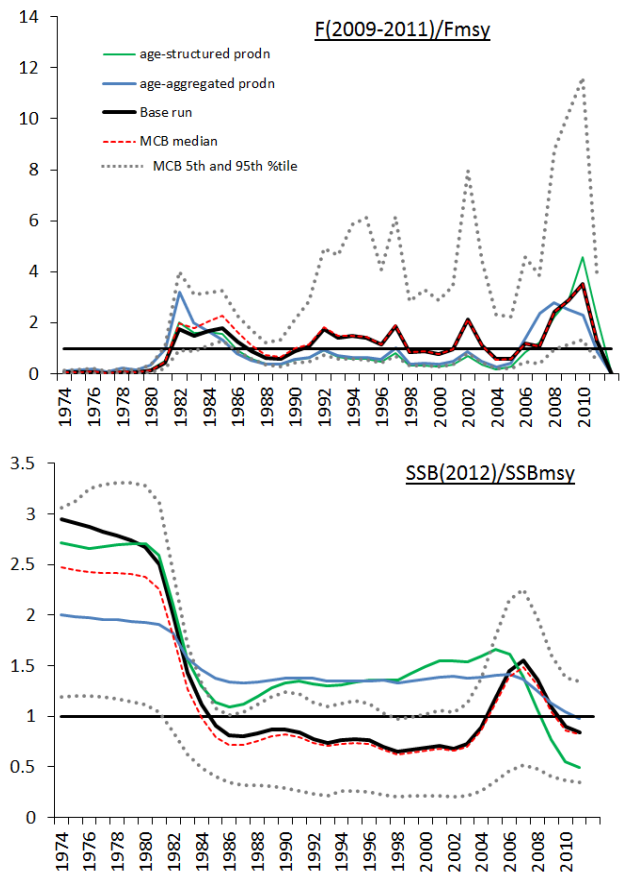


Figure 1. Trajectories of status benchmarks for the catch-at-age base case model, two biomass dynamics model runs, and the MCB analysis. Refer to key for explanation.

3.1.3 Evaluate the assessment findings with respect to the following:

a) <u>Are abundance, exploitation, and biomass estimates reliable, consistent with input data and population biological characteristics, and useful to support status inferences?</u>

The review panel examined the consistency of input data and population biological characteristics with abundance estimates, exploitation and biomass estimates. Overall the Review Panel agreed with the AW view that the base run provided the best representation of stock status and the use of MCB for projection estimates. The outputs are generally consistent with the inputs. The

review panel also noted that the MCB median estimates of biomass may also provide bases for evaluating stock status.

Initial examination of the sensitivity likelihoods suggested that a higher M may represent a creditable run and should be considered. However, additional information provided to the review panel by the assessment team supported the use of the base model M as input over the alternative of a higher M.

The Panel discussed the estimate of fishing mortality F and thought it may have been over estimated for projections because of changes in regulations and closures. However, examining the preliminary 2012 landings showed a substantial increase from 2011, thereby justifying the high F. Consideration might be given to using actual landings for future projections where 2012 is replaced with catch figures. As well the Panel suggests that 2011 be removed from three year estimate of F for 2013 and 2014.

b) <u>Is the stock overfished? What information helps you reach this conclusion?</u>

The RW Panel endorses the AW recommendation to determine stock status using the BAM base configuration. Based on the model estimates of SSB, the South Atlantic BLT is overfished by definition. Spawning biomass in 2011 is estimated as 445 thousand pounds, which is less than the estimate of Minimum Stock Size Threshold (489 thousand pounds), so the stock is overfished. SSB has been below SSB_{msy} for the past two years (2010-2011). The majority of viable sensitivity runs indicate that the SSB_{2011} was < SSB_{msy}. The only exception was the increase in M which indicated the SSB was greater than the SSB_{msy}. This was considered unlikely based on additional sensitivity runs requested by the Review Panel regarding M. Production model outputs of population status generally agreed with the catch-at-age model and indicate a B/B_{msy} of less than 1 in 2011.

c) <u>Is the stock undergoing overfishing? What information helps you reach this conclusion?</u>

Based on the Beaufort Assessment Model (BAM) base run estimate of fishing mortality (F), overfishing is occurring for the South Atlantic BLT. The geometric mean F over the past 3 years (F(2009-2011)/Fmsy) was greater (2.37) than 1.0

and has been for the past several years. The dramatic decrease in F(2011) to 1.30 was primarily the result of a fishery closure. Production model outputs all indicate an average F/Fmsy well in excess of 1.0.

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

The Review Panel concurs with the AW use of the Beverton-Holt spawner recruit relationship to predict the recruitment of age -1 fish with a note of caution. The stock recruitment relationship was considered a major source of uncertainty. Recruitment estimates and MSY management quantities are based on a steepness that could not be estimated and was fixed at 0.84. Alternative proxies for MSY such as $F_{X\%}$ were examined but they too require an assumption about steepness.

e) Are the quantitative estimates of the status determination criteria for this stock reliable? If not, are there other indicators that may be used to inform managers about stock trends and conditions?

Based on the methods explored and the information available to the Review Panel, the quantitative estimates for determination of stock status were considered reliable and within the bounds of the uncertainties identified in the Assessment Document and the Review Panel's report.

3.1.4. Evaluate the stock projections, addressing the following:

a) Are the methods consistent with accepted practices and available data?

The methods used by the AW for projections are consistent with accepted practices in the region and elsewhere, and the available data. Initially the Review Panel had several concerns regarding the use of the Monte Carlo and bootstrap (MCB) approach as a measure of precision and to compute uncertainty. The MCB analysis is considered an approximation of uncertainty for an individual run. For BLT unconverged and unrealistic runs were removed (3200 reduced to 3043) from the analysis, however, there was still the possibility of including nonsense variable inputs that individually could occur within the established parameter bounds, but combined (biologically) could not, resulting in unrealistic outputs of R_0 and F_{msy}. All unfiltered runs were given equal weight and were included in the

estimate of uncertainty. These limitations were identified in the assessment report. In addition, there was the mixing of deterministic and stochastic parameters, the latter introducing process error. The review panel concluded that although the MCB approach is a common approach used in SEDAR assessments to estimate uncertainty, the results may be different if a true Bayesian approach was applied.

b) Are the methods appropriate for the assessment model and outputs?

In essence, the MCB process generates a stochastic version of the BAM model by introducing process error to the model components of natural mortality and steepness. The means of management quantities (MSY, B_{MSY}, F_{MSY}) from the MCB runs do not equal estimates from base run. Comparing estimates from the deterministic and stochastic version of the Schaefer population model indicated that the deterministic solutions for F_{MSY}, were not correct for the stochastic version (Bousquet et al. 2008). In fact, the direction of the differences observed between the MCB based estimates and those of the base run are in the direction predicted by the equations for the Schaeffer model. That is, F_{MSY} from the stochastic runs will be less than the deterministic estimates from the base run, MSY will be slightly higher for the stochastic estimates and B_{MSY} slightly lower. The size of the differences will be a function of the amount stochastic error in the model. These differences will not be apparent when looking only at ratio benchmarks as in Figure 1. It is important to note that for consistency, if MCB is used for projections, the MCB estimates of the management quantities should also be used for evaluating stock status.

c) Are the results informative and robust, and useful to support inferences of probable future conditions?

Projection results are informative and robust within the range of observations and inputs from the MCB. Currently F is estimated as the mean of the three previous years. Given the observed rapid changes in F and the preliminary landings estimates for 2012 and 2013 consideration might be given to using actual landings for future projections or to drop the 2011 from the estimate of F for 2013 and 14.

d) Are key uncertainties acknowledged, discussed, and reflected in the projection results?

The assessment report identified and evaluated uncertainties associated with the assessment through the MCB approach. The report identifies the degree of uncertainty associated with M, ageing error, steepness, model component weights, indices and recruitment deviations. Some concern was expressed by the review panel about the appropriateness of using the mean F (high relative to the time series) for the previous three years given the high F's of 2009 and 2010 and the low value for 2011 for projections. However examination of the preliminary landings for 2012 and 2013 support the use of a large F. Preliminary landings data were requested the Review Panel and provided by the assessment team during the meeting.

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

Uncertainty was explored in the assessment models using extensive sensitivity runs and likelihood profiling, retrospective analyses and Monte Carlo Bootstrapping (MCB). All of the methods used are standard and commonly employed in stock assessments. The AW reported on the various analyses with more material being provided and used in discussion at the RW. The application of methods appeared to be comprehensive and appropriately focused. Sensitivity runs as variants of the base case run were numerous and good information was provided on the impacts on fits (through detailed likelihood components and also weighting diagnostics, SDNRs, likelihood profiles, etc). However, such runs can only look at what the model structure accommodates and cannot consider, for example, processes such as fishery or environmentally induced geographic changes in distribution of the stock or fishery induced local depletion. There was much discussion at the RW on these issues and on data inclusion or exclusion in indices to represent stock abundance. Ultimately, the stock assessment assumed a single dynamic pool of fish and there was insufficient data at this time to support investigating alternative hypotheses. With the exception of this structural uncertainty, the other uncertainties in the assessment and its outputs have been appropriately and comprehensively considered.

Issues considered in sensitivity runs include variations in M and steepness, alternative maturity vector, adjustment of model weights and exclusion of each series of indices, allowing catchability to vary, inclusion of ageing error, and allowing recreational selectivity to be dome shaped. Issues of uncertainty not

covered explicitly in sensitivity tests include the quantum of landings assigned to recreational landings and especially discards in 2005-2007.

For the MCB approach a total of 3200 realizations were made using M and h values drawn from specified pdf's and with the landings, indices and age composition data bootstrapped. Each realization 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 3200 realizations). However, it should be noted that reweighting can have major implications for fitting and parameter estimation and each realization may not be feasible. The degree to which this may, or may not, matter is model and data specific. As all realizations are afforded equal weight in determining distributions of outputs there is in general need for care in interpreting MCB results. For BLT, the SDNRs for all sensitivity tests are surprisingly good when runs are made using the base case weights. This is encouraging; however, this is no guarantee that the base case weights would in any way be appropriate for a specific M and h combination drawn from the pdfs, some may be incompatible.

Notwithstanding, the RW was comfortable that the AW had fully explored uncertainty to the extent possible and that the characterization of benchmark trajectories (Figure 1) and hence stock status (ToR 3) and projections (ToR 4) are suitable for informing management decisions.

3.1.6. Consider the research recommendations provided by the Data and Assessment workshops and make any additional recommendations or prioritizations warranted.

Research recommendations for BLT were provided in the data and assessment Workshop documents (see reports) and were reviewed by the Panel in the context of the assessment. The Panel noted that many of these recommendations were broad in scope and reflected concerns across a range of deep-water species. The review Panel confined their attention to those specific to the stock assessment of South Atlantic BLT.

While the panel recognizes the necessity for research on stock structure, it recommended starting with the available information on describing the differences in demographics/life history characteristics of the species over the

range of the management area, before embarking on a broad scale genetic study. Additionally, the available information on habitat in the areas listed should be evaluated before initiating any new studies.

The Review Panel concluded that given this is an age-based assessment, comparison and calibration studies for the age determination should receive high priority along with marginal increment analysis to determine if the opaque zone is formed annually. In other words, conduct an age validation study. Protocols should be established for ageing, improved precision and the inclusion of age data from multiple readers/labs. Many species would probably benefit from expanding the MRIP program to include the collection of hard parts for aging for sampling.

The collection of information to better describe spawning season and spawning periodicity could probably start with fishery-dependent sources, but will need data from fishery-independent programs to cover the range of the species. The latter program would probably have to be tailored to provide samples across the deep-water snapper/grouper complex.

Studies of discard mortality should be a low priority given the current negligible discard rate in the commercial fishery. The collection of additional information on discards and catch (e.g, lengths, ageing material) is important especially for the areas north of Cape Hatteras, but would likely require an observer program be developed for all fisheries focusing on the deep-water snapper/grouper complex.

The BAM model is reliant on historical information and any data on size compositions, maximum size, etc., which can be retrieved from historical recreational fishing photos, could be quite useful. One of the main issues raised about the recreational fishery concerned the high landings in the mid-late 2000s, especially the high landing and discard estimates for 2007. Closer scrutiny of these estimates requires data at higher resolution than was apparently available for this stock assessment.

Developing a fishery-independent survey, sampling of deep-water habitats may elucidate habitat characteristics, and spatial distributions of BLT and other deep-water reef fishes. If a sufficient time series is developed, a fishery-independent index may be developed. However, the small size of the fishery may be prohibitive to the development of a fishery independent index of abundance for this species.

Recommendations/suggestions on possible ways to improve the SEDAR process are discussed under Conclusion and Recommendations (Section 4.0) of this report and are applicable to both stock assessments reviewed under SEDAR 32.

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

The South Atlantic BLT assessment relies upon fishery dependent indexes of abundance to inform the Beaufort Assessment Model. No fishery independent indices are available for this stock. As such, the geographical distribution, seasonal movement, spawning, and consistency of the fishery over time have all had an impact on the indices and contribute to the uncertainty associated with the assessment. Whether or not the stock is truly a single spawning population distributed through the described range, or a series of multiple spawning components is, unknown given its broad spatial occurrence along the Atlantic coast. Changes in the state proportional contribution to total landings/catches from the commercial handline and longline fisheries implies a divergence from a more southerly dominated (Florida and South Carolina) fishery during the 1980's to a northern (North Carolina, especially above Cape Hatteras) focused fishery in more recent years. The reason(s) for these observed changes in landings are unknown. The changes in catch and subsequent catch rates used as indices of abundance may be a function of population dynamics, serial depletion, or a northerly migration in response to environmental variability. Further investigation of this issue should be undertake before the next assessment to insure the current commercial indices represent changes in abundance and not the adaption of the fishing fleets to availability. Development of a fishery independent index of abundance would help to resolve some of these issues, but is unrealistic given the small size of the fishery.

During the initial review and presentation of the stock assessment it was unclear that the commercial CPUE indices were being truncated or trimmed at Cape Hatteras, thereby excluding the effort data north of this area. Landings data used in the assessment model included all reported catches taken throughout the entire range of the stock. Given that a large portion of recent landings are being reported north of Cape Hatteras are not included in the commercial CPUE indices the effects these omissions on the abundance indices are unknown. The review panel suggests the increased catches be addressed and that this

apparent inconsistency between the indices and the fishery be resolved before the next assessment.

The BLT assessment uses three CPUE indices based on information from the Headboat (1980-92), handline (1993-10) and longline (1993-04) fisheries, with no data for 2011 due to a commercial and recreational closure. The recreational headboat time series was terminated due to the low number of trips/catches. No overlapping years between the Headboat index and the other two indices were used in the assessment suggesting uncertainty in the scaling of the indices. Limited information was, however, available for the headboat fishery over the entire time series, although sampling was poor after 1992. During the review the Panel requested additional analysis of the headboat time series to investigate if there were consistencies in CPUE patterns. When the headboat CPUE estimates were binned into three year averages the data generally tracked the ups and downs of the other indices, supporting the observed trends in abundance from the two commercial time series. The headboat data should be investigated further to determine if the times series can be extended, especially given the recent increases in headboat catches and sampling since 2008.

3.2 Gulf of Mexico menhaden assessment review

3.2.1. Evaluate the data used in the assessment, addressing the following:

a) <u>Are data decisions made by the Assessment Workshop sound and robust?</u>

The Review Panel agreed that the data decisions made by the Data and Assessment Workshops were generally sound and robust. Furthermore, after a thorough review of the data and a few requests for additional information the Panel concurred that the data generally were applied properly. Uncertainties in data inputs were also appropriately acknowledged.

b) <u>Are data uncertainties acknowledged, reported, and within normal or expected levels?</u>

Uncertainties in the data were acknowledged and reported within normal or expected levels. The Review Panel discussed several data concerns during the

workshop and requested additional information from the assessment team for clarification, to address the issues and to verify the uncertainties. These concerns fell into the broad categories of stock structure, landings, reproductive biology, and ageing and are discussed below.

The Gulf of Mexico menhaden stock ranges from western Florida through the northern Gulf of Mexico (GOM) to Campeche, Mexico, but their abundance is greatest in the north central GOM. After reviewing the information available the Review Panel agreed with the conclusion that no evidence exists to contradict the assumption that the population in the north central GOM constitutes a unit stock. However, there was some uncertainty as to whether population trends and demographics were similar in eastern and western portions of the species' range as the assessment and data tended to focus mostly on the central portion of the range where reduction fishery is concentrated.

Landings estimates were judged to be accurate as the reduction fishery is responsible for reporting nearly all landings and there has been a log system in place since 1964 for that fishery, including daily catch records. Cooperation by industry in supplying information to NMFS is impressive (weekly electronic reporting, 100% participation in voluntary program, access for port sampling and provision of freezer space for samples). The decision to start the landings time series in 1977 was quite reasonable given concerns about data quality for age composition data prior to 1977, inexplicable truncated age distributions in the early 1970s, species identification/composition and other issues with these early data as noted in past stock assessments. However, sensitivity analyses were conducted with the longer times series of age composition included.

The protocol for sampling menhaden to estimate length and age composition of the reduction fishery landings involves taking a haphazard sample from the top of a given boat's hold. Members of the Review Panel questioned if such a method provided a representative sample of the catch. Results from a 2012 study involving alternative sampling protocols suggest that sampling only from the top of a hold provides a biased sample of the catch, specifically underestimating numbers of older fish in the catch. For example, age-3 fish constituted less than 3% of the catch when sampled with the traditional method, while they were approximately 20% of samples taken from the start, middle, or end of hold pump-out. No age-4 fish were present in samples taken with the traditional method, but they constituted approximately 5% of landings sampled during the start or middle of pump-out.

There was some discussion about the lack of older fish in the estimated catch-at-age being due to older fish being less vulnerable to the fishery, if the spatial distribution of fish is age-specific. Major grounds for the fishery occur within 10 miles of the coast, but the species is estimated to extend out to 60 miles. Therefore, if older fish are found farther offshore or in smaller, non-targeted schools, then they may not be vulnerable to the fishery. This would conflict with the assumption of a logistic selectivity function for the reduction fishery. However, based on early-season catches that are generally taken farther offshore (10-20 miles), older fish do not appear to be farther offshore during the fishing months.

The Review Panel requested additional information on the spatial distribution of the fishery. The analysis, presented by the assessment team, on fishery hotspots composites for 2008, 2009 and 2011 fishing years was informative, but a longer time series of year-specific hotspots would have provided information on the spatial overlap between fishery- and fishery-independent indices of abundance used in the assessment. Plotting these hotspots may provide insight into the potential distribution of older fish off western Louisiana, as well as to the east of Alabama/Mississippi, areas not covered by either the seine or gillnet survey indices used in the assessment.
Fecundity was used as a metric for reproductive potential to compute a proxy for spawning stock biomass. A relationship produced in the early 1980s relating numbers of eggs to female length was used in this assessment to estimate length-specific fecundity in the model, thus larger, assumed older, fish are estimated to produce more eggs per individual than younger fish. Ovarian egg number may be a reliable index of SSB if all the ovary samples were at the same stage of reproductive development, but that would seem unlikely for existing menhaden fecundity data. Furthermore, Gulf menhaden have a protracted spawning season and are assumed to be an indeterminate batch spawner. If older fish produce more batches or higher quality eggs, then their contribution to stock-specific fecundity would be underestimated using the current approach. Lastly, it was noted that while fecundity is a common metric of reproductive potential in the region, it is not specified in the management plan as part of the stock status determination criteria.

Several issues exist with the aging protocols. Multiple scale readers aged fish in the 1960s to early 1970s, but only a single reader has aged fish since the 1970s. No formal protocol for aging appears to exist. Three informal analyses of aging accuracy or repeatability produced questionable results (e.g., 71% agreement

between otolith and scale derived age estimates; 82% agreement between age estimates from scales aged in 2005 and again in 2012; and, substantial disagreement in age estimates from the 1970s versus contemporary re-ageing of those samples). Given the short-lived nature of the fish, reader error of even one year can cause substantial bias in an age-based assessment. While the computed aging error matrix did not indicate directional bias, the assumption that the error was symmetric about ages precluded any other error pattern from being estimated. In most fishes, age of older individuals tends to be underestimated by scales as annuli pack at the scale margin and become difficult to discern. In fact, the assessment team conveyed that aging older menhaden (>2 yrs) with scales is more difficult than aging younger fish.

There was evidence of a shift in the estimated age composition of landings from mostly age-1 fish in the1960s-80s to mostly age-2 fish in more recent decades. Several hypotheses for the shift are discussed in the AW Report (e.g., habitat alteration affecting recruitment of juvenile fish in estuaries, decreased fishing mortality, recent contractions in the spatial distribution of the fishery, changing spatial distribution of age-1 menhaden, or the influence of hypoxic habitats on spatial distribution). However, re-aging of a sub-sample of scales from three years among each decade from 1970s to the 2000s indicated ages of fish sampled in the early portion of the time series, when multiple scale readers existed, may have been underestimated. Therefore, the AW removed the earliest years of the time series. No other bias in the ageing was identified.

c) Are data applied properly within the assessment model?

All things considered the data were applied properly within the assessment using standard approaches for standardizing variables and estimating the unknowns.

d) Are input data series reliable and sufficient to support the assessment approach and findings?

Natural Mortality (M) was estimated from an extensive tagging study conducted in the early 1970s (Ahrenholz 1991). The resultant estimate of M (1.22 y-1) was then scaled with the Lorenzen (1996) function to estimate declining M with age. After some discussion the RW concluded this approach was sound.

The Gulf of Mexico menhaden assessment is based on two reliable and apparently representative indices of abundance: a juvenile seine index and an adult gillnet index. A number of available fishery dependent and independent indices of abundance were excluded from being used in the model for various reasons by the AW. The SEAMAP juvenile trawl index which was highly correlated with the seine index was included in the SEDAR 27 assessment model, but dismissed here because it was judged that trawls are not efficient for pelagic species, the spatial extent of the survey was not appropriate for the resource, and the western portion of the survey has species identification problems. A research recommendation was included in the AW report for genetic sampling by size to resolve the species identification problem. Some concern was expressed by the RW that the gillnet index was limited to the Louisiana series. Data from the western and eastern portions of the resource area were excluded because of mixed species catches and species identification problems. Many of the potential surveys lacked ages (i.e., collection of hard parts for ageing). A larval survey was not used because of poor winter coverage, complex recruitment dynamics from larvae to fishery recruitment, and problems with species identification. Members of the Review Panel questioned why some of these indices were excluded prior to assessing their impact on model fit, such as through likelihood profiling.

A question arose about whether there could be a cryptic biomass of older (>3 years) fish that is not encountered by the fishery. Amy Schueller, the assessment lead, responded that older fish are captured in the gillnet survey. Further, if fish school by size or age, then small schools of larger, older fish may not be targeted by purse seiners.

Overall the Review Panel felt that the data input series were utilized appropriately and are sufficient to support the assessment outputs.

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

a) <u>Are methods scientifically sound and robust?</u>

The Beaufort Assessment Model (BAM) was used as the principal assessment tool for the Gulf of Mexico menhaden stock. 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. Use of the BAM permitted the inclusion of all available data types, including total annual removals from the commercial fleets (and the very small recreational catches), age and length compositions, and indices of biomass abundance, with appropriate error distributions and use of priors on parameters. Decisions on *a priori* data inclusion and exclusion are considered under ToR-1. The specified assessment model used standard approaches to predicting landings and modeling recruitment, and the BAM allowed an exploration of catchability and selectivity options.

