SEDAR

Southeast Data, Assessment, and Review

SEDAR 50

Atlantic Blueline Tilefish

SECTION V: Review Workshop Report

October 2017

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 Background Documents and Review Workshop Working Papers 5
2. Review Panel Report 12
EXECUTIVE SUMMARY 12
2.1. Statements addressing each TOR. 13
2.2. References 34

1. Introduction

1.1 Workshop Time and Place

The SEDAR 50 Review Workshop for Atlantic Blueline Tilefish was held August 29-31, 2017 in Atlantic Beach, NC.

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 appropriately 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? Do the methods follow accepted scientific practices?
   b) Are assessment models configured appropriately and applied consistent with accepted scientific practices?
   c) Are the methods appropriate for the available data?
3. Evaluate the assessment findings with respect to the following:
   a) Are population estimates (model output – e.g. abundance, exploitation, biomass) reliable, consistent with input data and population biological characteristics, and useful to support status inferences?
   b) Is the stock overfished? What information helps you reach this conclusion?
   c) Is the stock undergoing overfishing? What information helps you reach this conclusion?
   d) Is there an informative stock recruitment relationship? Is the stock recruitment curve reliable and useful for evaluation of productivity and future stock conditions?
   e) Are the quantitative estimates of the status determination criteria for this stock appropriate for management use? If not, are there other indicators that may be used to inform managers about stock trends and conditions?
4. 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 all sources of uncertainty in the population, data sources, and assessment methods
   - Are the implications of uncertainty in technical conclusions 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 suggestions on improvements in data or modeling approaches which should be considered when scheduling the next assessment.
8. Prepare a Peer Review Summary of the Panel's evaluation of the stock assessment, 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 PANEL

Scott Crosson	Review Panel Chair	SAFMC SSC
Churchill Grimes	Reviewer	SAFMC SSC
Yan Jiao	Reviewer	MAFMC SSC
Patrick Cordue	CIE Reviewer	CIE
Jamie Gibson	CIE Reviewer	CIE
Paul Medley	CIE Reviewer	CIE

ANALYTICAL REPRESENTATIVES

Nikolai Klibansky	Lead analyst	SEFSC Beaufort
Kevin Craig	Assessment Team	SEFSC Beaufort
Paul Nitschke*	Assessment team	NEFSC
Kyle Shertzer	Assessment Team	SEFSC Beaufort
Erik Williams	Assessment Team	SEFSC Beaufort

APPOINTED OBSERVERS

Skip Fuller*	For-hire	VA
Jeff Gutman*	For-hire	NJ
Rusty Hudson	Recreational/Commercial	FL / SFA
Andy Piland	For-hire	NC

COUNCIL REPRESENTATIVES

Tony DiLernia*	Council member	MAFMC
Dewey Hemilright	Council member	MAFMC
Anna Beckwith	Council member	SAFMC
Mark Brown*	Council member	SAFMC

COUNCIL AND AGENCY STAFF

Julia Byrd	Coordinator	SEDAR
Kimberly Cole	Admin	SEDAR/SAFMC
Myra Brouwer*	SAFMC lead	SAFMC
John Carmichael*	SAFMC	SAFMC
Jason Didden/Matt Seeley	MAFMC lead	MAFMC
Mike Errigo	Fishery Biologist	SAFMC
Nick Farmer*/Jeff Pulver	Fishery Biologist	SERO

*Did not attend Review Workshop

Review Workshop Attendees

Alan Bianchi
Rob Cheshire
Michelle Duval
Eric Fitzpatrick
Amy Flowers
Jennifer Potts
Mike Schmidtke
Kate Siegfried
Amy Schueller

1.4 List of Background Documents and Review Workshop Working Papers

Atlantic Blueline Tilefish document list.

Document #	Title	Authors
Documents Prepared for the Data Workshop (DW)		
SEDAR50-DW01	Brief Summary – Habitat and Developing Spatial Species Information for Blueline Tilefish in the South Atlantic Region	Pugliese 2016
SEDAR50-DW02	Summary of the 2015 blueline tilefish cooperative-with-industry data collection project	Kellison 2016

SEDAR50-DW03	A Preliminary Assessment of Reproductive Parameters for Blueline Tilefish in Atlantic Waters from Virginia to Florida ****SEE SEDAR50-DW19 FOR FINAL REPRODUCTIVE ANALYSES**	Kolmos et al. 2016
SEDAR50-DW04	Distribution of scientifically collected blueline tilefish (*Caulolatilus microps*) in the Atlantic, and associated habitat	Klibansky 2016
SEDAR50-DW05	Summary of the results of a genetic-based investigation of blueline tilefish (*Caulolatilus microps*)	McDowell 2016
SEDAR50-DW06	Preliminary Genetic Population Structure of Blueline Tilefish *Caulolatilus microps* along the East Coast of the United States	O'Donnell and Darden 2016
SEDAR50-DW07	Description of age and growth for blueline tilefish, *Caulolatilus microps*, caught north and south of Cape Hatteras, NC	Schmidtke and Jones 2016
SEDAR50-DW08	Standard Operative Procedure for Embedding and Sectioning Blueline Tilefish (*Caulolatilus microps*)	Ostrowski 2016
SEDAR50-DW09	Summary of Northeast Fisheries Science Center Blueline Tilefish Survey Data	Nitschke and Miller 2016
SEDAR50-DW10	Summary of Mid-Atlantic Commercial Blueline Tilefish Data	Nitschke and Miller 2016
SEDAR50-DW11	Distribution of blueline tilefish (Caulolatilus microps) in the U.S. EEZ from fishery-dependent and fishery-independent data collections	Farmer and Klibansky 2016
SEDAR50-DW12	**Recommendations from the SEDAR 50 (Blueline Tilefish) Stock ID Work Group Meeting**	**SEDAR 50 Stock ID Work Group 2016**
SEDAR50-DW13	Comparison of Blueline Tilefish Otolith Derived Ages: Comparing Increment Counts Derived by Readers from NMFS SEFSC-Beaufort and SCDNR Age Laboratories	Ballenger 2017
SEDAR50-DW14	TBD	TBD
~~SEDAR50-DW15~~	SEDAR 50 Public Comments – visit the following link to view public comments submitted for SEDAR 50 https://safmc.wufoo.com/reports/sedar-50-public-comments/	

SEDAR 50-DW16	**SEDAR 50 Stock Identification Joint SSC