The base case model and rationale for modeling decisions are well described in the AW report and were further explored during the RW. 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. Natural mortality was estimated from tagging data, assumed to be constant through time, and was scaled among ages based on the method of Lorenzen (1996). Steepness of the Beverton-Holt spawner recruit (S-R) relationship was fixed at 0.7. Selectivities and catchabilities were all estimated as constant for the full assessment period (1977-2011).

b) <u>Are assessment models configured properly and used consistent with standard practices?</u>

The model was fit to the data using appropriate methods, consistent with standard practice. Analysis included iterative reweighting using the method of Francis (2011) and exploration of a variety of data configurations and parameterizations. The modeling processes and decision making that resulted in a proposed base case run and sensitivity testing are well described in the AW Report, which includes information on Likelihood components, weighting, SDNRs by data component and weight, likelihood profiles, etc. Further diagnostics were made available and elaborated during the RW.

c) <u>Are the methods appropriate for the available data?</u>

The treatment of the data and the relative importance given to the various components were well explored by the AW and at the RW and appear appropriate. The model structure is adequate to capture the main patterns in the data, thus the modeling procedures adopted appear to be robust. Landings and indices were fit using lognormal likelihoods. Age composition data were fit using

robust multinomial likelihoods. Landings were fit closely by the model, as were age composition data. Trends in abundance indices were generally fit by the model, but greater residuals existed for extreme index values (ie., those at the beginning or the end of the time series) that were not closely fit by the model.

In addition to the catch-at-age primary assessment, an age-aggregated biomass dynamics stock assessment was carried out using the ASPIC software. The biomass dynamics models was considered 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, hence is better able to capture higher frequency changes in indices (e.g., recent high indices and catches). The biomass dynamics model provides a useful comparison with the catch-at-age model, which it broadly supports without capturing recent population changes. A number of sensitivity tests were carried out on the biomass dynamics model which demonstrated the robustness of conclusions based upon it. The biomass dynamics model used, implemented with ASPIC, is well known and commonly used in fisheries assessment. The methods were appropriately configured and implemented.

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 specified pdf's. Outputs provided are the quantiles of the distribution resulting from application of the MCB simulations. Each simulation applies to a single BAM model using the weights developed for the base case run. No reweighting procedures are used for individual realizations.

The MCB approach was used to generate a stochastic version of the BAM model by introducing process error to the model components of natural mortality and steepness. Means of management quantities (MSY, B_{MSY}, F_{MSY}) from the MCB runs do not equal estimates from the base run. 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). F_{MSY} 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 B_{MSY} slightly lower. The size of the differences will be a function of the amount of stochastic error in the model. These differences will not be apparent when looking only at ratio benchmarks.

3.2.3. 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?

The RW Panel examined the consistency of input data and population biological characteristics with abundance estimates, exploitation, and biomass estimates. Panelists felt the base BAM parameterization chosen by the AW provided the best representation of stock status and felt the usage of MCB for projection estimates was appropriate.

The menhaden fishery landings are dominated by age-2 fish with fishing occurring after this age group has spawned at least once. However, the selectivity pattern for the reduction fishery was flat topped, and there was uncertainty about the presence of older fish (age-3 and older) in fishery-independent gillnet catches versus their general absence in reduction fishery landings.

Very high F's were estimated within time series considered, especially during the 1980s. Fishing mortality has subsequently declined to range between 1.0 and 3.5 y-1. The 2011 full F was 2.36 y-1, with much lower Fs estimated for the older ages.

b) Is the stock overfished? What information helps you reach this conclusion?

Currently there are no formal benchmarks established for Gulf of Mexico menhaden to evaluate stock status in terms of being overfished or overfishing. The assessment team presented a suite of potential options for the Review Panel to evaluate. For SSB_{2011}/SSB_{med}, $SSB_{2011}/SSB_{30\%}$, $SSB_{2011}/SSB_{35\%}$, and $SSB_{2011}/SSB_{40\%}$ all BAM base run values exceeded 1.0. A surplus production confirmed the evaluations. Therefore, it is unlikely the Gulf menhaden stock would be evaluated to be overfished given commonly applied benchmarks in the region. The Review Panel agrees with the AW statement that the Gulf menhaden stock is not overfished.

c) Is the stock undergoing overfishing? What information helps you reach this conclusion?

F_{msy} was defined as infinite because of the stock population dynamics and the nature of the fishery. This assumption is valid as long as the fishery selectivity remains unchanged. The surplus production model produced results relative to estimates of MSY with no indication of exceeding the criteria typically used to evaluate overfishing. The Review Panel agrees with the AW general statement that no overfishing is occurring.

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

No stock recruitment relationship was developed for this stock. Information on recruitment was based on the seine survey and the reproductive output based on population fecundity from BAM numbers at age.

e) Are the quantitative estimates of the status determination criteria for this stock reliable? If not, are there other indicators that may be used to inform managers about stock trends and conditions?

Managers are currently defining the goals and objectives for the Gulf menhaden fishery. Quantitative estimates for stock status determination are not defined and under discussion.

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

Uncertainty was explored in the Gulf menhaden assessment modeling using extensive sensitivity runs and likelihood profiling, retrospective analyses, and MCB. All of the methods used are standard and widely used. The AW reported on the various analyses. The assessment team provided additional material when requested, which was 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 with good information being

provided/reported 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 but 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.

Issues considered in sensitivity runs include scaling and the form of M, S-R steepness and form, adjustment of model weights and exclusion of each series of indices, alternative selectivity assumptions for the commercial reduction fishery, start year, inclusion/exclusion of indices, alternative weightings, and alternative growth specification.

The MCB is alluded to above under ToR-2. A total of 5,000 realizations were made using M and h values drawn from specified probability density functions (PDFs) and with the landings, indices, and composition data bootstrapped. A total of 4,068 realizations were used to compile the final MCB quantile plots with realizations discarded if they did not converge or showed other poor behavior. The process for discarding realizations was not discussed in detail. Each realization 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 4,068 realizations). It should be noted that reweighting can have major implications for fitting and parameter estimation and that each realization may not be feasible, possibly explaining why some realizations did not converge. The degree to which this may or may not matter is model and data-specific. As all realizations are afforded equal weight in determining distributions of outputs there is in general need for care in interpreting MCB results. For menhaden, the SDNRs for all sensitivity tests are surprisingly good (except for one case) when runs are made using the base case weights. However, this is no guarantee that for specific M and h combinations drawn from the PDFs, which may be incompatible, the base case weights would necessarily be appropriate.

Notwithstanding the above concern, the RW was comfortable that the AW had fully explored uncertainty to the extent possible and that the characterization of benchmark trajectories and hence stock status (ToR-3) are suitable for informing management decisions.

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

The RW panel suggested there should be an evaluation of the utility of using ovarian egg number as a proxy for SSB and notes that this will depend not only on biological considerations but also on age validation and errors, and selectivity determination. Ultimately, the utility of egg numbers versus SSB will depend on how status benchmarks and control rules are determined.

The Louisiana gillnet survey used in the menhaden assessment has a number of different mesh sizes and concern was expressed about developing a single index over these different mesh sizes, especially given the length frequencies presented in the assessment (AW Report, Fig. 5.44). The RW panel recommends evaluating the efficacy of developing separate indices by mesh or accounting for the different mesh sizes within the same index.

The panel did not see value in undertaking genetic studies to further elucidate Gulf menhaden population structure given the fishery operates in the center of the species distribution and it is unlikely that information gained would justify the expense of additional analyses. However, the RW panel did see considerable benefit in using simpler genetic techniques, such as DNA barcoding, to aid in species identification, which is currently problematic in several fishery-independent surveys conducted in peripheral range areas of Texas, Alabama, and Florida.

Throughout the course of the DW and AW, a number of items were identified as important research topics for future stock assessments. The RW Panel evaluated the various items listed and developed a consensus priority list that differs somewhat from those presented.

DATA ELEMENT	RECOMMENDATION	PRIORITY
FISHERY-INDEPENDENT ADULT INDEX	Collect Gulf menhaden ageing structures (scales and otoliths) from alternate fishing gears (e.g., gillnets and trawls) to determine gear selectivity. Need to expand efforts to age menhaden by state agencies. Determine readability of whole versus sectioned otoliths.	**Very High**
FISHERY-INDEPENDENT ADULT INDEX	Improve species identifications at the periphery of the Gulf menhaden's range in Texas and Alabama/Florida waters.	**Very High**
FISHERY-DEPENDENTSURVEYS	A Gulf-wide aerial survey may be a useful tool to measure adult Gulf menhaden abundance; "groundtruthing" for fish size and age and school size, would be a necessary adjunct to the survey.	**High**
FISHERY-DEPENDENTSURVEYS	Additional sampling needs to be conducted to address the homogeneity of the catch in the hold of the reduction fishery vessels at the four Gulf menhaden factories. Supplemental samples must be pulled from throughout the fishhold during the pumpout process to determine if the assumption that the traditional 'last set of the trip' accurately represents the age composition for the catch for the given port-week	**High**
FISHERY-INDEPENDENTJUVENILE INDEX	Improve species identifications at the periphery of the Gulf menhaden's range in Texas and Alabama/Florida waters.	**High**
FECUNDITY/MATURITY	The seminal study on fecundity and sexual maturity of Gulf menhaden was published thirty years ago (Lewis and Roithmayr 1981) with data from the late 1970s. It is recommended that a study should be initiated to re-examine the reproductive biology of gulf menhaden in the northern Gulf of Mexico, which includes updating fecundity estimates, maturity schedules(GSI), and sex ratios. Any study needs to reinvestigate whether gulf menhaden are determinant or in determinant spawners. Survey necessarily needs to include spawning from winter collections.	**High**
GENETICS AND STOCKSTRUCTURE	Identification of menhaden-specific nuclear DNA markers (preferably microsatellites or SNP's) using a lab-based DNA library screening techniques. Evaluation of these markers for use in genetic studies of Gulf menhaden	**Low**

GENETICS AND STOCK STRUCTURE	Identification in the Clupeid literature of potential new heterologous nuclear DNA markers (preferably microsatellites or SNP's) which will potentially enhance genetic sampling in Gulf menhaden.	**Low**
GENETICS AND STOCKSTRUCTURE	Reassessment of Gulf menhaden throughout its range using a larger, more informative genetic panel of markers than that described in Anderson (2006).	**Low**

Recommendations/suggestions on possible ways to improve the SEDAR process are discussed under Conclusion and Recommendations (Section 4.0) of this report and are applicable to both stock assessments reviewed under SEDAR 32.

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

The Review Panel expressed some concern about the selectivity associated with the Louisiana gillnet survey used as an index of adult abundance for the assessment model. Probability density functions of length samples depict an expected distribution pattern for the smaller mesh sizes; however, the larger mesh sizes show a broad size distribution uncharacteristic of this gear type. The gillnet index also samples larger, and presumed older, fish than the commercial reduction fishery. This implies that the large fish are not being captured by the fishery and supports the dome shaped reduction fishery selectivity of 0.35 for ages 3 and 4 in the BAM base run assessment parameterization. However, a recent study to investigate sampling protocols in the reduction fishery, albeit small, suggests that the traditional reduction fishery sampling method may be missing larger fish when samples are only collected from the top of the hold. Further investigation of traditional sampling protocols and potential sampling bias should be undertaken before the next assessment.

The index is used to characterize the coast-wide stock following the age specific selectivity vector within the model. Understanding of the gillnet selectivity and reduction fishery sampling could resolve several fitting problems with the index and uncertainties in the model and should be considered for the next scheduled assessment.

4.0 CONCLUSIONS AND RECOMMENDATIONS

The recommendations and conclusions in accordance with the ToR are described in detail in Section 3 of this report. Each term of reference and their sub-components identified for South Atlantic BLT and the Gulf of Mexico menhaden have been addressed. Section 4 contains a brief overview of the recommendations for each assessment reviewed. Many of the responses to the specific questions are redundant with the comments and recommendations presented in Section 3; however, they serve to reinforce the statements.

4.1 South Atlantic blueline tilefish

The Review Panel evaluated the data methods used in the assessment and concluded that overall the decisions made by the Data and Assessment Workshop for South Atlantic BLT were appropriate, generally sound and robust. The models used for the assessment are commonly employed to evaluate the stock status of other fisheries and the sensitivity runs undertaken were sufficient to estimate uncertainties in the input parameters and model outputs. All were within expected levels.

The assessment findings clearly show that, by benchmark definition, this stock is overfished and that overfishing is occurring. With the exception of one sensitivity run (M=0.15), all scenarios estimated $SSB_{2011} < SSB_{msy}$ and F_{2011} and $F_{(2009-2011)} > F_{msy}$. This evaluation of stock status is supported by both the BAM and ASPIC models. Spawning biomass in 2011 from the base run was estimated as 445 thousand pounds, which represents 91% estimate of Minimum Stock Size Threshold (489 thousand pounds). Fishing mortality was relatively high between 2008 and 2010, but dropped dramatically in 2011 due to a commercial and recreational closure. The 2011 F, however, remained above the defined threshold for overfishing in the base run and all sensitivity run, except for the higher M scenario.

The methods used by the AW for projections are consistent with accepted practices in the region and elsewhere, and the available data. Initially the Review Panel had a few concerns regarding the use of the MCB approach as a measure of precision and to compute uncertainty, but in the end concluded that the approach was appropriate. The review panel did, however, pointed out that although the MCB approach is a common approach used in SEDAR

assessments to estimate uncertainty, the results may be different if a true Bayesian approach was applied. After reviewing the preliminary 2012 and 2013 landings the Panel recommends that consideration be given using a different F in the projections. Current projections indicate that for an F=0 the stock should build to above MSY by 2014 and F<Fmsy in 1 to 2 years. Fishing at F=MSY and F=recovery the stock should gradually increase over 5-6 year. However, fishing at the current level will only lead to a continuing decline in SSB.

Uncertainty in the assessment models was explored using extensive sensitivity runs and likelihood profiling, retrospective analyses and Monte Carlo Bootstrapping (MCB). All of the methods used are standard and commonly employed in stock assessments. The Review Panel had an extended discussion and concern related to the geographical distribution of the resource and the single dynamic pool assumption for this stock (see section 3.0). Excluding the structural uncertainties, the uncertainties in the assessment and the outputs were deemed to have been appropriately and comprehensively considered. The RW felt that the AW had fully explored uncertainty to the extent possible and that the characterization of benchmark trajectories (Fig 1) and hence stock status and projections are suitable for informing management decisions.

The research recommendations provided by the Data and Assessment Workshops were reviewed and prioritized by the Panel (Section 3.1.6). The Panel noted that many of these recommendations were broad in scope and reflected concerns across a range of deep-water species; however, comments were restricted to those associated with South Atlantic BLT. The panel supported research recommends related to demographics, life history characteristics, and ageing of the species. The development of a fishery independent index would have benefits, but is likely unrealistic for such a small resource.

Several issues remain unclear for BLT related to stock structure and the indices which addressed would help to improve future assessments. The assessment assumes a single spawning population distributed throughout the described range, yet a series of multiple spawning components over its broad spatial occurrence could also explain many of the observations. Do the commercial indices used in the assessment represent changes in abundance or the adaption of the fishing fleets to availability? Further investigation of this issue should be undertaken. Currently a large portion of recent landings being reported north of Cape Hatteras are not included in the commercial CPUE indices. The review panel suggests the increased catches be addressed and that this apparent

inconsistency between the indices and the fishery be resolved before the next assessment. Finally, the headboat time series, terminated in 1992 should be revisited to determine if the series can be extended, especially given the recent increases in headboat catches and sampling since 2008.

4.2 Gulf of Mexico Menhaden

The Review Panel agreed that the data decisions made by the Data and Assessment Workshops were generally sound, robust and based on the best available data. Furthermore, after a thorough review of the data and a few requests for additional information, the Review Panel concurred that the data were generally applied properly. Uncertainties in data inputs were also appropriately acknowledged. The models used for the assessment are commonly employed to evaluate stock status of fisheries and the sensitivity runs undertaken sufficiently to estimate uncertainties in the input parameters. All were within expected levels.

Uncertainties in the assessment were acknowledged, examined, reported, and within normal or expected levels. The Review Panel discussed several data concerns during the workshop related to the broad categories of stock structure, landings, reproductive biology, and ageing that are briefly summarized below. All things considered, the data were applied properly within the assessment using standard approaches for standardizing variables and estimating the unknowns. Furthermore, the data input series were utilized appropriately and are sufficient to support the assessment outputs.

The Beaufort Assessment Model (BAM) was used as the principal assessment tool for the Gulf of Mexico menhaden stock. This permitted the inclusion of all available data types, including total annual removals from the commercial fleets (and small recreational catches), age and length compositions, and indices of biomass abundance, with appropriate error distributions and use of priors on parameters.

The base case model and rationale for modeling decisions are well described and fit to the data using appropriate methods, consistent with standard practice. The base case run and sensitivity testing includes information on likelihood components, weighting, SDNRs by data component and weight, and likelihood

profiles. Landings and indices were fit using lognormal likelihoods, while age composition data were fit using robust multinomial likelihoods. The treatment of the data and the relative importance given to the various components were well explored at both the Assessment and the Review workshops and appear appropriate. The Panel agreed that the model structure is adequate to capture the main patterns in the data, thus the modeling procedures adopted appear to be robust.

Uncertainty was explored in the Gulf menhaden assessment modeling using extensive sensitivity runs and likelihood profiling, retrospective analyses, and MCB. All of the methods used are standard and widely used.

For menhaden, the SDNRs for all sensitivity tests are surprisingly good (except for one case) when runs were made using the base case weights. Issues considered in sensitivity runs include scaling and the form of M, S-R steepness and form, adjustment of model weights and exclusion of each series of indices, alternative selectivity assumptions for the commercial reduction fishery, start year, inclusion/exclusion of indices, alternative weightings, and alternative growth specification. In the end the Panel felt comfortable that the AW had fully explored uncertainty to the extent possible and that the characterization of benchmark trajectories and hence stock status are suitable for informing management decisions.

Currently there are no formal benchmarks established for Gulf of Mexico menhaden to evaluate if the stock is overfished or if over fishing is occurring. The assessment team presented a suite of potential options used by other fisheries in the region. Comparison suggests that the stock is not over fished. A surplus production confirmed the evaluations. The Review Panel agreed with the AW conclusion on stock status. Fmsy was defined as infinite because of the stock population dynamics and the nature of the fishery. This assumption is valid as long as the fishery selectivity remains unchanged. The surplus production model showed no indication of exceeding the criteria typically used to evaluate overfishing. The Review Panel concurred with the AW general statement that it is unlikely overfishing is occurring.

The research recommendations provided by the Data and Assessment Workshops were reviewed and prioritized by the Panel (Section 3.2.5). Priority should be given to evaluation of the use egg number as a proxy for SSB and the

selectivity of gillnets used for the adult index of abundance. The Review Panel did not completely agree with the AW priorities. In particular the panel did not see value in undertaking genetic studies to further elucidate Gulf menhaden population structure, given the fishery operates in the center of the species distribution and it is unlikely that information gained would justify the expense of additional analyses. However, the RW panel did see considerable benefit in using simpler genetic techniques, such as DNA barcoding, to aid in species identification, which is currently problematic in several fishery-independent surveys conducted in peripheral range areas of Texas, Alabama, and Florida.

Key improvements in data and modeling approaches that may help with the next assessment are focused around the gillnet survey and sampling. The Review Panel expressed concern about selectivity associated with the Louisiana gillnet survey used as an index of adult abundance for the assessment model. Large fish taken in the survey are not being captured by the fishery suggesting a dome shaped in the reduction fishery selectivity for ages 3 and 4 in the BAM base run parameterization. However, traditional reduction fishery sampling methods appear to be missing larger fish when only collected from the top of the hold. Understanding of the gillnet selectivity and reduction fishery sampling could resolve several fitting problems with the index and uncertainties in the model and should be considered for the next scheduled assessment. It was also noted during the workshop that the seine survey may be discontinued. This could have serious implications for future assessments. The current surveys should be maintained and if possible improved.

4.3 The SEDAR Process

The Southeast Data, Assessment, Review (SEDAR) process provides an open and comprehensive approach to stock assessments. Through a series of three workshops, the data are reviewed, the assessment methods adopted and parameterized, and the end product peer reviewed by a panel of experts. This provides multiple opportunities for stakeholder input and discussion at all levels. SEDAR 32 was no exception. The Review Workshop examined two stock assessments; the South Atlantic BLT and the Gulf of Mexico menhaden. During (and before) the RW, vast amounts of background information via reports, scientific papers and presentations were provided to the Panel members. These documents formed the foundation of the assessments which were then

complemented by the assessment team's presentation and response to the Panel's questions.

The coordinators are to be congratulated on the selection of the Panel members. Between them they covered a broad spectrum of knowledge related to stock assessments, fish biology, surveying/sampling design and even local knowledge of the fishery. The chair was also well versed in stock assessment modeling methods and approaches. To his credit he ran a successful meeting and kept the entire group on track and on time. It was also a pleasure to have members of the fishery and the industry participates in the meeting. Their input on the fishery and local factors was extremely valuable when trying to understand some of the complexities or apparent inconsistencies in the data and the observations.

Overall the process was well coordinated and a positive experience as a reviewer with little room for improvement. The Panel members worked well together to come to agreement on issues and to form a consensus view. This may not always be the case depending upon the members. The only slightly negative aspect of the process was that all the detailed analysis and decisions regarding the assessments had been made prior to the Review Workshop. For several of the issues/discussions it would have been nice to explore the alternatives in more detail, as with most sources of uncertainty, the devil is in the detail. It was also noted that additional stocks were originally scheduled to be included in the review. Several Panel members felt strongly that two full stock assessments were about all that could be accommodated in the time allocated if a comprehensive review was expected. The recommendation would be to keep the number of stocks reviewed at a single meeting to a minimum, preferably no more than two, if they are as extensive as those for SEDAR 32.

DISCLAIMER

The information in this report has been provided for review purposes only. The author makes no representation, expressed or implied, as to the accuracy of the information and accepts no liability whatsoever for either its use or any reliance placed on it.

5.0 REFERENCES (only those acknowledged in this report)

Able, K.W., D.C. Twichell, C.B. Grimes and R.S. Jones. 1987. Tilefishes of the genus Caulolatilus construct burrows in the sea floor. Bull. Mar. Sci. 40(1):1-10.

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 254:65–75,

Charnov, E.L., H. Gislason and J.G. Pope. 2012. Evolutionary assembly rules for fish life histories. Fish and Fisheries.

Grimes, C. B., C. F. Idelberger, K. W. Able, and S. C. Turner. 1988. The reproductive biology of tilefish, Lopholatilus chamaeleonticeps Goode and Bean, from the United States Mid-Atlantic Bight, and the effects of fishing on the breeding system. Fish. Bull. 86:745–762.

Harris, P.J., D.M. Wyanski, and P.T. Powers Mikell. 2004. Age, growth and reproduction of blueline tilefish along the southeastern coast of the United States, 1982-1999. Transactions of the American Fisheries Society 133:1190-1204.

Katz, S.J., C.B. Grimes and K.W. Able. 1983. Delineation of tilefish, Lopholatilus chamaeleonticeps, stocks along the U.S. east coast and in the Gulf of Mexico. Fish. Bull. 81(1):41-50.

McBride, R.S., T.E. Vidal and S.X. Cadrin. 2013. Changes in size and age at maturity of the northern stock of Tilefish (Lopholatilus chamaeleonticeps) after a period of overfishing. Fish. Bull. 111:161–174.

Appendix I: Bibliography of materials provided for review.

South Atlantic Blueline Tilefish

Document #	Title	Authors
Documents Prepared for the Review Workshop		
SEDAR32-RW01	The Beaufort Assessment Model (BAM) with application to blueline tilefish: mathematical description, implementation details, and computer code	NMFS-SFB 2013
SEDAR32-RW02	Catch Curves for blueline tilefish from the commercial handline and longline fleets	NMFS-SFB 2013
SEDAR 32 - AW	South Atlantic Blueline Tilefish – SECTION III: Assessment Workshop Report, August 2013.	SEDAR - 2013
Reference Documents		
SEDAR32-RD01	List of documents and working papers for SEDAR 4 (Caribbean – Atlantic Deepwater Snapper Grouper) – all documents available on the SEDAR website.	SEDAR 4
SEDAR32-RD02	Comparison of Reef Fish Catch per Unit Effort and Total Mortality between the 1970s and 2005–2006 in Onslow Bay, North Carolina	Rudershausen et al. 2008
SEDAR32-RD03	Source document for the snapper-grouper fishery of the South Atlantic region.	SAFMC 1983
SEDAR32-RD04	FMP, regulatory impact review, and final environmental impact statement for the SG fishery of the South Atlantic region	SAFMC 1983
SEDAR32-RD05	Age, growth and reproductive biology of blueline tilefish along the southeastern coast of the United States, 1982-99	Harris et al. 2004
SEDAR32-RD06	List of documents and working papers for SEDAR 9 (Gulf of Mexico Gray Triggerfish, Greater Amberjack, and Vermillion Snapper)	SEDAR 9
SEDAR32-RD07	Estimated Conversion Factors for Adjusting MRFSS Gulf of Mexico Red Snapper Catch Estimates and Variances in 1981-2003 to MRIP Estimates and Variances	Rios et al. 2012
SEDAR32-RD08	Estimates of Historic Recreational Landings of Spanish Mackerel in the South Atlantic Using the FHWAR Census Method	Brennan and Fitzpatrick 2012
SEDAR32-RD09	Excerpt from ASMFC Atlantic Croaker Stock Assessment & Peer Review Reports	ASMFC 2003

	2003 – Information on Jacquard Index	
SEDAR32-RD10	Survival estimates for demersal reef fishes released by anglers	Collins 1994
SEDAR32-RD11	Indirect estimation of red snapper (Lutjanus campechanus) and gray triggerfish (Balistes capriscus) release mortality	Patterson et al. 2002
SEDAR32-RD12	Estimating discard mortality of black sea bass (Centropristis striata) and other reef fish in North Carolina using a tag-return approach	Rudershausen et al. 2010
SEDAR32-RD13	Commercial catch composition with discard and immediate release mortality proportions off the southeastern coast of the United States	Stephen and Harris 2010
SEDAR32-RD14	Migration and Standing Stock of Fishes Associated with Artificial and Natural Reefs on Georgia's Outer Continental Shelf	Ansley & Harris 1981
SEDAR32-RD15	Age, Growth, and Reproductive Biology of the Gray Triggerfish (*Balistes capriscus*) from the Southeastern United States, 1992-1997	Moore 2001
SEDAR32-RD16	Size, growth, temperature, and the natural mortality of marine fish	Gislason et al. 2010
SEDAR32-RD17	Evolutionary assembly rules for fish life histories	Charnov et al. 2012
SEDAR32-RD18	A Review for Estimating Natural Mortality in Fish Populations	Siegfried & Sansó

Gulf of Mexico Menhaden

Document #	Title
SEDAR32A - **1.1**	2012 Dec 07 SEDAR32 Analyst CC summary (file type: Word)
SEDAR32A - **1.2**	2012 Dec 07 SEAMAP INDEX presentation 1 (file type: PowerPoint)
SEDAR32A - **1.3**	2012 Dec 07 Revised CDFR CPUE Index (file type: PowerPoint)
SEDAR32A - **1.4**	2012 Dec 07 Gillnet index (file type: PowerPoint)
SEDAR32A - **2.1**	2012 Dec 11 SEDAR32A Analyst CC summary (file type: Word)
SEDAR32A - **2.2**	2012 Dec 11 Langseth SEAMAP INDEX presentation 1 (file type: PowerPoint)
SEDAR32A - **3.1**	2013 Feb 05 SEDAR32A Analyst CC summary (file type: Word)
SEDAR32A - **3.2**	2013 Feb 05 Seine Index02-2013 (file type: PowerPoint)
SEDAR32A - **3.3**	2013 Feb 05 SEAMAP INDEX presentation 4 (file type: PowerPoint)

SEDAR32A - **3.4**	2013 Feb 05 Gillnet index and other data (file type: PowerPoint)
SEDAR32A - **4.1**	2013 March 8 SEDAR32A CC summary (file type: Word)
SEDAR32A - **4.2**	2013 March 8 Data webinar_draft (file type: PowerPoint)
SEDAR32A - **4.3**	2013 March 8 SEAMAP INDEX presentation 5 (file type: PowerPoint)
SEDAR32A - **5.1**	2013 March 26 SEDAR32A CC summary (file type: Word)
SEDAR32A - **5.2**	2013 March 26 Data webinar_final (file type: PowerPoint)
SEDAR32A - **6.1**	2013 May 9 SEDAR32A Conf Call Summary (file type: Word)
SEDAR32A - **6.2**	2013 May 9 BAM AW_webinar 1 (file type: PowerPoint)
SEDAR32A - **7.1**	2013 June 4 SEDAR32A Conf Call Summary (file type: Word)
SEDAR32A - **7.2**	2013 June 4 BAM AW_webinar 2 (file type: PowerPoint)
SEDAR32A - **7.3**	2013 June 4 Surplus production models for Gulf menhaden (file type: PowerPoint)
SEDAR32A - **8.0**	SEDAR32A AW Summary (file type: PDF)
SEDAR32A – **RW01**	The Beaufort Assessment Model (BAM) with application to Gulf menhaden: mathematical description, implementation details, and computer code (file type: PDF)
SEDAR32A - **RW02**	Benchmarks in Excel (file type: Excel)
SEDAR32A	Assessment Report
SEDAR32A	Assessment Supplement (assessment report figures, tables, appendices ONLY)

Appendix II: Statement of Work for Dr. Gary Melvin

External Independent Peer Review by the Center for Independent Experts

SEDAR 32 South Atlantic blueline tilefish and Gulf of Mexico menhaden assessment review

BACKGROUND

Scope of Work and CIE Process: The National Marine Fisheries Service's (NMFS) Office of Science and Technology coordinates and manages a contract providing external expertise through the Center for Independent Experts (CIE) to conduct independent peer reviews of NMFS scientific projects. The Statement of Work (SoW) described herein was established by the NMFS Project Contact and Contracting Officer's Representative (COR), and reviewed by CIE for compliance with their policy for providing independent expertise that can provide impartial and independent peer review without conflicts of interest. CIE reviewers are selected by the CIE Steering Committee and CIE Coordination Team to conduct the independent peer review of NMFS science in compliance the predetermined Terms of Reference (ToRs) of the peer review. Each CIE reviewer is contracted to deliver an independent peer review report to be approved by the CIE Steering Committee and the report is to be formatted with content requirements as specified in **Annex 1**. This SoW describes the work tasks and deliverables of the CIE reviewer for conducting an independent peer review of the following NMFS project. Further information on the CIE process can be obtained from www.ciereviews.org.

SCOPE

Project Description SEDAR 32 will be a compilation of data, an assessment of the stock, and an assessment review conducted for South Atlantic blueline tilefish and Gulf of Mexico menhaden. The CIE peer review is ultimately responsible for ensuring that the best possible assessment has been provided through the SEDAR process. The South Atlantic blueline tilefish stock is within the jurisdiction of the South Atlantic Fisheries Management Council and the state waters of North Carolina, South Carolina, Georgia, and Florida. The Gulf of Mexico menhaden stock is within the jurisdiction of the Gulf States Marine Fisheries Commission and the state waters of Texas, Louisiana, Mississippi, Alabama, and Florida. . The Terms of Reference (ToRs) of the peer review are attached in **Annex 2a and 2b**.

OBJECTIVES

Requirements for CIE Reviewers: Three CIE reviewers shall have the necessary qualifications to complete an impartial and independent peer review in accordance with the tasks and ToRs described in the SoW herein. The CIE reviewers shall have expertise in stock assessment, statistics, fisheries science, and marine biology sufficient to complete the tasks of the scientific peer-review described herein. Each CIE reviewer's

duties shall not exceed a maximum of 14 days to complete all work tasks of the peer review described herein.

Location of Peer Review: Each CIE reviewer shall participate during a panel review meeting to conduct the independent peer review in Morehead City, North Carolina, from 27-30 August 2013.

Statement of Tasks: Each CIE reviewer shall complete the following tasks in accordance with the SoW and Schedule of Milestones and Deliverables herein.

Tasks prior to the meeting: The contractor shall independently select qualified reviewers that do not have conflicts of interest to conduct an independent scientific peer review in accordance with the tasks and ToRs within the SoW. Upon completion of the independent reviewer selection by the contractor's technical team, the contractor shall provide the reviewer information (full name, title, affiliation, country, address, email, and FAX number) to the contractor officer's representative (COR), who will forward this information to the NMFS Project Contact no later than the date specified in the Schedule of Milestones and Deliverables. The contractor shall be responsible for providing the SoW and stock assessment ToRs to each reviewer. The NMFS Project Contact will be responsible for providing the reviewers with the background documents, reports, foreign national security clearance, and other information concerning pertinent meeting arrangements. The NMFS Project Contact will also be responsible for providing the Chair a copy of the SoW in advance of the panel review meeting. Any changes to the SoW or ToRs must be made through the COR prior to the commencement of the peer review.

Foreign National Security Clearance: Foreign National Security Clearance will not be necessary for this review because the panel review meeting will be conducted at a non-governmental facility.

Pre-review Background Documents: Approximately two weeks before the peer review, the NMFS Project Contact will send (by electronic mail or make available at an FTP site) to the COR the necessary background information and reports (i.e., working papers) for the reviewers to conduct the peer review, and the COR will forward these to the contractor. In the case where the documents need to be mailed, the NMFS Project Contact will consult with the COR on where to send documents. The reviewers are responsible only for the pre-review documents that are delivered to the contractor in accordance to the SoW scheduled deadlines specified herein. The reviewers shall read all documents deemed as necessary in preparation for the peer review.

Tasks during the panel review meeting: Each reviewer shall conduct the independent peer review in accordance with the SoW and stock assessment ToRs, and shall not serve in any other role unless specified herein. **Modifications to the SoW and ToRs shall not be made during the peer review, and any SoW or ToRs modifications prior to the peer review shall be approved by the COR and contractor.** Each reviewer shall

actively participate in a professional and respectful manner as a member of the meeting review panel, and their peer review tasks shall be focused on the stock assessment ToRs as specified herein. The NMFS Project Contact will be responsible for any facility arrangements (e.g., conference room for panel review meetings or teleconference arrangements). The NMFS Project Contact will also be responsible for ensuring that the Chair understands the contractual role of the reviewers as specified herein. The contractor can contact the COR and NMFS Project Contact to confirm any peer review arrangements, including the meeting facility arrangements.

Tasks after the panel review meeting: Each reviewer shall prepare an independent peer review report, and the report shall be formatted as described in **Annex 1**. This report should explain whether each stock assessment ToR was or was not completed successfully during the SEDAR meeting. If any existing BRP or their proxies are considered inappropriate, each independent report shall include recommendations and justification for suitable alternatives. If such alternatives cannot be identified, then the report shall indicate that the existing BRPs are the best available at this time. Additional questions and pertinent information related to the assessment review addressed during the meetings that were not in the ToRs may be included in a separate section at the end of an independent peer review report.

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

1) Conduct necessary pre-review preparations, including the review of background material and reports provided by the NMFS Project Contact in advance of the peer review.
2) Participate during the panel review meeting in Morehead City, North Carolina, from 27-30 August 2013.
3) Conduct an independent peer review in accordance with the ToRs (**Annex 2a and 2b**).
4) No later than September 13, 2013, each CIE reviewer shall submit an independent peer review report addressed to the "Center for Independent Experts," and sent to Mr. Manoj Shivlani, CIE Lead Coordinator, via email to shivlanim@bellsouth.net, and Dr. David Sampson, CIE Regional Coordinator, via email to david.sampson@oregonstate.edu. Each CIE report shall be written using the format and content requirements specified in Annex 1, and address each ToR in **Annex 2**.

Schedule of Milestones and Deliverables: CIE shall complete the tasks and deliverables described in this SoW in accordance with the following schedule.

22 July 2013	CIE sends reviewer contact information to the COR, who then sends this to the NMFS Project Contact

12 August 2013	NMFS Project Contact sends the stock assessment report and background documents to the CIE reviewers.
27-30 August 2013	Each reviewer participates during panel review meeting and conducts an independent peer review
13 September 2013	CIE reviewers submit draft CIE independent peer review reports to the CIE Lead Coordinator and CIE Regional Coordinator
27 September 2013	CIE submits CIE independent peer review reports to the COR
4 October 2013	The COR distributes the final CIE reports to the NMFS Project Contact and regional Center Director

Modifications to the Statement of Work: This 'Time and Materials' task order may require an update or modification due to possible changes to the terms of reference or schedule of milestones resulting from the fishery management decision process of the NOAA Leadership, Fishery Management Council, and Council's SSC advisory committee. A request to modify this SoW must be approved by the Contracting Officer at least 15 working days prior to making any permanent changes. The Contracting Officer will notify the COR within 10 working days after receipt of all required information of the decision on changes. The COR can approve changes to the milestone dates, list of pre-review documents, and ToRs within the SoW as long as the role and ability of the CIE reviewers to complete the deliverable in accordance with the SoW is not adversely impacted. The SoW and ToRs shall not be changed once the peer review has begun.

Acceptance of Deliverables: Upon review and acceptance of the CIE independent peer review reports by the CIE Lead Coordinator, Regional Coordinator, and Steering Committee, these reports shall be sent to the COR for final approval as contract deliverables based on compliance with the SoW and ToRs. As specified in the Schedule of Milestones and Deliverables, the CIE shall send via e-mail the contract deliverables (CIE independent peer review reports) to the COR (William Michaels, via William.Michaels@noaa.gov).

Applicable Performance Standards: The contract is successfully completed when the COR provides final approval of the contract deliverables. The acceptance of the contract deliverables shall be based on three performance standards:
(1) The CIE report shall completed with the format and content in accordance with **Annex 1**,
(2) The CIE report shall address each ToR as specified in **Annex 2**,
(3) The CIE reports shall be delivered in a timely manner as specified in the schedule of milestones and deliverables.

Distribution of Approved Deliverables: Upon acceptance by the COR, the CIE Lead Coordinator shall send via e-mail the final CIE reports in *.PDF format to the COR. The COR will distribute the CIE reports to the NMFS Project Contact and Center Director.

Support Personnel:

William Michaels, Program Manager, COR
NMFS Office of Science and Technology
1315 East West Hwy, SSMC3, F/ST4, Silver Spring, MD 20910
William.Michaels@noaa.gov Phone: 301-427-8155

Manoj Shivlani, CIE Lead Coordinator
Northern Taiga Ventures, Inc.
10600 SW 131st Court, Miami, FL 33186
shivlanim@bellsouth.net Phone: 305-383-4229

Roger W. Peretti, Executive Vice President
Northern Taiga Ventures, Inc. (NTVI)
22375 Broderick Drive, Suite 215, Sterling, VA 20166
RPerretti@ntvifederal.com Phone: 571-223-7717

Key Personnel:

NMFS Project Contact:

Julia Byrd, SEDAR Coordinator
4055 Faber Place Drive, Suite 201
North Charleston, SC 29405
julia.byrd@safmc.net Phone: 843-571-4366

Annex 1: Format and Contents of CIE Independent Peer Review Report

1. The CIE independent report shall be prefaced with an Executive Summary providing a concise summary of the findings and recommendations, and specify whether the science reviewed is the best scientific information available.

2. The main body of the reviewer report shall consist of a Background, Description of the Individual Reviewer's Role in the Review Activities, Summary of Findings for each ToR in which the weaknesses and strengths are described, and Conclusions and Recommendations in accordance with the ToRs. The CIE independent report shall be a stand-alone document for others to understand the weaknesses and strengths of the science reviewed. The CIE independent report shall be an independent peer review of each ToRs.

3. The reviewer report shall include the following appendices:

   Appendix 1: Bibliography of materials provided for review
   Appendix 2: A copy of the CIE Statement of Work

Annex 2a: Terms of Reference for the Peer Review

SEDAR 32 South Atlantic blueline tilefish assessment review

1. Evaluate the data used in the assessment, addressing the following:
   e) Are data decisions made by the Data Workshop and Assessment Workshop sound and robust?
   f) Are data uncertainties acknowledged, reported, and within normal or expected levels?
   g) Are data applied properly within the assessment model?
   h) 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.
   d) Are methods scientifically sound and robust?
   e) Are assessment models configured properly and used consistent with standard practices?
   f) Are the methods appropriate for the available data?
3. Evaluate the assessment findings with respect to the following:
   f) Are abundance, exploitation, and biomass estimates reliable, consistent with input data and population biological characteristics, and useful to support status inferences?
   g) Is the stock overfished? What information helps you reach this conclusion?
   h) Is the stock undergoing overfishing? What information helps you reach this conclusion?
   i) Is there an informative stock recruitment relationship? Is the stock recruitment curve reliable and useful for evaluation of productivity and future stock conditions?
   j) Are the quantitative estimates of the status determination criteria for this stock reliable? If not, are there other indicators that may be used to inform managers about stock trends and conditions?
4. Evaluate the stock projections, addressing the following:
   e) Are the methods consistent with accepted practices and available data?
   f) Are the methods appropriate for the assessment model and outputs?
   g) Are the results informative and robust, and useful to support inferences of probable future conditions?

h) Are key uncertainties acknowledged, discussed, and reflected in the projection results?

5. Consider how uncertainties in the assessment, and their potential consequences, are addressed.
   - 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
   - Ensure that the implications of uncertainty in technical conclusions are clearly stated.
6. Consider the research recommendations provided by the Data and Assessment workshops and make any additional recommendations or prioritizations warranted.
   - Clearly denote research and monitoring that could improve the reliability of, and information provided by, future assessments.
   - Provide recommendations on possible ways to improve the SEDAR process.
7. Provide guidance on key improvements in data or modeling approaches which should be considered when scheduling the next assessment.
8. Prepare a Peer Review Summary summarizing the Panel's evaluation of the stock assessment and addressing each Term of Reference. The CIE reviewers are contracted to conduct an independent peer review, therefore the contractual responsibilities of the CIE reviewers do not include the preparation of the Peer Review Summary.
   - Each CIE reviewer may assist the Chair of the panel review meeting with contributions to the Summary Report, based on the terms of reference of the review.
   - Each CIE reviewer is not required to reach a consensus, and should provide a brief summary of the reviewer's views on the summary of findings and conclusions reached by the review panel in accordance with the ToRs.

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 2b: Terms of Reference for the Peer Review

SEDAR 32A Gulf of Mexico menhaden assessment review

1. Evaluate the data used in the assessment, addressing the following:
   e) Are data decisions made by the Assessment Workshop sound and robust?
   f) Are data uncertainties acknowledged, reported, and within normal or expected levels?
   g) Are data applied properly within the assessment model?
   h) 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.
   d) Are methods scientifically sound and robust?
   e) Are assessment models configured properly and used consistent with standard practices?
   f) Are the methods appropriate for the available data?
3. Evaluate the assessment findings with respect to the following:
   f) Are abundance, exploitation, and biomass estimates reliable, consistent with input data and population biological characteristics, and useful to support status inferences?
   g) Is the stock overfished? What information helps you reach this conclusion?
   h) Is the stock undergoing overfishing? What information helps you reach this conclusion?
   i) Is there an informative stock recruitment relationship? Is the stock recruitment curve reliable and useful for evaluation of productivity and future stock conditions?
   j) Are the quantitative estimates of the status determination criteria for this stock reliable? 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.
   - 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
   - Ensure that the implications of uncertainty in technical conclusions are clearly stated.

5. Consider the research recommendations provided by the Assessment workshop and make any additional recommendations or prioritizations warranted.
   - Clearly denote research and monitoring that could improve the reliability of, and information provided by, future assessments.
   - 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 Panel's evaluation of the stock assessment and addressing each Term of Reference. The CIE reviewers are contracted to conduct an independent peer review, therefore the contractual responsibilities of the CIE reviewers do not include the preparation of the Peer Review Summary.
   - Each CIE reviewer may assist the Chair of the panel review meeting with contributions to the Summary Report, based on the terms of reference of the review.
   - Each CIE reviewer is not required to reach a consensus, and should provide a brief summary of the reviewer's views on the summary of findings and conclusions reached by the review panel in accordance with the ToRs.

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 32/32A South Atlantic Blueline Tilefish and Gulf of Mexico Menhaden Review Workshop

Tuesday
9:00 a.m. Convene
9:00a.m. – 9:30a.m. Introductions and Opening Remarks **Coordinator**
- Agenda Review, TOR, Task Assignments
9:30a.m. – 12:00p.m. Assessment Presentation and Discussion (BLT*) TBD
12:00p.m. – 1:30p.m. Lunch Break
1:30 p.m. - 3:30 p.m. Panel Discussion **Chair**
- *Assessment Data & Methods*
- Identify additional analyses, sensitivities, corrections
3:30p.m. – 3:45 p.m. Break
3:30 p.m. - 5:00 p.m. Panel Discussion **Chair**
-Continue deliberations
5:00p.m. – 6:00p.m. Panel Work Session **Chair**
***Tuesday Goals*:** Initial BLT* presentation completed, sensitivities and modifications identified.

Wednesday
8:30 a.m. – 12:00 p.m. Assessment Presentation and Discussion (GM) TBD**
12:00 a.m. – 1:30 p.m. Lunch Break
1:30 p.m. - 3:30 p.m. Panel Discussion **Chair**
- *Assessment Data & Methods*
- Identify additional analyses, sensitivities, corrections
3:30p.m. – 3:45 p.m. Break
3:30 p.m. - 5:00 p.m. Panel Discussion **Chair**
-Continue deliberations
5:00p.m. – 6:00p.m. Panel Work Session **Chair**
***Wednesday Goals*:** Initial GM** presentation completed, sensitivities and modifications identified.

Thursday
8:30 a.m. – 12:00 p.m. Panel Discussion **Chair**
- Review additional analyses, sensitivities
12:00 a.m. – 1:30 p.m. Lunch Break
1:30 p.m. – 3:30 p.m. Panel Discussion **Chair**
-Continue deliberations
3:30 p.m. – 3:45 p.m. Break
3:45 p.m. – 5:00 p.m. Panel Discussion **Chair**
-Consensus recommendations and comments
5:00 p.m. - 6:00 p.m. Panel Work Session **Chair**
Thursday Goals: Final sensitivities identified, preferred models selected, projection approaches approved, Summary report drafts begun.

Friday
8:00 a.m. – 10:30 a.m. Panel Discussion **Chair**
- Review additional analyses, final sensitivities
- Projections reviewed.
10:30 a.m. – 10:45 p.m. Break
10:45 a.m. – 1:00 p.m. Panel Discussion or Work Session **Chair**
- Review Consensus Reports
1:00 p.m. ADJOURN
Friday Goals: Complete assessment work and discussions. Final results available. Draft Summary Report reviewed.

Appendix III: List of Participants

The following provides a list of participants in attendance at the SEDAR 32 Review Workshop for each stock assessment.

South Atlantic Blueline Tilefish.

Review Workshop Panelists

Steve Cadrin	Review Panel Chair	SAFMC SSC
Churchill Grimes	Reviewer	SAFMC SSC
Will Patterson	Reviewer	GSMFC Appointee
Gary Melvin	Reviewer	CIE
Stephen Smith	Reviewer	CIE
Kevin Stokes	Reviewer	CIE

Analytical Team

Kevin Craig	Lead analyst, SA BLT	NMFS Beaufort
Amy Scheuller	Lead analyst, GoM menhaden	NMFS Beaufort
Kyle Shertzer	Assessment Team	NMFS Beaufort
Erik Williams	Assessment Team	NMFS Beaufort
Katie Andrew	Assessment Team	NMFS Beaufort
Rob Cheshire	Assessment Team	NMFS Beaufort
Robert Leaf	Assessment Team	USM

Observers

Dewey Hemilright	Fishing Industry	Commercial, NC
Robert Johnson	Fishing Industry	Charter/Headboat

Council Representative

Michelle Duval	Council Member	SAFMC

Council and Agency Staff

Julia Byrd	SEDAR Coordinator	SEDAR
Julie O'Dell	Administration	SEDAR/SAFMC
Michael Errigo	Fishery Biologist	SAFMC Staff
Steve VanderKooy	IJF Program Coordinator	GSMFC
Jessica Stephen	Fishery Biologist	SERO
Brian Langseth	Observer	SEFSC Beaufort
Joe Smith	Observer	NOAA

Gulf of Mexico menhaden.

Review Workshop Panelists

Steve Cadrin	Review Panel Chair	SAFMC SSC

Churchhill Grimes	Reviewer	SAFMC SSC
Will Patterson	Reviewer	GSMFC Appointee
Gary Melvin	Reviewer	CIE
Stephen Smith	Reviewer	CIE
Kevin Stokes	Reviewer	CIE

Analytical Team

Kevin Craig	Lead analyst, SA BLT	NMFS Beaufort
Amy Scheuller	Lead analyst, GoM menhaden	NMFS Beaufort
Kyle Shertzer	Assessment Team	NMFS Beaufort
Erik Williams	Assessment Team	NMFS Beaufort
Katie Andrew	Assessment Team	NMFS Beaufort
Rob Cheshire	Assessment Team	NMFS Beaufort
Robert Leaf	Assessment Team	USM

Observers

Dewey Hemilright	Fishing Industry	Commercial, NC
Robert Johnson	Fishing Industry	Charter/Headboat

GSMFC Menhaden Advisory Committee

John Mareska,	ADCNR-MRD
Ron Lukens,	Omega Protein, Inc.
Behzad Mahmoud,I	FL FWC
Matt Hill,	MDMR
Jerry Mambretti,	TPWD
Harry Blanchet,	LDWF
Borden Wallace,	Daybrook Fisheries

Council Representative

Michelle Duval	Council Member	SAFMC

Council and Agency Staff

Julia Byrd	SEDAR Coordinator	SEDAR
Julie O'Dell	Admin.	SEDAR/SAFMC
Michael Errigo	Fishery Biologist	SAFMC Staff
Steve VanderKooy	IJF Program Coordinator	GSMFC
Jessica Stephen	Fishery Biologist	SERO
Brian Langseth	Observer	SEFSC Beaufort
Joe Smith	Observer	SEFSC Beaufort

Report for the Center of Independent Experts on the SEDAR 32 South Atlantic blueline tilefish and Gulf of Mexico menhaden assessment review (August 27 to 30, 2013)

Stephen J. Smith
383 Portland Hills Drive
Dartmouth, Nova Scotia
Canada B2W 6R4

Executive Summary

The Southeast Data, Assessment, and Review (SEDAR) 32 Review Panel met from 27 to 30 August 2013, in Morehead City, NC to review the data and assessments for Atlantic blueline tilefish and Gulf of Mexico menhaden. The panel consisted of a Chair, three CIE reviewers, and two independent reviewers. This was the first assessment for blueline tilefish in the SEDAR process while Gulf of Mexico menhaden was last reviewed at SEDAR 27 in 2011. The results of the age-based and age-aggregated models all indicate that the Atlantic blueline tilefish stock is overfished and overfishing is occurring. The results of the stock assessment were judged to be the best scientific information available, however the reliance on commercial and recreational fishery based indices of abundance limited the assessment team's ability to judge whether the recent increase in landings north of Cape Hatteras was due to a northward shift in distribution or a newly discovered but previously un-fished part of the population. In addition, the lack of a recruitment index made it impossible to verify recruitment estimates from the model that were not supported by the age compositions of the landings.

There was no evidence for menhaden of overfishing or of the stock being overfished given commonly applied benchmarks in the region and based on the results from the age-based and age-aggregated models. Managers are in the process of developing the goals and objectives for the menhaden fishery including biomass and F benchmarks for this fishery. Without established thresholds, it is not possible to provide quantitative estimates of stock status. Landings data for this fishery were of high quality and fishery-independent indices for recruitment and adults were also available for this assessment. The assessment was also of high quality and represents the best scientific information available. More fishery-independent indices may become available for future assessments once a rapid method for resolving species identification has been developed. The lack of older fish in the catch relative to their presence in the Louisiana gillnet index for adult fish was of concern with respect to estimating productivity of the stock.

Background

The review workshop of the 32nd Southeast Data, Assessment, and Review (SEDAR) process was convened in Morehead City, NC from August 27 to 30, 2013. The purpose of the workshop was to review stock assessments for Atlantic blueline tilefish and Gulf of Mexico menhaden. The South Atlantic blueline tilefish stock is within the jurisdiction of the South Atlantic Fisheries Management Council and the state waters of North Carolina, South Carolina, Georgia, and Florida. The Gulf of Mexico menhaden stock is within the jurisdiction of the Gulf States Marine Fisheries Commission and the state waters of Texas, Louisiana, Mississippi, Alabama, and Florida.

The SEDAR 32 Assessment Process was held via a series of webinars from April through July 2013. The pre-assessment webinar was held April 17, 2013. Specific assessment webinar dates were May 8, May 23, June 5, June 19, July 10, and July 24, 2013.

Blueline tilefish had not been assessed in the SEDAR process prior to this assessment while Gulf of Mexico menhaden was last assessed in 2011 at SEDAR 27.

Description of Individual Reviewer's Role in Review Activities

Background information, meeting arrangements and other material were made available to the reviewers on July 29, 2013 either via email or through an ftp site. The menhaden stock assessment report was available as of August 6, while the blueline tilefish stock assessment document was made available on August 9. I reviewed these two main assessment documents accessing the background information from the ftp as necessary to get more detail on the data used or analyses that were carried out. On August 21, I participated in a one-hour conference call with available reviewers and assessment leads hosted by Julia Byrd (South Atlantic Fishery Management Council) and Steve Cadrin (panel chair) to go over arrangements, agenda, etc., and also to go over any questions or clarifications concerning the assessment documents. I identified some issues with the surplus production model results for blueline tilefish that were addressed later in that same week by the assessment team.

The review meeting was held August 27 to 30 at the Crystal Coast Civic Center in Morehead City, NC. The panel review chair assigned me to develop text for the review report sections on the Data term of reference (TOR) (TOR 1 for both species) and Research Recommendations TOR (TOR 6 for blueline and TOR 5 for menhaden) based on my notes and those contributed by other panelists. The other two CIE panelists were given similar assignments. The chair and the two non-CIE panelists were responsible for the compiling all of the text into the final review report.

The first day of the meeting was devoted to the presentation of the material on blueline, while the presentation on menhaden took up most of the second day. On the Thursday, the two assessment teams returned with presentations dealing with their responses to issues and questions that the panel had raised during the original presentations. The panel spent Friday morning drafting the report and reviewing the draft material as a group.

Summary of Findings

SEDAR 32 South Atlantic blueline tilefish assessment review

1. Evaluate the data used in the assessment.
   a) Are data decisions made by the Data Workshop 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 review panel focused attention on the definition of the stock area, accuracy of aging data, the decisions to include age compositions but exclude length compositions from the model, the reliability of the commercial and recreational landings data, and the choice of fishery-dependent abundance indices used in the model.

   The management area for blueline tilefish extends from Florida to Rhode Island and all landings of this species in this area were included. Genetic or tagging data are not available for this species and it was assumed that the population would exhibit a Gulf of Mexico/Atlantic split similar to many other species in this same area. However, it was also noted that previous work on the related golden tilefish indicated a split north and south of Hatteras. The fishery for blueline tilefish appears to be focused in a few smaller areas, the locations of which have changed over time and a larger percentage of the recent landings are now coming from North Carolina waters north of Cape Hatteras. This species is also known to burrow in soft bottom habitats and this fine scale structure may result in local depletion. There was some discussion about whether increased landings in the area north of Cape Hatteras represented a previously untapped area for blueline tilefish or were due to a northward change in stock area but there was no information available to decide between the two possibilities.

   Age data were obtained from sampling recent commercial fisheries landings that appeared to target a very narrow range of ages (3-5 for recreational and 5-8 for commercial fisheries). There were no age composition data for landings in the earlier part of the series when it was expected that larger/older fish should have been at a higher proportion in the population given the assumption of maximum age of 43 years. The von Bertalanffy growth curve for the recent data indicated that 98% of total growth had been completed by age 15, and therefore ages 15 and older were adopted as a plus group.

   Assumptions about the initial age composition raised issues about the current estimates of natural mortality (M) and fishing mortality (F), as well the assumption of flat-topped selectivity. Natural mortality at age was estimated using the methods of Charnov et al. (2012) which are based on estimates of K and L_∞ from Von Bertalanffy growth curves and therefore highly dependent upon the quality of the recent age data. Considerable uncertainty in age determination for blueline tilefish was documented by Harris et al. (2004). A maximum M of 0.15 and a minimum of 0.05 were used for sensitivity training based upon a CV of 54% from the Hoenig method. While scaling the mean rate over the older ages to 0.1 was reasonable given the Hoenig estimate based on maximum age, the lack of fish of age 15 years and older in the recent landings suggests that either M may be higher because the maximum age of 43 is questionable due to the uncertainty in ageing, or fishing mortality was much higher than assumed. This suggests that the higher M alternative should receive more attention in the

sensitivity analysis than the lower M, and perhaps M estimates higher than 0.15 should be considered.

Maturity-at-age was based upon estimates for golden tilefish with 50% mature at age 3 and 100% mature at age 4. While these results indicated a relatively younger maturity than may be expected for such a long-lived fish, similar results have been reported for other long-lived species in the region. However, maturity studies of golden tilefish suggest that functional maturity may occur at ages older than histological maturity because of territoriality, dominance and mate choice (Grimes et al. 1988, McBride et al. 2013). If this is also true for blueline tilefish, then the apparent truncation of age composition due to harvesting may result in a decline in the size of males that gain access to the females for spawning. It is not known what impact this decline in size may have on stock productivity.

The available age composition data representing the recent years do not appear to track year-classes even though high recruitment was estimated to have occurred prior to the period that the bulk of these data were collected. This increased recruitment was not actually observed but was estimated by the model to account for recent increases in the adult handline index and recent catches.

While the age compositions were included in fitting the model, the length compositions were removed from the analysis due to preliminary results indicating lack of fit. In light of the uncertainties associated with the ageing data, it seemed strange that the length composition data would not be better fitted by the model. However, sensitivity runs and estimated length compositions from the base run in which the length compositions were not part of the objective function demonstrated that including length composition data resulted in poorer fits to the age compositions and the abundance indices. Varying sampling coverage in time and space was one of the main reasons suggested for the lack of information in the length composition data. The review panel agreed with the assessment team, noting that the residual patterns from model runs with length compositions were not acceptable.

The landing data were considered to be reliable since 1974 and discarding for the commercial fishery was assumed to be negligible, consistent with there being no regulatory reasons for discarding (e,g., size limits). The recreational catch was sporadic and low relative to the commercial catch until 2006. There was considerable discussion about the reliability of the recreational landings estimate for 2006 to 2008, including the very high discard estimates in 2007. Most of these landings appeared to have occurred in North Carolina waters and there was a suggestion that the development of a “deep-drop” fishery may have driven the increase, with the decrease in 2011 due to the implementation of a deep water closure. A quick look at the MRIP data indicated that CVs for 2006 to 2011 decreased relative to the period before and the number of sample intercepts increased, both indicative of increased fishing activity. However, the magnitude of the landings relative to the commercial landings in those same years still seemed to be unprecedented and industry participants questioned the reliability of the recreational estimates.

The commercial and recreational headboat catch rate information were key data for both the Beaufort Assessment Model (BAM) and ASPIC models. These were the only annual abundance indices available and were developed using the standard approach of fitting delta-GLM models to filter out annual trends from other factors associated with these data. The recreational index used here represented the earlier period when the SSB was being fished down but this index actual represents very low levels of catch. There was no overlap between this index and the two commercial indices. A three-year running smooth of headboat catch rate information including

data after 1992 was presented, suggesting somewhat similar trends to the commercial indices in the later years.

While the landings data were taken from the whole area, the catch rate abundance indices were confined to data between 28° and 35° N latitude to more reflect the core stock area. As noted above, the model interpreted recent increases in catch and the handline index to be due to high recent recruitment. The validity of this assumption will be important for forecasting future productivity.

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 Beaufort Assessment Model (BAM), implemented in AD Model Builder software (Fournier et al., 2012) was used to develop a statistical age-based forward projecting assessment model of the population. In addition, two production type models were also fit to the data. An age-based production model was produced using BAM with the recruitment deviations option turned off. An age-aggregated surplus-production model implemented using the ASPIC package (Prager, 2005) was also used for comparative purposes.

The BAM base case model and rationale for modeling decisions are well described in the assessment report and were further explored during the Review Workshop. The base case run included commercial and recreational landings, age composition data and three indices of abundance (recreational head boats, commercial long line and hand line). Natural mortality varied by age and was assumed constant through time. Steepness was fixed at 0.84 based on meta-analyses (Myers et al., 2002; Shertzer and Conn, 2012). Selectivities and catchabilities were all estimated as constant for the full assessment period (1974–2011).

The model was fit to the data using appropriate methods, consistent with standard practice. Analysis included iterative reweighting using the method of Francis (2011) and exploration of a variety of data configurations and parameterizations. The modeling and decisions made to develop the base case run and the sensitivity testing were all well described in the Assessment Report and supporting working documents, and were further elaborated during the SEDAR 32 Review Workshop where additional diagnostics (likelihood components, weights, likelihood profiles) were made available. The modeling procedures adopted appeared to be robust. Landings and indices were fit using lognormal likelihoods. Age composition data were fit using robust multinomial likelihoods. The treatment of the data and the relative importance given to the various components were well explored and appeared appropriate. The model structure was adequate to capture the main patterns in the data.

The production models provided useful comparisons with the base case catch-at-age model results. The main point of difference between these models and the base case was that the production models did not estimate an increase in recruitment in the most recent years, and instead estimated a higher fishing mortality. Despite this difference, the results of the production models did suggest similar stock status to the base case in terms of MSY benchmarks. The production models were appropriately configured and implemented here, and are standard tools for stock assessment.

The Monte Carlo Bootstrapping (MCB) method was used to characterize the uncertainty around the estimates and stock status outputs from the base case model. This method simulates replications of the data using parametric bootstrapping of the landings and indices data, conditional on the distributional assumptions used in the model. The length composition replicate data were resampled from the original data. In addition, values for M and steepness were drawn from probability distribution functions representing possible ranges of likely values for these parameters. Uncertainties were presented as quantiles of the frequency distributions of the various outputs from the model fits to 3043 accepted replicate draws of the above data and parameters. Each individual model fit used the weights developed for the base case run.

There was some discussion about whether all combinations of M and steepness values based on random draws would be biologically appropriate. This is a subject that needs further study for the benefit of this and other assessments that use this technique. It was also noted that the introduction of random variation to M and steepness was essentially adding process error to what was an observation error model fitting approach. Estimates of the management quantities (MSY, B_{MSY}, F_{MSY}) in the base run were estimated using estimates based on a deterministic model structure. The impact of having a stochastic model structure with process error on estimating these management quantities has been investigated for surplus-production models by Bousquet et al. (2008) who showed that F_{MSY} from the stochastic model will be less than the deterministic estimates, estimates of MSY will be higher, and those for B_{MSY} lower. The size of the differences will be a function of the amount of stochastic error in the model. Means of management quantities from the MCB runs do not equal estimates from the base run and differed in the same direction as predicted above for the surplus-production models. While these differences may not be always apparent when comparing ratio benchmarks, for consistency sake, the MCB median estimates of the benchmarks should be used in the ratios for evaluating stock status from the MCB model results.

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 reliable? If not, are there other indicators that may be used to inform managers about stock trends and conditions?

   All of the reviewers agreed that the BAM base run provided the best representation of stock status. The model was evaluated through a series of sensitivity runs that explored a number of issues with the data, model structure and assumptions. The two production models arrived at the same stock status despite interpreting recent changes in stock size differently. The median status results from the MCB run also resulted in the same stock status.

   Spawning biomass in 2011 was estimated as 445 thousand pounds., which was less than the estimate of Minimum Stock Size Threshold (489 thousand pounds), so the stock is overfished.

SSB has been below SSB_{MSY} for the past two years (2010–2011). The majority of viable sensitivities runs indicate that the SSB2011 was < SSB_{MSY}. The only exception was the increase in M run which indicated the SSB was greater than the SSB_{MSY}. This was considered unlikely based on additional sensitivity runs requested by the Review Panel. Production model outputs of population status generally agree with the catch-at-age model and indicate a B/B_{MSY} of less than 1 in 2011.

Based on the BAM base run fishing mortality (F) estimates, overfishing is occurring for the South Atlantic Blueline tilefish. The ratio of the geometric mean F over the past 3 years to F_{MSY} was greater (2.37) than 1.0 and has been for the past several years. The dramatic decrease in F_{2011} was primarily the result of a fishery closure. Production model outputs all indicate an average F/F_{MSY} well in excess of 1.0.

The stock/recruitment relationship does not appear to be very informative. There is no information on steepness in the data and there are large positive deviations in the early to mid-2000s to accommodate the increased catches and handline index estimates in the mid to late 2000s. In the terminal three years of the assessment, estimated recruitment did not deviate from the spawner-recruit curve. The recruitment used for the projections was taken from the curve and represents the mean recruitment, not including the high years in the early to mid-2000s. This approach was considered to be reasonable given the data.

The quantitative estimates of stock status appear to be reliable given the agreement on stock status amongst the different models used and the results of the sensitivity runs.

4. Evaluate the stock projections, addressing the following:
   a) Are the methods consistent with accepted practices and available data?
   b) Are the methods appropriate for the assessment model and outputs?
   c) Are the results informative and robust, and useful to support inferences of probable future conditions?
   d) Are key uncertainties acknowledged, discussed, and reflected in the projection results?

   The projection methods used in this stock assessment were consistent with accepted practices in the region and elsewhere, and the available data. Initially the review panel had several concerns regarding the use of MCB approach as a measure of uncertainty. The MCB analysis is considered an approximation of uncertainty for the base run. A number of the limitations were identified in the assessment report. In addition, there was the point raised above about using the median estimates of the management quantities rather than those from the base run to evaluate stock status from the MCB results.

   Projection results were informative and robust within the range of observations and inputs from the MCB. Currently F is estimated as the geometric mean of the three previous years. Given the observed rapid changes in F and the preliminary landings estimates for 2012 and 2013, consideration might be given to using actual landings for future projections or drop the 2011 from the estimate of F for 2013 and 2014.

5. 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 modeling using extensive sensitivity runs and likelihood profiling, retrospective analyses and MCB. All of the methods used are standard stock assessment methods. Issues considered in sensitivity runs include variations in M and steepness, alternative maturity vector, adjustment of model weights and exclusion of each series of indices, allowing catchability to vary, inclusion of ageing error, and allowing recreational selectivity to be dome shaped. The sensitivity runs of the base case explored variants of the current model structure but cannot include the impact of other processes such as environmental or geographic effects that are not part of the current structure. However, very useful information was presented on the various sensitivity runs and the panel was satisfied that there had been sufficient exploration of the assessment uncertainties.

6. Consider the research recommendations provided by the Data and Assessment workshops 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.

Research recommendations for blueline tilefish were provided in the data and assessment working group documents. The Panel noted that many of these recommendations reflected concerns across a range of deep-water species and therefore confined their attention to those specific to the stock assessment of blueline tilefish.

While the panel supports work on stock structure, we recommend starting with the available information on describing the differences in demographics/life history characteristics over the range of the management area. Additionally, the available information on habitat in the areas listed should be evaluated before initiating any new studies.

Given that this is an age-based assessment, the comparison and calibration studies for the age determination should receive high priority, along with the marginal increment analysis to determine if the opaque zone is formed annually. Many species would probably benefit from expanding the MRIP program to include age sampling.

The collection of information to better describe spawning season and spawning periodicity could probably start with fishery-dependent sources but will need data from fishery-independent programs to cover the range of the species. The latter program would probably have to be tailored to provide samples across the deep-water snapper/grouper complex.

Studies of discard mortality should be low priority given the current negligible discard rate in the commercial fishery. The collection of additional information on catch (e.g., lengths, ageing material) is important, especially for the areas north of Hatteras, but would likely require an observer program developed for all fisheries focusing on the deep-water snapper/grouper complex.

The BAM model is reliant on historical information, and any data on size compositions, maximum size, etc., obtained from historical recreational fishing photos could be quite useful. One of the main issues raised about the recreational fishery concerned the high landings in the mid-late 2000s, especially the high landing and discard estimates for 2007. Closer scrutiny of these estimates requires data at higher resolution than was apparently available for this stock assessment.

With respect to developing a fishery-independent survey, sampling of deep-water habitats may elucidate habitat characteristics, and spatial distributions of blueline tilefish and other deep-water reef fishes. If a sufficient time series is developed, then a useful fishery-independent index may be available for the stock assessment.

Overall, the material provided to the panel and the presentations made at the SEDAR 32 meeting were of excellent quality. The assessment team members were responsive to all requests made for additional work and provided complete responses to all requests. The amount of material provided for both blueline tilefish and Gulf of Mexico menhaden was extensive and a three and one half day meeting may not have been long enough to consider all of the material to the same level of detail.

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

The reliance of the assessment on fishery-dependent abundance indices and the lack of a recruitment index were identified as weaknesses of the current approach that could be improved upon. Having an area-wide fishery-independent survey could provide information on geographic changes in distribution and on validation of recruitment trends, both identified as issues with this assessment. While the size of this fishery may not by itself warrant the cost of implementing such a survey, there may be broader advantages in designing a survey for the complex of deep-water species.

SEDAR 32A Gulf of Mexico menhaden assessment review

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

a) Are data decisions made by the 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 landings were judged to be accurate as the largest portion had been due to the reduction fishery and there has been a log system in place including daily catch records since 1964. Cooperation by industry with supplying information to NMFS is impressive (weekly electronic reporting, 100% participation in the voluntary program, access for port sampling and provision of freezer space for samples). The decision to start the series in 1977 was quite reasonable given the concerns about the data quality for age composition data prior to 1977, inexplicable truncated age distribution in the early 1970s and other issues with these early data as noted in

past stock assessments. Sensitivity analyses were used to evaluate the impact of including the longer time series of age compositions.

More information on the spatial distribution of the fishery was requested. The analysis presented on fishery hotspots for 2008, 2009 and 2011 was quite informative and a longer time series would provide information on the spatial overlap between the fishery and the fishery-independent indices used in the assessment. These data may provide insight into the potential distribution of older fish off of western Louisiana and to the east of Alabama/Mississippi; areas not covered by the seine or gillnet survey indices used in the assessment.

The use of population fecundity as a proxy for spawning stock biomass was discussed. A relationship produced in the early 1980s relating numbers of eggs to female length was used in this assessment to estimate length-specific fecundity in the model, thus larger, older fish are estimated to produce more eggs per capita than younger fish. This fish has a protracted spawning season and is assumed to be an indeterminate batch spawner. If older fish produce more batches or higher quality eggs, then their contribution to stock-specific fecundity would be underestimated using the current approach. Ovarian egg number could be a reliable index to SSB if all the ovary samples were at the same stage of reproductive development, but that would seem unlikely to be the case. However, accounting for the relationship between size and fecundity was recognized as a step in the right direction.

There was also discussion about the lack of older fish in the catches being due to the potential for older fish being less vulnerable to the fishery as a function of age-specific spatial distributions. The major grounds for the fishery are within 10 miles of the coast, but the resource distribution is out to 60 miles. Although the fishery may be constrained by spotter planes pilots being reluctant to go offshore, the majority of the stock was considered to be inshore during the warmer months. Based on early-season catches that are further offshore (10–20 miles), there do not appear to be older fish offshore during the fishery. However, the fact that the fishery may target more abundant schools of smaller and younger fish could provide another explanation of the lack of older fish in the catch.

Results from a 2012 study with alternative sampling protocols suggest that sampling the top of hold only did not accurately represent catch, particularly with respect to the presence of older fish in the catch at age estimated from the samples. The study had limited sample size and poor coverage of the port-week strata, and the results suggested that older fish were less than 5% of the catch in the alternative-design samples. However, the lack of older fish in the commercial catch was of concern given that older fish do appear in the gillnet survey used in the assessment.

Several issues were identified with the age data. Multiple age-readers aged fish in the 1960s–early 1970s, but only a single age-reader has aged fish since the 1970s. No formal protocol for ageing quality control appears to exist. Three informal analyses of ageing accuracy or repeatability produced questionable results (e.g., 71% agreement between otolith and scale derived age estimates; 82% agreement between age estimates from scales aged in 2005 and again in 2012; and, substantial disagreement in age estimates from the 1970s versus contemporary re-ageing of those samples). Given the short-lived nature of the fish, reader error of even one year can cause substantial bias in an age-based assessment.

An ageing error matrix was included in the BAM model but was based upon comparisons of

age estimates from scales and otoliths, with the otolith ages being assumed to be the true ages. However, there was also error in the age estimates from otoliths, perhaps just as much as in the scales of short-lived species like menhaden. The ageing error matrix also assumed that the error is symmetric about age. In most fishes older fish tend to be under aged with scales as annuli pack at the scale margin and become difficult to discern.

There was evidence of a shift in age composition in the landings from mostly age-1 in the1960–1980s to mostly age-2 in the most recent years. Several hypotheses for the shift were discussed in the assessment report (habitat alteration affecting recruitment of juvenile fish into estuaries, decreased fishing mortality, recent contractions in the spatial distribution of the fishery, changing spatial distribution of age-1 menhaden, influence of hypoxic habitats on spatial distribution) but ageing drift was ruled out based on age determinations from re-reading archived scale samples.

A number of available abundance indices were excluded from being used in the model. A juvenile trawl index, which was highly correlated with the seine index, was included in the SEDAR 27 assessment model, but dismissed here because it was judged that trawls are not efficient for pelagic fish, the spatial extent of the survey was not appropriate for the resource, and the western portion of the survey has species identification problems. A research recommendation was included in the assessment report for genetic sampling by size to solve the species identification problem. The gillnet index used in the assessment was limited to the Louisiana series. Data from the western and eastern portions of the resource area were excluded because of mixed species catches and species identification problems. A larval survey was not used because of poor winter coverage, complex recruitment dynamics from larvae to fishery recruitment, and problems with species identification. Members of the Review Panel questioned why some of these indices were excluded prior to assessing their impact on model fit, such as through likelihood profiling.

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 Beaufort Assessment Model (BAM), implemented in AD Model Builder software (Fournier et al., 2012) was used to develop a statistical age-based forward projecting assessment model of the population. In addition, an age-aggregated surplus-production model, implemented with the ASPIC package (Prager, 2005), was also used for comparative purposes. The base case model and rationale for modeling decisions were well described in the assessment report and were further explored during the Review Workshop. The base case run included commercial and recreational landings, age and length composition data and two indices of abundance, one representing recruits and the other adult fish. 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 0.75. Selectivities and catchabilities were all estimated as constant for the full assessment period (1977–2011).

The model was fit to the data using appropriate methods, consistent with standard practice. Analysis included iterative reweighting using the method of Francis (2011) and exploration of a

variety of data configurations and parameterizations. The modeling processes and decision making that resulted in a proposed base case run and sensitivity testing are well described in the Assessment Report and supporting working documents and were further elaborated during the SEDAR 32 Review Workshop where additional diagnostics (Likelihood components, weights, likelihood profiles) were made available. The modeling procedures adopted appear to be robust. Landings and discards were fit closely. Landings and indices were fit using lognormal likelihoods. Age composition data were fit using robust multinomial likelihoods. The treatment of the data and the relative importance given to the various components were well explored and appeared appropriate. The model structure was adequate to capture the main patterns in the data.

The production model provided useful comparisons with the base case catch-at-age model results. The main point of difference between this model and the base case was that the production model did not fit the higher gillnet index estimates in 2008, 2009 and 2011. The higher recruitment deviations estimated by the BAM model to support those years could not be accommodated in the production model with a constant intrinsic rate of growth over the time period. Despite this difference, the results of the production models did suggest similar stock status to the base case in terms of MSY benchmarks. The production model was appropriately configured and implemented here, and is a standard tool for stock assessment.

The Monte Carlo Bootstrapping (MCB) method was used to characterize the uncertainty around the estimates and stock status outputs from the base case model. This method simulates replications of the data using parametric bootstrapping of the landings and indices data conditional on the distributional assumptions used in the model. Replications of the length composition data were resampled from the original data. In addition, values for M and steepness were drawn from probability distributions functions representing possible ranges of likely values for these parameters. Uncertainties were presented as quantiles of the frequency distributions of the various outputs from the model fits to 4068 accepted replicate draws of the data and parameters. Each individual model fit used the weights developed for the base case run.

There was some discussion about whether all combinations of M and steepness values based on random draws would be biologically appropriate. This is a subject that needs further study for the benefit of this and other assessments that use this technique. It was also noted that the introduction of random variation to M and steepness was essentially adding process error to what was an observation error model fitting approach. Estimates of the management quantities (MSY, B_{MSY}, F_{MSY}) in the base run were estimated using estimates based on a deterministic model structure. The impact of having a stochastic model structure with process error on estimating these management quantities has been investigated for surplus-production models by Bousquet et al. (2008) who showed that F_{MSY} from the stochastic model will be less than the deterministic estimates, estimates of MSY will be higher, and those for B_{MSY} lower. The size of the differences will be a function of the amount of stochastic error in the model. Means of management quantities from the MCB runs do not equal estimates from the base run and differed in the same direction as predicted above for the surplus-production models. While these differences may not be always apparent when comparing ratio benchmarks, for consistency sake, the MCB median estimates of the benchmarks should be used in the ratios for evaluating stock status from the MCB model results.

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 reliable? If not, are there other indicators that may be used to inform managers about stock trends and conditions?

The review panel examined the consistency of the input data and population biological characteristics with the abundance estimates, exploitation, and biomass estimates. Panelists agreed that the base BAM parameterization chosen by the assessment team provided the best representation of stock status and also felt the usage of MCB for projection estimates was appropriate.

Fishery landings were dominated by age-2 fish with fishing occurring after this age group has spawned at least once. However, the selectivity pattern for the reduction fishery was flat topped, and there is uncertainty about the presence of older fish (age-3 and older) in the reduction fishery landings given that they have been observed in fishery-independent gillnet catches.

Very high F estimates were estimated during time series considered, especially during the 1980s. Fishing mortality has subsequently declined to range between 1.0 and 3.5 y^{-1}. The 2011 full F was 2.36 y^{-1}, with much lower F estimates for the older ages.

Currently there are no formal benchmarks established for Gulf menhaden to evaluate stock status. The assessment team presented a suite of potential options for the Review Panel to evaluate. Values of SSB_{2011}/SSB_{MED}, $SSB_{2011}/SSB_{30\%SPR}$, $SSB_{2011}/SSB_{35\%SPR}$, $SSB_{2011}/SSB_{40\%SPR}$ from the BAM base run exceeded 1.0. Results from the surplus production model also estimated SSB_{2011}/SSB_{MSY} to be much greater than 1.0. Therefore, it is unlikely the Gulf menhaden stock would be evaluated to be overfished given commonly applied benchmarks in the region.

F_{MSY} was undefined because all of the fish mature and spawn at least once before being harvested. The surplus production model produced results relative to estimates of MSY with no indication of exceeding the criteria typically used to evaluate overfishing. The review panel agrees with the assessment that it is unlikely the Gulf menhaden stock is experiencing overfishing given commonly applied benchmarks in the region.

Managers are currently defining the goals and objectives for the Gulf menhaden fishery, as well as establishing biomass and F benchmarks. Without established thresholds, it is not possible to provide quantitative estimates of stock status.

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 modeling using extensive sensitivity runs and likelihood profiling, retrospective analyses and MCB. All of the methods used are standard stock assessment methods. Issues considered in sensitivity runs included scaling and the form of M, S-R steepness and form, adjustment of model weights and exclusion of each series of indices, alternative selectivity assumptions for the commercial reduction fishery, change in the start year, alternative weightings and alternative growth specification. The sensitivity runs of the base case explored variants of the current model structure but cannot include the impact of other processes such as environmental or geographic effects that are not part of the current model structure. However, very useful information was presented on the various sensitivity runs and the panel was satisfied that there had been sufficient exploration of the assessment uncertainties.

5. Consider the research recommendations provided by the 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 panel provided the following comments on the research recommendations that given in the assessment documents.

Several issues were identified with ageing for menhaden including the lack of formal protocols for inter-reader comparisons and calibration/reference data sets. Given the short-lived nature of the fish, reader error of even one year can cause substantial bias in an age-based assessment. Given the pending retirement of the single ager, assessment of the accuracy of ageing and the establishment of formal protocols should be done as soon as possible.

It was not apparent to the panel that stock structure was an issue in the stock assessment and the panel did not see value in undertaking genetic studies on stock structure. However, the panel did see considerable benefit in using simpler genetic techniques such as DNA barcoding to aid species identification, which is currently problematic in peripheral range areas as sampled in the Texas, Alabama, and Florida surveys. Resolution of species identification and any other measures to ensure more consistency across the many state surveys that were excluded from the assessment could provide a more representative basis for monitoring abundance.

The recommendation to consider an aerial survey should be pursued, although the turbid waters close to the Mississippi may limit detectability of fish schools. This kind of survey offers an opportunity to form a partnership between the states, federal government and the fishing industry in a monitoring program to ensure sustainability.

The panel recommended that addressing the sampling of the catch throughout the holds of the reduction fishery vessels be rated as very high priority given concerns about the selectivity of larger fish to the catch. The 2012 study indicated that sampling only the top of the hold may underestimate the proportion of older fish in the catch and given the use of fecundity for spawning stock biomass result in an underestimate of productivity (see below).

While the studies proposed to update knowledge about the reproductive biology of Gulf menhaden would be nice to do, the panel felt that the current approach is adequate for now and more priority should be given to resolving the selectivity pattern of older fish to the fishery so that their reproductive contribution to the population can be better accounted for.

Overall, the material provided to the panel and the presentations made at the SEDAR 32 meeting were of excellent quality. The assessment team members were responsive to all requests made for additional work and provided complete responses to all requests. The amount of material provided for both blueline tilefish and Gulf of Mexico menhaden was extensive and a three and one half day meeting may not have been long enough to consider all of the material to the same level of detail.

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

The Louisiana gillnet survey used in the menhaden assessment has a number of different mesh sizes and concern was expressed about developing a single index over these different mesh sizes, especially given the length frequencies presented in the assessment (Figure 5.44, menhaden assessment). The panel recommends evaluating the efficacy of developing separate indices by mesh size or accounting for the different mesh sizes within the same index.

The Louisiana seine survey was used as a recruitment index for the menhaden in this assessment. Starting in late 2010, the state has reduced the sampling for this survey to a core set of stations on a quarterly basis due to budgetary reasons and to accommodate other priorities. Given the importance of this survey index to the assessment, the panel recommended that the survey return to the former sampling frequency and geographic coverage.

Conclusions and Recommendations

The results of the age-based and age-aggregated models all indicate that the Atlantic blueline tilefish stock is overfished and overfishing is occurring. The assessment and data were judged to be the best scientific data available for the evaluation of stock status. The stock assessment is completely reliant on commercial and recreational indices for abundance and as a result does not have any information on the stock in areas that are not being fished. The recent increase in landings in the areas north of Cape Hatteras are a case in point where it is uncertain whether this area contains a newly discovered biomass of blueline tilefish that had not been previously exploited or if there has been a general northward movement in the stock due to changing climate conditions. The lack of a recruitment index makes it difficult to verify if the increased recruitment in the mid-2000s estimated by the model was real, especially given the lack of evidence for this recruitment in the age compositions of the landings. Studies on reproductive biology including information on spawning season and spawning periodicity

was recommended. The establishment of a fishery-independent survey for the deep-water reef fish complex could in time provide useful information on habitat and distribution of blueline tile fish and help resolve questions about year-class strength.

According to the results of age-based and age-aggregated models the stock status of Gulf of Mexico menhaden was determined to be not overfished and overfishing was not occurring given commonly applied benchmarks in the region. However, the goals and objectives for the Gulf menhaden fishery, including biomass and F benchmarks, are still being defined for this fishery. Without established thresholds, it is not possible to provide quantitative estimates of stock status. The assessment and data were judged to be the best scientific data available for the evaluation of stock status. There were a number of fishery-independent abundance indices considered for this stock assessment but all but two were rejected due to one or more issues of species identification, spatial coverage or seasonal coverage. The lack of older fish (3+ years) in the catch was a concern given the presence of older fish in the gillnet abundance index. Sampling only the top part of the hold of fishing vessels for age and size composition was suggested as a possible reason for the lack of estimates of older fish in the catch. The lack of older fish could also be due to the fishery targeting on the more abundant schools of one and two year olds. Resolving the species identification issue could result in more fishery-independent indices being used in the stock assessment model.

References

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 254:65–75,

Charnov. E.L., Gislason, H. and Pope, J.G. 2012. Evolutionary assembly rules for fish life histories. Fish and Fisheries. 14: 213–224.

Fournier, D. A., Skaug, H. J., Ancheta, J, Ianelli, J., Magnusson, A., Maunder, M.N., Nielsen, A., and Sibert, J. 2012. AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. Optimization Methods and Software 27:233–249.

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

Grimes, C. B., Idelberger, C. F., Able, K. W., and Turner, S. C. 1988. The reproductive biology of tilefish, *Lopholatilus chamaeleonticeps* Goode and Bean, from the United States Mid-Atlantic Bight, and the effects of fishing on the breeding system. Fisheries Bulletin. 86:745–762.

Harris, P.J., Wyanski, D.M., and Powers Mikell, P.T. 2004. Age, growth and reproduction of blueline tilefish along the southeastern coast of the United States, 1982-1999. Transactions of the American Fisheries Society. 133: 1190–1204.

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

McBride, R.S., Vidal, T.E., and Cadrin, S.X. 2013. Changes in size and age at maturity of the northern stock of Tilefish (*Lopholatilus chamaeleonticeps*) after a period of overfishing. Fisheries Bulletin. 111:161–174.

Myers, R., Barrowman, N., Hilborn, R., and Kehler, D. 2002. Inferring Bayesian priors with limited direct data: applications to risk analysis. North American Journal of Fisheries Management. 22: 351–364.

Prager, M. H., 2005. User's Manual for ASPIC: A Stock-Production Model Incorporating Covariates (ver. 5) And Auxiliary Programs. National Marine Fishery Service, Beaufort Laboratory Document BL-2004-01, Beaufort, NC.

Shertzer, K.W., and Conn, P.B. 2012. Spawner–recruit relationships of demersal marine fishes: prior distribution of steepness. Bulletin of Marine Science. 88: 39–50.

Appendix 1: Bibliography of materials provided for the review

Background documents available on both species available at the ftp site for this review were given listed in the following documents.

SEDAR 32 South Atlantic Gray Triggerfish and Blueline Tilefish Workshop Document (ListS32_BLT_DocumentList_6.4.2013.pdf)

SEDAR 32A Gulf of Mexico Menhaden Document List (S32A_MenhadenDocumentList_8.6.2013.pdf)

Main documents provided for review:

SEDAR 32. South Atlantic Blueline Tilefish. SECTION II: Data Workshop Report. April 18, 2013. 146 p.

SEDAR 32. South Atlantic Blueline Tilefish. SECTION III: Assessment Workshop Report. August 9, 2013 (Revised August 22, 2013). 140 p.

SEDAR 32A Gulf of Mexico Menhaden. Draft Assessment Report. August 2013. A. Schueller, J. Smith, and S. VanderKooy (eds) xxvii + 372 p.

Appendix 2: CIE Statement of Work

External Independent Peer Review by the Center for Independent Experts

SEDAR 32 South Atlantic blueline tilefish and Gulf of Mexico menhaden assessment review

BACKGROUND

Scope of Work and CIE Process: The National Marine Fisheries Service's (NMFS) Office of Science and Technology coordinates and manages a contract providing external expertise through the Center for Independent Experts (CIE) to conduct independent peer reviews of NMFS scientific projects. The Statement of Work (SoW) described herein was established by the NMFS Project Contact and Contracting Officer's Representative (COR), and reviewed by CIE for compliance with their policy for providing independent expertise that can provide impartial and independent peer review without conflicts of interest. CIE reviewers are selected by the CIE Steering Committee and CIE Coordination Team to conduct the independent peer review of NMFS science in compliance the predetermined Terms of Reference (ToRs) of the peer review. Each CIE reviewer is contracted to deliver an independent peer review report to be approved by the CIE Steering Committee and the report is to be formatted with content requirements as specified in **Annex 1**. This SoW describes the work tasks and deliverables of the CIE reviewer for conducting an independent peer review of the following NMFS project. Further information on the CIE process can be obtained from www.ciereviews.org.

SCOPE

Project Description SEDAR 32 will be a compilation of data, an assessment of the stock, and an assessment review conducted for South Atlantic blueline tilefish and Gulf of Mexico menhaden. The CIE peer review is ultimately responsible for ensuring that the best possible assessment has been provided through the SEDAR process. The South Atlantic blueline tilefish stock is within the jurisdiction of the South Atlantic Fisheries Management Council and the state waters of North Carolina, South Carolina, Georgia, and Florida. The Gulf of Mexico menhaden stock is within the jurisdiction of the Gulf States Marine Fisheries Commission and the state waters of Texas, Louisiana, Mississippi, Alabama, and Florida. . The Terms of Reference (ToRs) of the peer review are attached in **Annex 2a and 2b**.

OBJECTIVES

Requirements for CIE Reviewers: Three CIE reviewers shall have the necessary qualifications to complete an impartial and independent peer review in accordance with the tasks and ToRs described in the SoW herein. The CIE reviewers shall have expertise in stock assessment, statistics, fisheries science, and marine biology sufficient to complete the tasks of the scientific peer-review described herein. Each CIE reviewer's duties shall not exceed a maximum of 14 days to complete all work tasks of the peer review described herein.

Location of Peer Review: Each CIE reviewer shall participate during a panel review meeting to conduct the independent peer review in Morehead City, North Carolina, from 27-30 August 2013.

Statement of Tasks: Each CIE reviewer shall complete the following tasks in accordance with the SoW and Schedule of Milestones and Deliverables herein.

Tasks prior to the meeting: The contractor shall independently select qualified reviewers that do not have conflicts of interest to conduct an independent scientific peer review in accordance with the tasks and ToRs within the SoW. Upon completion of the independent reviewer selection by the contractor's technical team, the contractor shall provide the reviewer information (full name, title, affiliation, country, address, email, and FAX number) to the contractor officer's representative (COR), who will forward this information to the NMFS Project Contact no later than the date specified in the Schedule of Milestones and Deliverables. The contractor shall be responsible for providing the SoW and stock assessment ToRs to each reviewer. The NMFS Project Contact will be responsible for providing the reviewers with the background documents, reports, foreign national security clearance, and other information concerning pertinent meeting arrangements. The NMFS Project Contact will also be responsible for providing the Chair a copy of the SoW in advance of the panel review meeting. Any changes to the SoW or ToRs must be made through the COR prior to the commencement of the peer review.

Foreign National Security Clearance: Foreign National Security Clearance will not be necessary for this review because the panel review meeting will be conducted at a non-governmental facility.

Pre-review Background Documents: Approximately two weeks before the peer review, the NMFS Project Contact will send (by electronic mail or make available at an FTP site) to the COR the necessary background information and reports (i.e., working papers) for the reviewers to conduct the peer review, and the COR will forward these to the contractor. In the case where the documents need to be mailed, the NMFS Project Contact will consult with the COR on where to send documents. The reviewers are responsible only for the pre-review documents that are delivered to the contractor in accordance to the SoW scheduled deadlines specified herein. The reviewers shall read all documents deemed as necessary in preparation for the peer review.

Tasks during the panel review meeting: Each reviewer shall conduct the independent peer review in accordance with the SoW and stock assessment ToRs, and shall not serve in any other role unless specified herein. **Modifications to the SoW and ToRs shall not be made during the peer review, and any SoW or ToRs modifications prior to the peer review shall be approved by the COR and contractor.** Each reviewer shall actively participate in a professional and respectful manner as a member of the meeting review panel, and their peer review tasks shall be focused on the stock assessment ToRs as specified herein. The NMFS Project Contact will be responsible for any facility arrangements (e.g., conference room for panel review meetings or teleconference arrangements). The NMFS Project Contact will also be responsible for ensuring that the Chair understands the contractual role of the reviewers as specified herein. The contractor can contact the COR and NMFS Project Contact to confirm any peer review arrangements, including the meeting facility arrangements.

Tasks after the panel review meeting: Each reviewer shall prepare an independent peer review report, and the report shall be formatted as described in **Annex 1**. This report should explain whether each stock assessment ToR was or was not completed successfully during the SEDAR meeting. If any existing BRP or their proxies are considered inappropriate, each independent report shall include recommendations and justification for suitable alternatives. If such alternatives cannot be identified, then the report shall indicate that the existing BRPs are the best available at this time. Additional questions and pertinent information related to the assessment review addressed during the meetings that

were not in the ToRs may be included in a separate section at the end of an independent peer review report.

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

1) Conduct necessary pre-review preparations, including the review of background material and reports provided by the NMFS Project Contact in advance of the peer review.
2) Participate during the panel review meeting in Morehead City, North Carolina, from 27-30 August 2013.
3) Conduct an independent peer review in accordance with the ToRs (**Annex 2a and 2b**).
4) No later than September 13, 2013, each CIE reviewer shall submit an independent peer review report addressed to the "Center for Independent Experts," and sent to Mr. Manoj Shivlani, CIE Lead Coordinator, via email to shivlanim@bellsouth.net, and Dr. David Sampson, CIE Regional Coordinator, via email to david.sampson@oregonstate.edu. Each CIE report shall be written using the format and content requirements specified in Annex 1, and address each ToR in **Annex 2**.

Schedule of Milestones and Deliverables: CIE shall complete the tasks and deliverables described in this SoW in accordance with the following schedule.

22 July 2013	CIE sends reviewer contact information to the COR, who then sends this to the NMFS Project Contact
12 August 2013	NMFS Project Contact sends the stock assessment report and background documents to the CIE reviewers.
27-30 August 2013	Each reviewer participates during panel review meeting and conducts an independent peer review
13 September 2013	CIE reviewers submit draft CIE independent peer review reports to the CIE Lead Coordinator and CIE Regional Coordinator
27 September 2013	CIE submits CIE independent peer review reports to the COR
4 October 2013	The COR distributes the final CIE reports to the NMFS Project Contact and regional Center Director

Modifications to the Statement of Work: This 'Time and Materials' task order may require an update or modification due to possible changes to the terms of reference or schedule of milestones resulting from the fishery management decision process of the NOAA Leadership, Fishery Management Council, and Council's SSC advisory committee. A request to modify this SoW must be approved by the Contracting Officer at least 15 working days prior to making any permanent changes. The Contracting Officer will notify the COR within 10 working days after receipt of all required information of the decision on changes. The COR can approve changes to the milestone dates, list of pre-review documents, and ToRs within the SoW as long as the role and ability of the CIE reviewers to complete the deliverable in accordance with the SoW is not adversely impacted. The SoW and ToRs shall not be changed once the peer review has begun.

Acceptance of Deliverables: Upon review and acceptance of the CIE independent peer review reports by the CIE Lead Coordinator, Regional Coordinator, and Steering Committee, these reports shall be sent to the COR for final approval as contract deliverables based on compliance with the SoW and ToRs. As specified in the Schedule of Milestones and Deliverables, the CIE shall send via e-mail the contract deliverables (CIE independent peer review reports) to the COR (William Michaels, via William.Michaels@noaa.gov).

Applicable Performance Standards: The contract is successfully completed when the COR provides final approval of the contract deliverables. The acceptance of the contract deliverables shall be based on three performance standards:
(1) The CIE report shall completed with the format and content in accordance with **Annex 1**,
(2) The CIE report shall address each ToR as specified in **Annex 2**,
(3) The CIE reports shall be delivered in a timely manner as specified in the schedule of milestones and deliverables.

Distribution of Approved Deliverables: Upon acceptance by the COR, the CIE Lead Coordinator shall send via e-mail the final CIE reports in *.PDF format to the COR. The COR will distribute the CIE reports to the NMFS Project Contact and Center Director.

Support Personnel:

William Michaels, Program Manager, COR
NMFS Office of Science and Technology
1315 East West Hwy, SSMC3, F/ST4, Silver Spring, MD 20910
William.Michaels@noaa.gov Phone: 301-427-8155

Manoj Shivlani, CIE Lead Coordinator
Northern Taiga Ventures, Inc.
10600 SW 131st Court, Miami, FL 33186
shivlanim@bellsouth.net Phone: 305-383-4229

Roger W. Peretti, Executive Vice President
Northern Taiga Ventures, Inc. (NTVI)
22375 Broderick Drive, Suite 215, Sterling, VA 20166
RPerretti@ntvifederal.com Phone: 571-223-7717

Key Personnel:

NMFS Project Contact:

Julia Byrd, SEDAR Coordinator
4055 Faber Place Drive, Suite 201
North Charleston, SC 29405
julia.byrd@safmc.net Phone: 843-571-4366

Annex 1: Format and Contents of CIE Independent Peer Review Report

1. The CIE independent report shall be prefaced with an Executive Summary providing a concise summary of the findings and recommendations, and specify whether the science reviewed is the best scientific information available.

2. The main body of the reviewer report shall consist of a Background, Description of the Individual Reviewer's Role in the Review Activities, Summary of Findings for each ToR in which the weaknesses and strengths are described, and Conclusions and Recommendations in accordance with the ToRs. The CIE independent report shall be a stand-alone document for others to understand the weaknesses and strengths of the science reviewed. The CIE independent report shall be an independent peer review of each ToRs.

3. The reviewer report shall include the following appendices:

   Appendix 1: Bibliography of materials provided for review
   Appendix 2: A copy of the CIE Statement of Work

Annex 2a: Terms of Reference for the Peer Review

SEDAR 32 South Atlantic blueline tilefish assessment review

1. Evaluate the data used in the assessment, addressing the following:
   e) Are data decisions made by the Data Workshop and Assessment Workshop sound and robust?
   f) Are data uncertainties acknowledged, reported, and within normal or expected levels?
   g) Are data applied properly within the assessment model?
   h) 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.
   α) Are methods scientifically sound and robust?
   β) Are assessment models configured properly and used consistent with standard practices?
   χ) Are the methods appropriate for the available data?
3. Evaluate the assessment findings with respect to the following:
   f) Are abundance, exploitation, and biomass estimates reliable, consistent with input data and population biological characteristics, and useful to support status inferences?
   g) Is the stock overfished? What information helps you reach this conclusion?
   h) Is the stock undergoing overfishing? What information helps you reach this conclusion?
   i) Is there an informative stock recruitment relationship? Is the stock recruitment curve reliable and useful for evaluation of productivity and future stock conditions?
   j) Are the quantitative estimates of the status determination criteria for this stock reliable? If not, are there other indicators that may be used to inform managers about stock trends and conditions?
4. Evaluate the stock projections, addressing the following:
   e) Are the methods consistent with accepted practices and available data?
   f) Are the methods appropriate for the assessment model and outputs?
   g) Are the results informative and robust, and useful to support inferences of probable future conditions?
   h) Are key uncertainties acknowledged, discussed, and reflected in the projection results?
5. Consider how uncertainties in the assessment, and their potential consequences, are addressed.
   - 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
   - Ensure that the implications of uncertainty in technical conclusions are clearly stated.
6. Consider the research recommendations provided by the Data and Assessment workshops and make any additional recommendations or prioritizations warranted.

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

7. Provide guidance on key improvements in data or modeling approaches which should be considered when scheduling the next assessment.
8. Prepare a Peer Review Summary summarizing the Panel's evaluation of the stock assessment and addressing each Term of Reference. The CIE reviewers are contracted to conduct an independent peer review, therefore the contractual responsibilities of the CIE reviewers do not include the preparation of the Peer Review Summary.
   - Each CIE reviewer may assist the Chair of the panel review meeting with contributions to the Summary Report, based on the terms of reference of the review.
   - Each CIE reviewer is not required to reach a consensus, and should provide a brief summary of the reviewer's views on the summary of findings and conclusions reached by the review panel in accordance with the ToRs.

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 2b: Terms of Reference for the Peer Review

SEDAR 32A Gulf of Mexico menhaden assessment review

1. Evaluate the data used in the assessment, addressing the following:
   e) Are data decisions made by the Assessment Workshop sound and robust?
   f) Are data uncertainties acknowledged, reported, and within normal or expected levels?
   g) Are data applied properly within the assessment model?
   h) 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.
   d) Are methods scientifically sound and robust?
   e) Are assessment models configured properly and used consistent with standard practices?
   f) Are the methods appropriate for the available data?
3. Evaluate the assessment findings with respect to the following:
   f) Are abundance, exploitation, and biomass estimates reliable, consistent with input data and population biological characteristics, and useful to support status inferences?
   g) Is the stock overfished? What information helps you reach this conclusion?
   h) Is the stock undergoing overfishing? What information helps you reach this conclusion?
   i) Is there an informative stock recruitment relationship? Is the stock recruitment curve reliable and useful for evaluation of productivity and future stock conditions?
   j) Are the quantitative estimates of the status determination criteria for this stock reliable? 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.
   - 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
   - Ensure that the implications of uncertainty in technical conclusions are clearly stated.
5. Consider the research recommendations provided by the Assessment workshop and make any additional recommendations or prioritizations warranted.
   - Clearly denote research and monitoring that could improve the reliability of, and information provided by, future assessments.
   - 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 Panel's evaluation of the stock assessment and addressing each Term of Reference. The CIE reviewers are contracted to conduct an independent peer review, therefore the contractual responsibilities of the CIE reviewers do not include the preparation of the Peer Review Summary.

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

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 32/32A South Atlantic Blueline Tilefish and Gulf of Mexico Menhaden Review Workshop

Morehead City, NC August 27-30, 2013

Tuesday

9:00 a.m. Convene

9:00a.m. – 9:30a.m. Introductions and Opening Remarks **Coordinator**

- *Agenda Review, TOR, Task Assignments*

9:30a.m. – 12:00p.m. Assessment Presentation and Discussion (BLT*) TBD

12:00p.m. – 1:30p.m. Lunch Break

1:30 p.m. - 3:30 p.m. Panel Discussion **Chair**

- *Assessment Data & Methods*
- *Identify additional analyses, sensitivities, corrections*

3:30p.m. – 3:45 p.m. Break

3:30 p.m. - 5:00 p.m. Panel Discussion **Chair**

-*Continue deliberations*

5:00p.m. – 6:00p.m. Panel Work Session **Chair**

***Tuesday Goals*:** Initial BLT* presentation completed, sensitivities and modifications identified.

Wednesday

8:30 a.m. – 12:00 p.m. Assessment Presentation and Discussion (GM) TBD**

12:00 a.m. – 1:30 p.m. Lunch Break

1:30 p.m. - 3:30 p.m. Panel Discussion **Chair**

- *Assessment Data & Methods*
- *Identify additional analyses, sensitivities, corrections*

3:30p.m. – 3:45 p.m. Break

3:30 p.m. - 5:00 p.m. Panel Discussion **Chair**

-*Continue deliberations*

5:00p.m. – 6:00p.m. Panel Work Session **Chair**

***Wednesday Goals*:** Initial GM** presentation completed, sensitivities and modifications identified.

Thursday

8:30 a.m. – 12:00 p.m. Panel Discussion **Chair**

- *Review additional analyses, sensitivities*

12:00 a.m. – 1:30 p.m. Lunch Break

1:30 p.m. – 3:30 p.m. Panel Discussion **Chair**

-*Continue deliberations*

3:30 p.m. – 3:45 p.m. Break

3:45 p.m. – 5:00 p.m. Panel Discussion **Chair**

-*Consensus recommendations and comments*

5:00 p.m. - 6:00 p.m. Panel Work Session **Chair**

Thursday Goals: Final sensitivities identified, preferred models selected, projection approaches approved, Summary report drafts begun.

Friday

8:00 a.m. – 10:30 a.m. Panel Discussion **Chair**

- *Review additional analyses, final sensitivities*
- *Projections reviewed.*

10:30 a.m. – 10:45 p.m. Break

10:45 a.m. – 1:00 p.m. Panel Discussion or Work Session **Chair**

- *Review Consensus Reports*

1:00 p.m. ADJOURN

Friday Goals: Complete assessment work and discussions. Final results available. Draft Summary Report reviewed.

* BLT = South Atlantic blueline tilefish **GM = Gulf of Mexico menhaden

Report on the

Southeast Data, Assessment, and Review (SEDAR) 32
Gulf of Mexico Menhaden and South Atlantic Blueline Tilefish

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 CIE independent report shall be prefaced with an Executive Summary providing a concise summary of the findings and recommendations, and specify whether the science reviewed is the best scientific information available.

The Southeast Data, Assessment, and Review (SEDAR) 32 for Gulf of Mexico menhaden and South Atlantic blueline tilefish took place at the Crystal Coast Civic Center, Morehead City, NC from 27th to 31st August 2013. 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 Panel Report was provided according to the schedule. This report is an individual report that largely reflects the Panel Report although with some minor departures.

For blueline tilefish the RP reached agreement on what would constitute base case runs and sensitivity tests, and a basis for projections. For menhaden the RP reached agreement on what would constitute base case runs and sensitivity tests. No projections were undertaken for menhaden.

For blueline tilefish this was a first stock assessment. Standard Southeast regional assessment methods (Beaufort Assessment Model (BAM) and Monte Carlo Bootstrapping (MCB)) were applied. The data available are all fisheries-related and are limited in a variety of ways. That blueline tilefish is also a non-target species does not help matters. While I accept the assessment as the best available, I think there are two major areas of uncertainty. The first is fundamental – it is not clear that the stock can realistically be modeled using the traditional dynamic pool approach. Second, age data are sparse and poor quality and the scale of M is poorly defined. As stock status depends critically on M, this is potentially problematic. Notwithstanding, the Assessment Workshop (AW) conducted a careful and thoughtful assessment and has provided a strong basis for determining stock status. Most indications are that the stock is overfished and subject to overfishing.

For Gulf of Mexico menhaden, landings data are excellent but there is some concern about the limited fishery-independent information and bias in catch sampling. Difficulties with, and lack of clear protocols for, ageing also create potential biases. Nevertheless, the single species model has been well explored and appears to be reliable as a basis for informing decision making once goals and objectives have been agreed by the Gulf States Marine Fisheries Commission (GSFMC). Against traditional single species standards, the fishery appears neither to be overfished nor experiencing overfishing. However, as a key low trophic level species in the Gulf ecosystem, it is unclear what goals and objectives will be set and what this might mean for the standards set and consequent status of menhaden.

BACKGROUND

*The main body of the reviewer report shall consist of a **Background**, Description of the Individual Reviewer's Role in the Review Activities, Summary of Findings for each ToR in which the weaknesses and strengths are described, and Conclusions and Recommendations in accordance with the ToRs.*

Blueline Tilefish (SA BLT)

Blueline tilefish (*Caulolatilus microps*) is a demersal species, patchily distributed from as far north as Rhode Island (RI) but in greater abundance from around the North Carolina (NC) / Virgina (VA) border, south to the Campeche Banks of Mexico. It is generally found at depths between approximately 70m and 240m with a preference for sand, mud and shell-hash bottoms in which adults form burrows and appear to move little. The lifespan of blueline tilefish is possibly of the order of 40 to 45 years, reaching maturity at 3-4 years of age and with fecundity increasing with length/age. Maximum size appears to be reached by about age 15. Eggs are broadcast and pelagic.

The management region for blueline tilefish considered in this review covers the US coast from the NC/VA border southward to the SAFMC/GMFMC boundary although the stock assessment spans the entire US Southeast coast south from RI and down to Florida (FL) east coast and the FL Keys. Linkages with blueline tilefish in US Gulf of Mexico and Mexican waters are assumed to be negligible but are unknown.

Commercial catches of blueline tilefish were negligible prior to about 1972 but grew quickly from 1980 and reached a peak of about 450 mt in 1982. Catches then fluctuated in the range 45-90 mt until 2007 before increasing to around 180 mt in 2008-2010. Recreational catches have been primarily by headboat and charters but with some private boat contribution. The pattern by state has varied through time but the majority of estimated recreational catches have been from NC. Estimates of recreational discards have been high in recent years.

Blueline tilefish has not previously been assessed. As part of SEDAR 04 in 2004 data were assembled but no assessment was conducted. As reported to SEDAR 32, some studies have suggested that increases in total mortality (Z) since the 1970s and declines in mean length may be due to increased harvest in the snapper-grouper fishery (Ross and Huntsman, 1982; Harris et al., 2004; Rudershausen et al., 2008).

Blueline tilefish are managed under the Fishery Management Plan (FMP) for the Snapper-Grouper Fishery of the South Atlantic Region, first approved and implemented in 1983. Blueline tilefish have not been managed directly under the FMP but a wide variety of FMP and regulatory amendments have affected blueline tilefish fisheries and will impact upon data interpretation.

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 Texas (TX). There is no evidence of substructure within this central area of concentration. Menhaden are filter feeders with juveniles favouring phytoplankton and adults zooplankton. Menhaden form dense schools near to the surface, particular 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, reaching maturity at two years of age and with fecundity increasing with length/age.

Commercial fisheries for menhaden were developed after WWII when companies involved in Atlantic Menhaden fisheries moved into 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 exceeded 600,000 mt. During the 1990s the number of operating companies, processing plants and vessels declined. The operational context has been stable for the past decade.

The Gulf menhaden fishery has been managed under a regional Fishery Management Plan (FMP) since 1978. The fishery was last assessed in 2007 and was then estimated to be not overfished and not subject to overfishing. Currently, there are no agreed benchmarks for Gulf menhaden and there are ongoing discussions as to goals and objectives for the stock/fishery. Gulf menhaden is considered a key, ecologically important species within the Gulf ecosystem.

REVIEW PROCESS

*The main body of the reviewer report shall consist of a Background, **Description of the Individual Reviewer's Role in the Review Activities**, Summary of Findings for each ToR in which the weaknesses and strengths are described, and Conclusions and Recommendations in accordance with the ToRs.*

ToR 8 (SA BLT) and ToR 7 (GoM M)

Prepare a Peer Review Summary summarizing the Panel's evaluation of the stock assessment and addressing each Term of Reference. The CIE reviewers are contracted to conduct an independent peer review, therefore the contractual responsibilities of the CIE reviewers do not include the preparation of the Peer Review Summary.

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

Southeast Data, Assessment, and Review (SEDAR) 32 for Gulf of Mexico menhaden and South Atlantic blueline tilefish took place at the Crystal Coast Civic Center, Morehead City, NC, from 27th to 31st August 2013.

Participants in the review are listed in Appendix 3. The SEDAR Panel comprised a SAFMC SSC appointed Chair (Cadrin), a further SSC member (Grimes), a Gulf States Marine Fisheries Council (GSMFC) appointed reviewer (Patterson) and three Center for Independent Experts (CIE) reviewers (Melvin, Smith and Stokes). The SEDAR Panel was tasked with providing separate reports for BLT and Gulf of Mexico menhaden (GM). The chair outlined the tasks for the two SEDAR Panel reports during the opening session. The chair assumed overall responsibility and asked the SSC member (Grimes) to coordinate the BLT report and the GSMFC member (Patterson) to coordinate the GM report. Contributions for text on all ToR for both assessments were split between the three CIE reviewers. 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 advance *via* a dedicated ftp server (see Appendix 1). Overall, administration of the review was sound.

The Terms of Reference (ToR) for the stock reviews are given in Appendix 2, Annexes 2a and 2b. The ToR are extensive and it is debatable whether three days (as provided in the agenda) of full sessions is sufficient for a thorough or adequate review of two stock assessments, including data inputs and emanating decision support materials. SEDAR 32 was originally intended to cover three stocks. I would strongly encourage the CIE not to contemplate reviews that cover more than two stocks and even then to ensure that sufficient time is available. Four days of full session, plus writing time, would have been preferable in this case and in general for two stocks. For difficult, contentious or critical assessments, concentration on single stocks would be advantageous. In general, however, covering two stocks does have the benefit of allowing analysts to work effectively on requests without causing downtime.

The meeting followed the general outline of the draft agenda (Appendix 2, Annex 3) but with sufficient flexibility to allow necessary responses from the two STAT. 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 the public, 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 time 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 32 resulted in informed and reasonable conclusions.

I note that the ToR for CIE reviewers as part of the SEDAR ToR (ToR 8 for BLT and ToR 7 for GM) are somewhat difficult to interpret and could perhaps benefit from a rewrite. The ToR for BLT and GM explicitly state that "…*responsibilities of the CIE reviewers do not include the preparation of the Peer Review Summary*", but also that *"Each CIE reviewer may assist the Chair of the panel review meeting with contributions to the Summary Report, based on the terms of reference of the review."* A straightforward interpretation of the ToR is that the chair, and possibly other, non-CIE, Panelists would prepare the SEDAR 32 reports for BLT and GM, although with "*contributions*" as useful from CIE members to "*assist*" the Chair. For SEDAR 32 the three CIE reviewers were publicly assigned all of the primary reporting tasks for both stocks under review. When asked to clarify the ToR the Chair stated that the Summary Report is the most important and that CIE reports were often not looked at and his preference would be for CIE reviewers to focus on the Summary Report. The SEDAR Coordinator further clarified that SEDAR cannot require CIE members to contribute to the Summary Report but strongly encouraged it, and also noted that the language used for the ToR could usefully be modified. In my view, there is a need for clarification of roles of panelists in different regions, especially where there are multiple panelists. This is not a complaint about the assignation of tasks for SEDAR 32, but I think it would be helpful for each region to clarify panel appointments and specific roles of CIE and other members.

REVIEWER'S 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, the reviewers are additionally tasked with contributing to Summary Reports for each of BLT and GM.

In addition to conduct(ing) necessary pre-review preparations, including the review of background material and reports provided by the NMFS Project Contact in advance of the peer review (SoW, ToR 1), I (Stokes) participated in all discussions (SoW ToR 2) and contributed sections on methods and uncertainty to the draft Summary Reports for BLT and GM, which were left with the Chair at close on 31st August. Due to illness, this (CIE) report (SoW ToR 3) has been provided later than the specified deadline, though with agreement of the CIE. I am grateful for that agreement. The Summary Reports were finalized and delivered by their due

date (20th September). Despite illness, I was able to provide input to the Summary Reports, including suggested edits and comments on the full drafts.

SUMMARY OF FINDINGS BY STOCK

The main body of the reviewer report shall consist of a Background, Description of the Individual Reviewer's Role in the Review Activities, ***Summary of Findings for each ToR*** *in which the weaknesses and strengths are described, and Conclusions and Recommendations in accordance with the ToRs.*

Blueline Tilefish

ToR 1 *Evaluate the data used in the assessment, addressing the following:*

a) *Are data decisions made by the Data Workshop 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?*

There are no specific data from genetics or tagging studies to allow stock definition for blueline tilefish. Consistent with the approach taken for many South Atlantic stocks for which no such data exist, the stock assessment area is defined to include all landings from Rhode Island to Florida. It was noted during the Review Workshop (RW) that many species tend to exhibit a Gulf of Mexico/Atlantic split with respect to stock structure; this is consistent with the defined southern boundary of the stock assessment area. It was further noted that work on related species has indicated a stock split at Cape Hatteras, implying that the defined stock assessment northern boundary could be too far north. However, as there have been very few landings north of Cape Hatteras this probably is not of concern for stock assessment and status determination. Of more concern from an assessment and management perspective is that catches tend to be concentrated in particular areas and that the fish are known to be relatively sedentary as adults, displaying burrowing behaviour. While this might suggest that a more refined spatial model would be useful to define status, the concern is mitigated by pelagic spawning, a long larval duration and a strong north-south flow dissipating in the mid-Atlantic. Overall, despite the paucity of information, arguments made during the RW seemed reasonably to support the overall stock area definition for assessment. However, it remains unclear if the single dynamic pool assumption underpinning the assessment is valid.

The pattern of time invariant natural mortality at age was estimated using the method of Charnov et al. (2012), which defines M at age based on life history parameters and meta analysis. The pattern of natural mortality was then scaled to provide the same fraction of fish surviving to the maximum age as for a constant M estimate using the standard method due to Hoenig. The approach seems reasonable but is subject to error due to uncertain ageing (see

below) and consequent uncertainty in estimates of K and L_∞ from Von Bertalanffy growth curves. As there is considerable uncertainty in age determination of blueline tilefish (Harris, 2004), this is an area of potential concern. I am content that the general approach taken was reasonable, as too was the use of lower and higher scalars (cf 0.05 and 0.15 against the base case scalar of 0.10) for sensitivity testing in the assessment. I am concerned, however, that there are few reported age readings near to the maximum defined age of 43, especially as the stock is only apparently lightly exploited and the plus group is set at 15 due to few fish observed at older ages and because 98% of the growth is estimated to have taken place by age 15, that there are no age compositions from the earlier catch history, and that age sampling from the more recent series is very restricted. *A priori*, I do think that this suggests greater weight should be given to the higher M assessment scenarios than to the low M one when considering sensitivity tests (below). Of course, it is possible that few fish are observed over 15 due to their being unavailable or not vulnerable to the fishery, implying a high cryptic biomass, with implications for the assumed and estimated selectivity. This possibility cannot be discounted given the sedentary and burrowing nature of the fish and the concentrated and restricted fishery areas. The issue of M is considered at other ToR (below).

In my view, these are the major issues relating to data. Issues relating to maturity at age, ageing error estimation and application in the model, all considered in the Review Panel report, are relatively minor. Similarly, I have no major issues with the landings data and abundance indices used. The landings data have a number of weaknesses and the lack of overlap between the two available abundance indices is unfortunate. But they are what they are and cannot readily be improved. They are appropriately treated in the assessment model. The one exception to this perhaps is that it would be helpful if either the recreational headboat index could be extended forward in time or the commercial indices further back to create a period of overlap. Some work was done on this during the RW (using multi-year binning of the headboat data) and it appeared to hold promise.

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

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?*

It is not clear why the Beaufort Assessment Model (BAM) and Monte Carlo Bootstrapping (MCB) are used rather than an integrated Bayesian model, implemented for example using SS3. I do not see any benefits of using BAM and MCB in this way and there is a risk that status

estimates and the portrayal of uncertainty may be incompatible, or that status estimates and projections may be incompatible. I think it would be useful to see a clear rationale set out for the use of BAM and MCB rather than adopting a more integrated, Bayesian approach. Notwithstanding the above comment, considering ToR 2 (a-c), I am confident that the methods used are scientifically sound and robust, models are properly configured and used consistent with standard practice in the region, but taking account of wider and recent experience (e.g. Francis, 2012), and that the methods are appropriate given the available data.

The 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. Use of the BAM permitted the inclusion of all available types of data, including total annual removals from commercial and recreational fleets (landings and discards), age and length compositions, and indices of biomass abundance, with appropriate error distributions and use of priors on parameters. Decisions on *a priori* data inclusion and exclusion are considered at ToR 1 and I am generally comfortable with the approaches taken by the Assessment Workshop (AW). The specified assessment model used standard approaches to predicting landings, modelling growth and recruitment, and the BAM allowed an exploration of catchability and selectivity options.

The base case model and rationale for modelling decisions are well described in the AW report (section 3) and were further explored during the Review Workshop. The base case run included commercial and recreational landings, age composition data and three indices of abundance (recreational head boats, commercial long line and hand line). As noted above, it is not ideal that the recreational and commercial abundance indices do not overlap, but this was explored during the RW and the general patterns do seem to be consistent. Length compositions were excluded by the AW due to concerns about inconsistent sampling and conflicts in fitting. The AW concluded that length composition data help to inform selectivity estimates but conflict with information in abundance indices, do not track year classes well, and add unnecessary noise. The RW panel was concerned at this exclusion and the issue was explored further during the RW by looking at shadow fits comparing the base case predicted (but not fit) length compositions with the data and by examining model fits to the length composition data. The RW concluded that the residual patterns in indices were not acceptable from the model that included length compositions, and the results could not be considered as a viable base case (or sensitivity run); the decision by the AW to exclude length composition data was therefore upheld. I agree with this conclusion. Natural mortality was assumed constant through time but age-specific and scaled consistent with maximum observed age (see ToR 1). Steepness was fixed at 0.84 based on meta-analyses (Myers et al., 2002; Shertzer and Conn, 2012). Selectivities and catchabilities were all estimated as constant for the full assessment period (1974-2011).

The model was fit to the data using appropriate methods, consistent with standard practice. Analysis included iterative reweighting using the method of Francis (2011) 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 AW Report and AW WDs and were further elaborated during the SEDAR 32 Review Workshop where additional diagnostics (Likelihood components, weights, likelihood profiles) were made available. The modelling procedures adopted appear to be robust. Landings and discards were fit closely, and age composition data and abundance indices were fit to the degree that they are compatible and as indicated using the reweighting procedures. Landings and indices were fit using lognormal likelihoods. Age composition data were fit using robust multinomial likelihoods. The treatment of the data and the relative importance given to the various components were well explored by the AW and at the RW and appear appropriate. The model structure is adequate to capture the main patterns in the data.

Overall, I think the AW has done a thorough and careful job and has made a good attempt at fitting less than ideal data. The AW Report does not include likelihood profiles on M, but these were provided to the RW meeting and are informative. Overall, the model prefers a higher M. This is driven, however, not by the data per se but more by model assumptions. In detail, the only data that can and do influence M estimation are the very limited age data. These data "want" M to be low but the likelihood range as M varies is not great. In contrast, the stock-recruit likelihood varies much more, "wants" M to be high, and dominates the total, penalized likelihood. I interpret this to mean i) there is not actually much information on M in the data and ii) there is an important sensitivity to M that does not get fully elaborated. Given that BRPs also depend on the stock-recruit assumptions and fit, the choice of M is critical in defining status.

In addition to the catch-at-age primary assessment, two biomass dynamics stock assessments were carried out using the ASPIC software, one fully age-aggregated and the other age structured. The biomass dynamics models were considered as confirmatory rather than alternative analyses, because the catch-at-age model makes fuller use of composition data and represents a more detailed investigation of population dynamics. The biomass dynamics models provide a useful comparison with the catch-at-age model results, which they broadly support, showing the similar status of the stock in relation to MSY benchmarks (ToR 3). The biomass dynamics models are well known and used methods and were appropriately configured and implemented.

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 specified pdf's. 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 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 (MSY, BMSY, FMSY) from the MCB runs do not equal estimates from the base run. As noted in the RW Report, 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 will not be apparent when looking only at ratio benchmarks.

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 reliable? If not, are there other indicators that may be used to inform managers about stock trends and conditions?*

a) All estimates are consistent with data inputs, given model structure and assumptions. Assuming a single dynamic pool and the base case M, the outputs are useful to support status determination. As noted above, it is not clear but M could be higher than assumed. During the RW it was accepted that the base case was appropriate but I consider this still to be moot. If this is so, then sensitivity tests reported by the AW (Fig. 3.41) suggest that the status determination could be overstated.
b) Based on the base case BAM and on sensitivity tests reported by the AW, the stock is likely overfished, with SSB in 2011 estimated as 91% of the MSST. Apart from 2005-2009, SSB is estimated to have been below MSST since the early 1980s, soon after substantial landings were reported and the estimated fast contraction of age structure. I think this raises a concern as to the appropriateness of the single dynamic pool assumption and the possibility that the (non-target) fishery has concentrated on limited pockets of a highly heterogenous distribution. I note that the majority of sensitivity tests confirm the base case status determination but that the only one of importance is the high M run which suggests the stock may not be overfished. However, I acknowledge that the production model results, both age-aggregated and disaggregated, support the status determination that the stock is overfished.
c) Based on the base case BAM and on sensitivity tests reported by the AW, the stock is likely being overfished. The pattern of estimated F suggests the stock has been subject to

F>>Fmsy since the beginning of the (non-target) fishery, but with large variation. As at ToR 3(b), I have concerns about the feasibility of the estimated SSB and F patterns and am concerned at the basic assumptions.

d) No, the estimated stock- recruitment assumption is not informative. The likelihood profiles on M presented during the RW show clearly that the S-R is dependent on the assumed M. Given the lack of information on M in the data, the status determination is driven substantially by assumptions about the stock-recruitment relationship and M. Better information on M is important to better definition of stock status.

e) As noted by the RW, the quantitative estimates for determination of stock status are reliable within the bounds of the uncertainties identified in the Assessment Document and the Review Panels report.

ToR 4 *Evaluate the stock projections, addressing the following:*

a) Are the methods consistent with accepted practices and available data?
b) Are the methods appropriate for the assessment model and outputs?
c) Are the results informative and robust, and useful to support inferences of probable future conditions?
d) Are key uncertainties acknowledged, discussed, and reflected in the projection results?

a) The MCB method used by the AW is an accepted practice in the region. As noted at ToR 2, I do not understand the use of BAM and MCB rather than an integrated approach and think there is a risk of inconsistency by use of the two (between status determination results and portrayed uncertainty, and as a basis for consistent projections). My main concern, however, is that the technical "art" of modeling is in the use of diagnostics to reweight and tune the base case and individual sensitivity runs. MCB is intended to provide an approximation of the uncertainty around a single run, allowing for the inclusion of error in the data (e.g. age data) and in parameters. For any MCB realization, however, bootstrapping data and drawing on parameters from input pdfs, the finely tuned weighting of the single run is retained even though it will in many realisations be inappropriate. This is acknowledged to an extent by the removal of unconverged and "unrealistic" runs (in this case CB realisations were sifted to leave 3043), but such filtering is not automated and is unclear, and all retained realisations are given equal weight in the MCB outputs and in calculating central tendencies. I would be more confident in the portrayal of uncertainty using MCB, and of using MCB for projections, if a comparison could be made against more common Bayesian approaches.

e) This is covered at ToR 2 (and in the RW Report). The bottom line is that if the MCB approach is being used for projections, the MCB estimates of the management quantities should be used for evaluating stock status to be consistent. My interpretation is that BAM might be used to investigate base cases and sensitivities, much like the use of MPD models, but full analyses should proceed using MCB to ensure consistency throughout.

Regarding this ToR, the methods are appropriate if used consistently but there is a potential inconsistency when used in combination with BAM.

f) Projection results are informative and reasonably robust. It is notable that for this fishery F is estimated to vary widely from year to year, and not always clearly related to management measures. It is not therefore clear if the use of three year averaged F is a good basis. Currently F is estimated as the mean of the three previous years. Given the observed rapid changes in F and the spatial restrictions imposed in 2011, care is needed to ensure projections are realistic. I agree with the RW conclusion that if possible it would be best to use preliminary landings estimates for 2012 and 2013 rather than model these using an assumed F.

g) Yes, key uncertainties are acknowledged, discussed, and reflected in the projection results (and see ToR 5).

ToR 5 *Consider how uncertainties in the assessment, and their potential consequences, are addressed.*

- *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. 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 of the methods used are standard and much used in the region and/or more widely. The AW reported widely on the various analyses and more materials were provided and used in discussion at the RW, notably likelihood profiles on M. The application of methods appears to be comprehensive and well 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, for example, processes such as fishery or environmentally induced geographic changes in distribution of the stock or fishery induced local depletion. Nor can they consider variations on the fundamental dynamic pool assumption. There was much discussion at the RW on these issues and on data inclusion or exclusion in indices to represent stock abundance. Ultimately, the stock assessment assumes a single dynamic pool of fish and there are insufficient data at this time to support investigating alternative hypotheses. With the exception of this potentially major structural uncertainty, the other uncertainties in the assessment and its outputs have been appropriately and comprehensively considered. I do think that the reporting of uncertainty should more fully highlight the possibility that the dynamic pool assumption may be flawed.

Issues considered in sensitivity runs include variations in M and steepness, alternative maturity vector, adjustment of model weights and exclusion of each series of indices, allowing

catchability to vary, inclusion of ageing error, and allowing recreational selectivity to be dome shaped. Issues of uncertainty not covered explicitly in sensitivity tests include the quantum of landings assigned to recreational landings and especially discards in 2007-9 (CHECK) (see ToR 1). As noted at ToR 2 and 3, I think the issue of M is problematic and am not convinced that the interpretation of sensitivity results should not put more weight on the higher M option.

The MCB is alluded to at ToR 2. A total of 3200 realisations were made using M and h values drawn from specified pdf's and with the landings, indices and age composition data bootstrapped. 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 3200 realisations). However, it should be noted that reweighting can have major implications for fitting and parameter estimation and that not all individual realisations may be feasible. The degree to which this may or may not matter is model and data specific. As all realisations are afforded equal weight in determining distributions of outputs there is in general need for care in interpreting MCB results. For blueline tilefish, the SDNRs for all sensitivity tests are surprisingly good when runs are made using the base case weights. This is encouraging. However, this is no guarantee that for specific M and h combinations drawn from the pdfs, which may be incompatible, the base case weights would in any way be appropriate. I note this is not a problem confined to MCB and is not one that will be quickly solved while model weighting/tuning remains an interactive process.

The RW was comfortable that the AW had fully explored uncertainty to the extent possible and that the characterisation of benchmark trajectories and hence stock status (ToR 3) and projections (ToR 4) are suitable for informing management decisions. I am in general agreement with this conclusion, noting the words "to the extent possible".

ToR 6 *Consider the research recommendations provided by the Data and Assessment workshops and make any additional recommendations or prioritizations warranted.*

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

The DW and AW made an impressive list of research recommendations. As noted at the RW, many of those recommendations are generally applicable to deepwater species and were not considered further. In my view, while there are a number of recommendations that might aid the blueline tilefish assessment through improved data acquisition, the most important relate to stock structure/life history and to ageing. I am always hesitant to say that specific recommendations should be prioritised as that depends on many factors beyond the specific stock assessment.

The DW and AW recommended genetic or other forms of stock identification across the Gulf of Mexico and the northwestern Atlantic. Of course such studies would be nice but they are also expensive. The issue in any case is not just whether or not there are genetic differences at different scales, but also how the stock(s) is(are) structured in space and time and throughout the life of the fish, and whether or not there are changes in relation to environmental or fishing pressures. The DW and AW also recommend habitat studies. I agree with the RW consensus that the starting place for any studies, perhaps considering a wider group of species, would be a full description and qualitative analysis of all information relating to the species – e.g., life history characteristics, known areas of occurrence, occurrences by age/size, known spawning areas (if any), habitat correlations, oceanographic considerations. The need is to get a better understanding of the potential alternative population structure characteristics to inform whether or not the single dynamic pool model is appropriate and also to aid in any possible fishery-independent survey design. Such work is also needed to understand the utility of the fishery-related abundance indices.

In addition, a key area for improvement is in ageing. The DW and AW made three recommendations related to ageing (validation, marginal increment analysis, and increased sampling from recreational fisheries). I would support all of these together with any attempts to find and read historical samples.

With regard to other DW and AW recommendations, I think all are lower priority, although it is undoubtedly the case that development of a fishery-independent abundance would be useful - in the long term. Given that blueline tilefish is part of a larger deepwater/grouper complex, it would seem sensible to consider index development in the wider context. I do not immediately see great value in the DW and AW recommendations related to statistical aspects of abundance index development. These are wider issues related to all such estimation and need to be viewed in that context.

Similarly, it appears that the various recommendations related to recreational fisheries are generic and need to be considered in a wider context.

The ToR asks for specific recommendations on how to improve the SEADR process. As stated above, I consider the process to be well organised and administered and open and transparent. As with many similar processes, I am concerned that too much is attempted in too short a time and was concerned at SEDAR 32 to hear that the original intention was to cover at least three stocks. From a reviewer perspective I think single stock reviews are the most valuable. I recognise, however, that considering two stocks allows for analysts to work on requests while review can continue. From a cost perspective that may be attractive but I am unconvinced it creates value and may in fact reduce it. I noted earlier that the terms of reference at least for CIE reviewers could usefully be clarified with respect to contributing to the Panel report and

individual reports. I understood this to be an intention of the SEDAR organisers and would encourage it.

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

A key modelling assumption is that of the simple dynamic pool. As noted above, this needs to be explored as fully as possible; in my view this should be done before any further assessment is attempted. The lack of understanding affects not just the fundamental assessment assumption but also the interpretability and potential utility of abundance indices within the single pool model. As indicated above, a comprehensive analysis would be in order and the use of qualitative approaches, as well as quantitative where possible, would be in order.

As noted at ToR 1, exploring the potential for overlapping abundance indices, most likely by extending the recreational series, could be useful. I would recommend putting more energy into this than developing a new index, although that would have longer-term benefits.

At the RW it was noted that i) commercial abundance indices were based on data only as far north as Cape Hatteras, whereas ii) the assessment includes landings as far north as Rhode Island. This was regarded as not problematic because the landings north of Cape Hatteras had been small. However, with more recent catches expanding north of Cape Hatteras, this needs to be reconsidered.

GoM Menhaden

ToR 1 *Evaluate the data used in the assessment, addressing the following:*

a) *Are data decisions made by the Data Workshop 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 seems 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 I see no cause for concern with the stock structure assumed for the assessment. Use of landings and index data are appropriate.

The methods used to estimate the pattern and scale of time invariant M are well described and seem reasonable. My only concern about M is that it may not be time invariant given the importance of menhaden with the multispecies fish complex of the Gulf. In discussion it was

noted that there is no indication of any trend in M but my concern in the multispecies context is more that there could be high inter-annual variability in M at age which would ideally be incorporated into uncertainty characterisation. Although it is not possible to import multispecies model results from another region, I wonder if it is at least possible to make any inferences about potential inter-annual variability in menhaden M at age from the MSVPA work in the mid-Atlantic.

Ageing of menhaden using scales is somewhat problematic, not helped by the use of multiple readers, but only one consistently for an extended period. While there seem to be plans for building a reference collection before that very experienced reader retires, it is of concern that there seems to be no clear reading protocol and that the current system seems to rely, as stated in the RW, on the fact that "we have Ethel". Given that informal analyses suggest relatively poor agreement between otolith and scale reading, between scales re-read in 2005 and 2012, and between historical readings and present readings, there is a clear need to sort out reading protocols and reader availability. With respect to the observed change in age proportions, there is considerable discussion in the AW report and the RW considered the issue in some detail. A wide set of drivers for real change were considered but re-reading a small sample of scales from the 1970s to compare with a contemporary sample suggested that age readings from the earlier period underestimated ages compared to re-read scales. Without this re-reading, if the assessment had used the full dataset, the model would have had to interpret the change in proportion at age through time. Because of the re-reading, the AW decided to remove the age data from the earlier period. This was appropriate although the (labour intensive) alternative of re-reading all scales and using the full dataset was not apparently considered.

I am unconvinced by the use of fecundity as a metric to be used in determining stock status. It is true that there will be greater fecundity as fish grow and age and this will increase faster than simple SSB. However, i) calculation of fecundity adds complexity and additional uncertainty in to any status determination, and ii) a standard (i.e. benchmark) can be set appropriately for any given metric (e.g., fecundity or SSB). I am aware that fecundity is used widely in the region but see no obvious advantage of using it for menhaden given uncertainty created by, for example, batch spawning. The GSFMC is currently deliberating on goals and objectives and has yet to agree standards, I would suggest concentrating on SSB unless it can be shown that there is a clear advantage of using fecundity (but see also ToR 4).

Removals data are good for Gulf menhaden due to lack of multiple fleets and sectors and long-term, high quality logbook system. The system was well reported and discussed at the RW and I see no areas of concern. It is good to see a fishery with such good fundamental data keeping to underpin assessment and management. The one area of concern is in the catch sampling for which it is clear the protocols lead to under-representation of older fish. This is potentially problematic and need to be improved but it is unclear how historic catch sample data can be

corrected. There is also some concern that the lack of older fish in the catch (as sampled) and the assumption of logistic selectivity may be incorrect if the stock distribution is age-related and the fishery does not target older fish (either because due to low spatial/temporal overlap or because schools of larger and older fish may be smaller and not targeted even if there is overlap). There was considerable discussion on this topic with useful input from industry, and it appears reasonable to assume the fishery does in fact operate in such a way that all ages of fish are available. Lack of older fish in the catches is therefore not likely related to fish availability, may be affected by the catch sampling protocols, or may (as inferred by the model) reflect very high fishing mortality rates. There is no way directly to investigate the effect of the biased catch sampling on the assessment and status determination but I am comfortable that the issues have been well explored and described by the DW and AW.

Despite the large number of potential abundance indices, only the Louisiana gillnet index was used in the assessment. All other surveys were excluded *a priori* for a variety of reasons. Of note is the exclusion of a previously used trawl survey for juveniles and the exclusion of gillnet survey data from the western and eastern peripheries of the defined stock (leaving just the Louisiana index). The trawl survey was excluded as being a poor sampling method for pelagic menhaden and because the spatial extent was not appropriate. The western (TX) and eastern (FL, AL) gillnet data were excluded due to difficulties distinguishing between species in those areas. It was suggested during the RW that exclusion of indices could have been made in the modelling process, using e.g. likelihood profiling. I am comfortable the approach taken by the DW and AW and think it is appropriate to judge the utility of indices on *a priori* considerations. I think the DW and AW did a good job in this respect. Considering all potential indices at the modelling stage would have been time consuming and likely unproductive.

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

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?*

As noted for Blueline Tilefish, it is not clear why the Beaufort Assessment Model (BAM) and Monte Carlo Bootstrapping (MCB) are used rather than an integrated Bayesian model, implemented for example using SS3. I do not see any benefits of using BAM and MCB in this way and there is a risk that status estimates and the portrayal of uncertainty may be incompatible, or that status estimates and projections may be incompatible. I think it would be useful to see a clear rationale set out for the use of BAM and MCB rather than adopting a more

integrated, Bayesian approach. Notwithstanding the above comment, considering ToR 2 (a-c), I am confident that the methods used are scientifically sound and robust, models are properly configured and used consistent with standard practice in the region, but taking account of wider and recent experience (e.g. Francis, 2012), and that the methods are appropriate given the available data.

It also needs to be noted that the assessment is for Gulf menhaden as a single stock. As management goals and objectives are still being developed, and given that menhaden is a key species in the Gulf ecosystem, it may be necessary in time to develop models that address this issue in order to provide decision support.

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. Use of the BAM permitted the inclusion of all available types of data, including total annual removals from the commercial fleets (and the very small recreational catches), age and length compositions, and indices of biomass abundance, with appropriate error distributions and use of priors on parameters. Decisions on *a priori* data inclusion and exclusion are considered at ToR 1 and I am generally comfortable with the approaches taken by the AW. The specified assessment model used standard approaches to predicting landings and modelling recruitment, and the BAM allowed an exploration of catchability and selectivity options.

The base case model and rationale for modelling decisions are well described in the AW report and were further explored during the RW. 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. 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 0.75. Selectivities and catchabilities were all estimated as constant for the full assessment period (1977-2011).

The model was fit to the data using appropriate methods, consistent with standard practice. Analysis included iterative reweighting using the method of Francis (2011) 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 AW Report, which includes information on Likelihood components, weighting, SDNRs by data component and weight, likelihood profiles, etc. Further diagnostics were made available and elaborated during the SEDAR 32 RW. The modelling procedures adopted appear to be robust. Landings were fit closely, as were age composition data, and abundance indices were fit to the degree that they are compatible and as indicated using the reweighting procedures. Landings and indices were fit using lognormal likelihoods. Age composition data were fit using robust multinomial likelihoods.

The treatment of the data and the relative importance given to the various components were well explored by the AW and at the RW and appear appropriate. The model structure is adequate to capture the main patterns in the data.

In addition to the catch-at-age primary assessment, an age-aggregated biomass dynamics stock assessment was carried out using the ASPIC software. The biomass dynamics models was considered 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 (e.g., recent high indices and catches). The biomass dynamics model provides a useful comparison with the catch-at-age model, which it broadly supports without capturing recent population changes. A number of sensitivity tests were carried out on the biomass dynamics model which demonstrated the robustness of conclusions based upon it. The biomass dynamics model used, implemented with ASPIC, is well known and used. The methods were appropriately configured and implemented.

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 specified pdf's. 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 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 (MSY, BMSY, FMSY) from the MCB runs do not equal estimates from the base run. 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. Of course, these differences will not be apparent when looking only at ratio benchmarks.

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 reliable? If not, are there other indicators that may be used to inform managers about stock trends and conditions?

a) All estimates are consistent with data inputs, given model structure and assumptions. During the RW it was accepted that the base case was appropriate; I agree with this conclusion. As noted above, however, there are some issues with ageing, catch sampling and possibly cryptic biomass. These could all affect model fitting and status determination. Also, the assessment is single species only and does not account in any way for the role of menhaden within the Gulf of Mexico ecosystem. It was noted in the primary presentation at the RW that menhaden is a "key critter". There are no clear standards for setting BRPs for key species or Low Trophic Level (LTL) species but the Marine Stewardship Council (MSC) Certification Requirements (CR) offer insight in to developing international practice for default standards or use of more complex models to define appropriate BRPs. At this stage I would say that the current stock assessment might be used to assess stock status against standard BRPs (e.g. SSBcurrent/SSB40%) or even developing LTL standards (e.g. the MSC CR default of a target at SSB75%). More complex multispecies or ecosystem models would be required to support more refined management goals and objectives.

b) There are no standards set for Gulf menhaden (see ToR 3d). The AW provided estimates of a variety of standard single species status determinants, all of which suggested the stock was not overfished. I am hesitant to say that I consider the stock not to be overfished as the judgment depends on what goals and objectives the GSFMC decides and how these are translated in to specific standards.

c) Similar to ToR 3b, the AW presented status information that suggested overfishing is not taking place when judged against a plausible set of single species standards. As above, I am hesitant to say that the stock is not being overfished because it will depend on agreement as to goals and objectives and whether or not these take account of wider ecosystem considerations. I do agree that against all potential single species standards the fish stock is not being overfished.

d) No.

e) As noted above, goals and objectives are currently being developed for Gulf menhaden. The potential status determinants presented by the AW are all reasonably reliable on a single species basis without wider ecosystem considerations. There are no multispecies models available for the Gulf, as there are for the Atlantic, and it is not possible readily to transfer multispecies model results from one region to another, in the absence of a Gulf-specific model.

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

- *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. 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 MCB. All of 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.

Issues considered in sensitivity runs include scaling and form of M, S-R steepness and form, adjustment of model weights and exclusion of each series of indices, alternative selectivity assumptions for the commercial reduction fishery, start year, inclusion/exclusion of indices, alternative weightings and alternative growth specification.

The MCB is alluded to at ToR 2. A total of 5,000 realisations were made using M and h values drawn from specified pdf's and with the landings, indices and composition data bootstrapped. A total of 4,068 realisations were used to compile the final MCB quantile plots with realisations discarded if they did not converge or showed other poor behaviour. The process for discarding realisations was not discussed in detail. 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. The degree to which this may or may not matter is model and data specific. As all realisations are afforded equal weight in determining distributions of outputs, there is in general need for care in interpreting MCB results. For menhaden, the SDNRs for all sensitivity tests are surprisingly good (except for one case) when runs are made using the base case weights. This is encouraging. However, this is no guarantee that for specific M and h combinations drawn from the pdfs, which may be incompatible, the base case weights would in any way be appropriate.

The RW was comfortable that the AW had fully 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. I agree with this conclusion.

ToR 5 *Consider the research recommendations provided by the Data and Assessment workshops and make any additional recommendations or prioritizations warranted.*

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

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

Catch sampling appears to be highly biased. The assessment model only has a few ages of menhaden represented and the sampling bias could seriously influence outcomes, including status determination. Although it is not easy to fix existing biased data, more work on clarifying potential existing biases and on improving the sampling to reduce bias should be regarded as a high priority.

The RW noted that the Louisiana seine survey was used as a recruitment index for the menhaden in this assessment. Starting in late 2010, the state has reduced the sampling for this survey to a core set of stations on a quarterly basis due to budgetary reasons and to accommodate other priorities. Given the importance of this survey index to the assessment, the panel recommended that the survey return to the former sampling frequency and geographic coverage. I regard this as a high priority issue.

Difficulties with species identification in the TX, FL and AL gillnet surveys precluded their use in abundance indices. It is clear that if the difficulties could be overcome that adult and juvenile gillnet indices could be improved with respect to stock wide representation. Cost effective methods to improve species identification, including simple genetic approaches, could be usefully developed. I note that this needs to be additional to maintaining the Louisiana index (see above).

As noted at ToR 1 and by the RW, a number of issues were identified with ageing for menhaden including the lack of formal protocols for inter-reader comparisons and calibration/reference data sets. Given the short-lived nature of the fish, reader error of even one year can cause substantial bias in an age-based assessment. Given the pending retirement of the single ager, assessment of the accuracy of ageing and the establishment of formal protocols should be done as soon as possible.

The ToR asks for specific recommendations on how to improve the SEDAR process. As stated above, I consider the process to be well organised and administered and open and transparent. As with many similar processes, I am concerned that too much is attempted in too short a time

and was concerned at SEDAR 32 to hear that the original intention was to cover at least three stocks. From a reviewer perspective I think single stock reviews are the most valuable. I recognise, however, that considering two stocks allows for analysts to work on requests while review can continue. From a cost perspective that may be attractive, but I am unconvinced it creates value and may reduce it. I noted earlier that the terms of reference at least for CIE reviewers could usefully be clarified with respect to contributing to the Panel report and individual reports. I understood this to be an intention of the SEDAR organisers and would encourage it.

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

The key issues for the single species Gulf menhaden assessment relate to signals about age compositions and selectivity. The detailed results of the Louisiana gillnet surveys (see AW Report Figure 5.44) suggest that there are larger and older fish in the fishery, at least as indicated by the catch sampling, which appears to be age-biased. The same figure also shows the wide range of net sizes used in the survey and the widely differing length distributions. It would be useful to account for the different mesh sizes in the development of a single stock index. Ideally, that index would also be developed to include data from the TX, FL and AL surveys following improved species identification. While a refined gillnet index would be useful, it is essential to improve the catch sampling to ensure that the assessment can compare unbiased population and fishery age compositions.

Perhaps more fundamentally, but depending on how the GSFMC sets goals and objectives, there may be a need to consider multispecies or ecosystem approaches to deal with the role of menhaden within the ecosystem.

GENERAL CONCLUSIONS AND RECOMMENDATIONS

Because of the wide scope (two stocks, consideration of both data collection and analysis, and stock assessment), it is highly likely that misinterpretation of some materials, presentations or discussions has been made. This is the fault of the reviewer, not of the many excellent STAT scientists who gave good presentations and made the review an enjoyable experience – to them, many thanks. Thanks also to the SEDAR organisers. The SEDAR system is well established, very well administered and transparent. I have seen it in action on a number of occasions and remain impressed.

Both assessments reviewed were carefully conducted and well reported. The Analytical Teams were clearly well on top of the many issues and provided excellent materials and presentations, as well as responding well to requests during the RW. It is clear that both teams (with much crossover) are well led, motivated and able.

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 Blueline tilefish I see two major issues. First, it is unclear if the standard dynamic pool model used is appropriate. The species is widely distributed but caught only patchily with spatial dominance of landings varying through time apparently in relation to effort distribution in other fisheries. Second, the only data available are fisheries related and are limited; it is hard conclusively to interpret these data. The generally limited data sets (restricted age data in time and across ages, uncertainty in age reading, non-overlapping and fishery dependent abundance indices) provide little information on M, which, in combination with stock-recruitment assumptions, is the key driver of stock status. Other issues are second order with regard to status determination and if any research or data priorities are made for the species, they should concentrate on these two major issues.

For Gulf of Mexico menhaden landings data are excellent but there is some concern about the limited fishery-independent information and bias in catch sampling. Difficulties with, and lack of clear protocols for, ageing also create potential biases. Nevertheless, the single species model has been well explored and appears to be reliable as a basis for informing decision-making once goals and objectives have been agreed. Against traditional single species standards, the fishery appears neither to be overfished nor experiencing overfishing. However, as a key low trophic level species in the Gulf ecosystem, it is unclear what goals and objectives will be set and what this might mean for the standards set and consequent status of menhaden.

APPENDIX 1

BIBLIOGRAPHY

Prior to the Workshop, extensive materials were provided *via* a dedicated, anonymous ftp server (ftp.safmc.net). The materials were extensive and relevant to all terms of reference in varying degrees.

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 the dedicated server, which was accessed using an open Wi-Fi connection throughout the meeting. Wi-Fi access was generally adequate.

REFERENCES

- 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 254:65–75,
- Charnov, E.L., Gislason, H, and J.G. Pope. 2013. Evolutionary assembly rules for fish life histories. Fish and Fisheries 14:213-224
- Grimes, C. B., C. F. Idelberger, K. W. Able, and S. C. Turner. 1988. The reproductive biology of tilefish, Lopholatilus chamaeleonticeps Goode and Bean, from the United States Mid-Atlantic Bight, and the effects of fishing on the breeding system. Fish. Bull. 86:745–762.
- Fournier, D. and C.P. Archibald. 1982. A General Theory for Analyzing Catch at Age Data. Canadian Journal of Fisheries and Aquatic Sciences 39(8):1195-1207.
- Fournier, D., H. J. Skaug, J. Ancheta, J. Ianelli, A. Magnusson, M. N. Maunder, A. Nielsen, and J. Sibert. 2012. AD Model Builder: Using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. Optimization Methods and Software 27(2):233-249.
- Francis, R.I.C.C. 2011. Data weighting in statistical fisheries stock assessment models. Canadian Journal of Fisheries and Aquatic Sciences 68:1124-1138.
- Harris, P.J., D.M. Wyanski, and P.T.P. Mikell. 2004. Age, growth, and reproductive biology of blueline tilefish along the southeastern coast of the United States, 1982-1999. Transactions of the American Fisheries Society 133:1190-1204.
- Lorenzen, K. 1996. The relationship between body weight and natural mortality in juvenile and adult fish: A comparison of natural ecosystems and aquaculture. Journal of Fish Biology 49:627-647.
- McBride, R.S., T.E. Vidal and S.X. Cadrin. 2013. Changes in size and age at maturity of the

northern stock of Tilefish (Lopholatilus chamaeleonticeps) after a period of overfishing. Fish. Bull. 111:161–174.

- Myers, RA, NJ Barrowman, R Hilborn, DG Kehler. 2002. Inferring Bayesian priors with limited direct data: application to risk analysis. North American Journal of Fisheries Management 22:351−364.
- Ross, J.L. and G.R. Huntsman. 1982. Age, growth and mortality of blueline tilefish from North Carolina and South Carolina. Transactions of the American Fisheries Society 111:585-592.
- Rudershausen, P.J., E.H. Williams, J.A. Buckel, J.C. Potts, and C.S. Manooch III. 2008. Comparison of reef fish catch-per-unit-effort and total mortality between the 1970s and 2005-2006 in Onslow Bay, North Carolina. Transactions of the American Fisheries Society 137:1389-1405.
- SEDAR, 2004. SEDAR 4: Stock Assessment of the Deepwater Snapper-Grouper Complex in the South Atlantic.
- Shertzer, K.W. and Conn, P.B. 2012. Spawner-Recruit Relationships of Demersal Marine Fishes: Prior Distribution of Steepness. Bulletin Marine Science 88:39-50

APPENDIX 2

Attachment A: Statement of Work for Dr. Kevin Stokes

External Independent Peer Review by the Center for Independent Experts

SEDAR 32 South Atlantic blueline tilefish and Gulf of Mexico menhaden assessment review

BACKGROUND

Scope of Work and CIE Process: The National Marine Fisheries Service's (NMFS) Office of Science and Technology coordinates and manages a contract providing external expertise through the Center for Independent Experts (CIE) to conduct independent peer reviews of NMFS scientific projects. The Statement of Work (SoW) described herein was established by the NMFS Project Contact and Contracting Officer's Representative (COR), and reviewed by CIE for compliance with their policy for providing independent expertise that can provide impartial and independent peer review without conflicts of interest. CIE reviewers are selected by the CIE Steering Committee and CIE Coordination Team to conduct the independent peer review of NMFS science in compliance the predetermined Terms of Reference (ToRs) of the peer review. Each CIE reviewer is contracted to deliver an independent peer review report to be approved by the CIE Steering Committee and the report is to be formatted with content requirements as specified in **Annex 1**. This SoW describes the work tasks and deliverables of the CIE reviewer for conducting an independent peer review of the following NMFS project. Further information on the CIE process can be obtained from www.ciereviews.org.

SCOPE

Project Description SEDAR 32 will be a compilation of data, an assessment of the stock, and an assessment review conducted for South Atlantic blueline tilefish and Gulf of Mexico menhaden. The CIE peer review is ultimately responsible for ensuring that the best possible assessment has been provided through the SEDAR process. The South Atlantic blueline tilefish stock is within the jurisdiction of the South Atlantic Fisheries Management Council and the state waters of North Carolina, South Carolina, Georgia, and Florida. The Gulf of Mexico menhaden stock is within the jurisdiction of the Gulf States Marine Fisheries Commission and the state waters of Texas, Louisiana, Mississippi, Alabama, and Florida. . The Terms of Reference (ToRs) of the peer review are attached in **Annex 2a and 2b**.

OBJECTIVES

Requirements for CIE Reviewers: Three CIE reviewers shall have the necessary qualifications to complete an impartial and independent peer review in accordance with the tasks and ToRs described in the SoW herein. The CIE reviewers shall have expertise in stock assessment, statistics, fisheries science, and marine biology sufficient to complete the tasks of the scientific peer-review described herein. Each CIE reviewer's duties shall not exceed a maximum of 14 days to complete all work tasks of the peer review described herein.

Location of Peer Review: Each CIE reviewer shall participate during a panel review meeting to conduct the independent peer review in Morehead City, North Carolina, from 27-30 August 2013.

Statement of Tasks: Each CIE reviewer shall complete the following tasks in accordance with the SoW and Schedule of Milestones and Deliverables herein.

Tasks prior to the meeting: The contractor shall independently select qualified reviewers that do not have conflicts of interest to conduct an independent scientific peer review in accordance with the tasks and ToRs within the SoW. Upon completion of the independent reviewer selection by the contractor's technical team, the contractor shall provide the reviewer information (full name, title, affiliation, country, address, email, and FAX number) to the contractor officer's representative (COR), who will forward this information to the NMFS Project Contact no later than the date specified in the Schedule of Milestones and Deliverables. The contractor shall be responsible for providing the SoW and stock assessment ToRs to each reviewer. The NMFS Project Contact will be responsible for providing the reviewers with the background documents, reports, foreign national security clearance, and other information concerning pertinent meeting arrangements. The NMFS Project Contact will also be responsible for providing the Chair a copy of the SoW in advance of the panel review meeting. Any changes to the SoW or ToRs must be made through the COR prior to the commencement of the peer review.

Foreign National Security Clearance: Foreign National Security Clearance will not be necessary for this review because the panel review meeting will be conducted at a non-governmental facility.

Pre-review Background Documents: Approximately two weeks before the peer review, the NMFS Project Contact will send (by electronic mail or make available at an FTP site) to the COR the necessary background information and reports (i.e., working papers) for the reviewers to conduct the peer review, and the COR will forward these to the contractor. In the case where the documents need to be mailed, the NMFS Project Contact will consult with the COR on where to send documents. The reviewers are responsible only for the pre-review documents that are delivered to the contractor in accordance to the SoW scheduled deadlines specified herein. The reviewers shall read all documents deemed as necessary in preparation for the peer review.

Tasks during the panel review meeting: Each reviewer shall conduct the independent peer review in accordance with the SoW and stock assessment ToRs, and shall not serve in any other role unless specified herein. **Modifications to the SoW and ToRs shall not be made during the peer review, and any SoW or ToRs modifications prior to the peer review shall be approved by the COR and contractor.** Each reviewer shall actively participate in a professional and respectful manner as a member of the meeting review panel, and their peer review tasks shall be focused on the stock assessment ToRs as specified herein. The NMFS Project Contact will be responsible for any facility arrangements (e.g., conference room for panel review meetings or teleconference arrangements). The NMFS Project Contact will also be responsible for ensuring that the Chair understands the contractual role of the reviewers as specified herein. The contractor can contact the COR and NMFS Project Contact to confirm any peer review arrangements, including the meeting facility arrangements.

Tasks after the panel review meeting: Each reviewer shall prepare an independent peer review report, and the report shall be formatted as described in **Annex 1**. This report should explain whether each stock assessment ToR was or was not completed successfully during the SEDAR meeting. If any existing BRP or their proxies are considered inappropriate, each independent report shall include recommendations and

justification for suitable alternatives. If such alternatives cannot be identified, then the report shall indicate that the existing BRPs are the best available at this time. Additional questions and pertinent information related to the assessment review addressed during the meetings that were not in the ToRs may be included in a separate section at the end of an independent peer review report.

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

1) Conduct necessary pre-review preparations, including the review of background material and reports provided by the NMFS Project Contact in advance of the peer review.
2) Participate during the panel review meeting in Morehead City, North Carolina, from 27-30 August 2013.
3) Conduct an independent peer review in accordance with the ToRs (**Annex 2a and 2b**).
4) No later than September 13, 2013, each CIE reviewer shall submit an independent peer review report addressed to the "Center for Independent Experts," and sent to Mr. Manoj Shivlani, CIE Lead Coordinator, via email to shivlanim@bellsouth.net, and Dr. David Sampson, CIE Regional Coordinator, via email to david.sampson@oregonstate.edu. Each CIE report shall be written using the format and content requirements specified in Annex 1, and address each ToR in **Annex 2**.

Schedule of Milestones and Deliverables: CIE shall complete the tasks and deliverables described in this SoW in accordance with the following schedule.

22 July 2013	CIE sends reviewer contact information to the COR, who then sends this to the NMFS Project Contact
12 August 2013	NMFS Project Contact sends the stock assessment report and background documents to the CIE reviewers.
27-30 August 2013	Each reviewer participates during panel review meeting and conducts an independent peer review
13 September 2013	CIE reviewers submit draft CIE independent peer review reports to the CIE Lead Coordinator and CIE Regional Coordinator
27 September 2013	CIE submits CIE independent peer review reports to the COR
4 October 2013	The COR distributes the final CIE reports to the NMFS Project Contact and regional Center Director

Modifications to the Statement of Work: This 'Time and Materials' task order may require an update or modification due to possible changes to the terms of reference or schedule of milestones resulting from the fishery management decision process of the NOAA Leadership, Fishery Management Council, and Council's SSC advisory committee. A request to modify this SoW must be approved by the Contracting Officer at least 15 working days prior to making any permanent changes. The Contracting Officer will notify the COR within 10 working days after receipt of all required information of the decision on

changes. The COR can approve changes to the milestone dates, list of pre-review documents, and ToRs within the SoW as long as the role and ability of the CIE reviewers to complete the deliverable in accordance with the SoW is not adversely impacted. The SoW and ToRs shall not be changed once the peer review has begun.

Acceptance of Deliverables: Upon review and acceptance of the CIE independent peer review reports by the CIE Lead Coordinator, Regional Coordinator, and Steering Committee, these reports shall be sent to the COR for final approval as contract deliverables based on compliance with the SoW and ToRs. As specified in the Schedule of Milestones and Deliverables, the CIE shall send via e-mail the contract deliverables (CIE independent peer review reports) to the COR (William Michaels, via William.Michaels@noaa.gov).

Applicable Performance Standards: The contract is successfully completed when the COR provides final approval of the contract deliverables. The acceptance of the contract deliverables shall be based on three performance standards:
(1) The CIE report shall completed with the format and content in accordance with **Annex 1**,
(2) The CIE report shall address each ToR as specified in **Annex 2**,
(3) The CIE reports shall be delivered in a timely manner as specified in the schedule of milestones and deliverables.

Distribution of Approved Deliverables: Upon acceptance by the COR, the CIE Lead Coordinator shall send via e-mail the final CIE reports in *.PDF format to the COR. The COR will distribute the CIE reports to the NMFS Project Contact and Center Director.

Support Personnel:

William Michaels, Program Manager, COR
NMFS Office of Science and Technology
1315 East West Hwy, SSMC3, F/ST4, Silver Spring, MD 20910
William.Michaels@noaa.gov Phone: 301-427-8155

Manoj Shivlani, CIE Lead Coordinator
Northern Taiga Ventures, Inc.
10600 SW 131st Court, Miami, FL 33186
shivlanim@bellsouth.net Phone: 305-383-4229

Roger W. Peretti, Executive Vice President
Northern Taiga Ventures, Inc. (NTVI)
22375 Broderick Drive, Suite 215, Sterling, VA 20166
RPerretti@ntvifederal.com Phone: 571-223-7717

Key Personnel:
NMFS Project Contact:

Julia Byrd, SEDAR Coordinator
4055 Faber Place Drive, Suite 201
North Charleston, SC 29405
julia.byrd@safmc.net Phone: 843-571-4366

Annex 1: Format and Contents of CIE Independent Peer Review Report

1. The CIE independent report shall be prefaced with an Executive Summary providing a concise summary of the findings and recommendations, and specify whether the science reviewed is the best scientific information available.

2. The main body of the reviewer report shall consist of a Background, Description of the Individual Reviewer's Role in the Review Activities, Summary of Findings for each ToR in which the weaknesses and strengths are described, and Conclusions and Recommendations in accordance with the ToRs. The CIE independent report shall be a stand-alone document for others to understand the weaknesses and strengths of the science reviewed. The CIE independent report shall be an independent peer review of each ToRs.

3. The reviewer report shall include the following appendices:

   Appendix 1: Bibliography of materials provided for review
   Appendix 2: A copy of the CIE Statement of Work

Annex 2a: Terms of Reference for the Peer Review

SEDAR 32 South Atlantic blueline tilefish assessment review

1. Evaluate the data used in the assessment, addressing the following:
   e) Are data decisions made by the Data Workshop and Assessment Workshop sound and robust?
   f) Are data uncertainties acknowledged, reported, and within normal or expected levels?
   g) Are data applied properly within the assessment model?
   h) 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.
   d) Are methods scientifically sound and robust?
   e) Are assessment models configured properly and used consistent with standard practices?
   f) Are the methods appropriate for the available data?
3. Evaluate the assessment findings with respect to the following:
   f) Are abundance, exploitation, and biomass estimates reliable, consistent with input data and population biological characteristics, and useful to support status inferences?
   g) Is the stock overfished? What information helps you reach this conclusion?
   h) Is the stock undergoing overfishing? What information helps you reach this conclusion?
   i) Is there an informative stock recruitment relationship? Is the stock recruitment curve reliable and useful for evaluation of productivity and future stock conditions?
   j) Are the quantitative estimates of the status determination criteria for this stock reliable? If not, are there other indicators that may be used to inform managers about stock trends and conditions?
4. Evaluate the stock projections, addressing the following:
   h) Are the methods consistent with accepted practices and available data?
   i) Are the methods appropriate for the assessment model and outputs?
   j) Are the results informative and robust, and useful to support inferences of probable future conditions?
   k) Are key uncertainties acknowledged, discussed, and reflected in the projection results?
5. Consider how uncertainties in the assessment, and their potential consequences, are addressed.
   - 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
   - Ensure that the implications of uncertainty in technical conclusions are clearly stated.
6. Consider the research recommendations provided by the Data and Assessment workshops and make any additional recommendations or prioritizations warranted.

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

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

8. Prepare a Peer Review Summary summarizing the Panel's evaluation of the stock assessment and addressing each Term of Reference. The CIE reviewers are contracted to conduct an independent peer review, therefore the contractual responsibilities of the CIE reviewers do not include the preparation of the Peer Review Summary.

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

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 2b: Terms of Reference for the Peer Review

SEDAR 32A Gulf of Mexico menhaden assessment review

1. Evaluate the data used in the assessment, addressing the following:
   a) Are data decisions made by the 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 reliable? 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.
   - 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
   - Ensure that the implications of uncertainty in technical conclusions are clearly stated.
5. Consider the research recommendations provided by the Assessment workshop and make any additional recommendations or prioritizations warranted.
   - Clearly denote research and monitoring that could improve the reliability of, and information provided by, future assessments.
   - 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 Panel's evaluation of the stock assessment and addressing each Term of Reference. The CIE reviewers are contracted to conduct an independent

peer review, therefore the contractual responsibilities of the CIE reviewers do not include the preparation of the Peer Review Summary.

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

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 32/32A South Atlantic Blueline Tilefish and Gulf of Mexico Menhaden Review Workshop

Morehead City, NC August 27-30, 2013

Tuesday

9:00 a.m. Convene

9:00a.m. – 9:30a.m. Introductions and Opening Remarks **Coordinator**

- Agenda Review, TOR, Task Assignments

9:30a.m. – 12:00p.m. Assessment Presentation and Discussion (BLT*) TBD

12:00p.m. – 1:30p.m. Lunch Break

1:30 p.m. - 3:30 p.m. Panel Discussion **Chair**

- Assessment Data & Methods

- Identify additional analyses, sensitivities, corrections

3:30p.m. – 3:45 p.m. Break

3:30 p.m. - 5:00 p.m. Panel Discussion **Chair**

-Continue deliberations

5:00p.m. – 6:00p.m. Panel Work Session **Chair**

***Tuesday Goals*:** Initial BLT* presentation completed, sensitivities and modifications identified.

Wednesday

8:30 a.m. – 12:00 p.m. Assessment Presentation and Discussion (GM) TBD**

12:00 a.m. – 1:30 p.m. Lunch Break

1:30 p.m. - 3:30 p.m. Panel Discussion **Chair**

- Assessment Data & Methods

- Identify additional analyses, sensitivities, corrections

3:30p.m. – 3:45 p.m. Break

3:30 p.m. - 5:00 p.m. Panel Discussion **Chair**

-Continue deliberations

5:00p.m. – 6:00p.m. Panel Work Session **Chair**

***Wednesday Goals*:** Initial GM** presentation completed, sensitivities and modifications identified.

Thursday

8:30 a.m. – 12:00 p.m. Panel Discussion **Chair**

- Review additional analyses, sensitivities

12:00 a.m. – 1:30 p.m. Lunch Break

1:30 p.m. – 3:30 p.m. Panel Discussion **Chair**

-Continue deliberations

3:30 p.m. – 3:45 p.m. Break

3:45 p.m. – 5:00 p.m. Panel Discussion **Chair**

-Consensus recommendations and comments

5:00 p.m. - 6:00 p.m. Panel Work Session **Chair**

Thursday Goals: Final sensitivities identified, preferred models selected, projection approaches approved, Summary report drafts begun.

Friday

8:00 a.m. – 10:30 a.m. Panel Discussion **Chair**

- Review additional analyses, final sensitivities

- Projections reviewed.

10:30 a.m. – 10:45 p.m. Break

10:45 a.m. – 1:00 p.m. Panel Discussion or Work Session **Chair**

- Review Consensus Reports

1:00 p.m. ADJOURN

Friday Goals: Complete assessment work and discussions. Final results available. Draft Summary Report reviewed.

* BLT = South Atlantic blueline tilefish **GM = Gulf of Mexico menhaden

APPENDIX 3
PERTINENT INFORMATION FROM THE REVIEW

1) Participants List

Review Workshop Panelists

Steve Cadrin	Review Panel Chair	SAFMC SSC
Churchill Grimes	Reviewer	SAFMC SSC
Will Patterson	Reviewer	GSMFC Appointee
Gary Melvin	Reviewer	CIE
Stephen Smith	Reviewer	CIE
Kevin Stokes	Reviewer	CIE

Analytical Team

Kevin Craig	Lead analyst, SA BLT	NMFS Beaufort
Amy Scheuller	Lead analyst, GoM menhaden	NMFS Beaufort
Kyle Shertzer	Assessment Team	NMFS Beaufort
Erik Williams	Assessment Team	NMFS Beaufort
Katie Andrew	Assessment Team	NMFS Beaufort
Rob Cheshire	Assessment Team	NMFS Beaufort
Robert Leaf	Assessment Team	USM

Observers

Dewey Hemilright	Fishing Industry	Commercial, NC
Robert Johnson	Fishing Industry	Charter/Headboat, FL

Council Representative

Michelle Duval	Council Member	SAFMC

Council and Agency Staff

Julia Byrd	SEDAR Coordinator	SEDAR
Julie O'Dell	Administration	SEDAR/SAFMC
Michael Errigo	Fishery Biologist	SAFMC Staff
Steve VanderKooy	IJF Program Coordinator	GSMFC
Jessica Stephen	Fishery Biologist	SERO
Brian Langseth	Observer	SEFSC Beaufort
Joe Smith	Observer	NOAA

GSMFC Menhaden Advisory Committee

John Mareska, ADCNR-MRD
Behzad Mahmoud, FL FWC

Jerry Mambretti, TPWD
Borden Wallace, Daybrook Fisheries
Ron Lukens, Omega Protein, Inc.
Matt Hill, MDMR
Harry Blanchet, LDWF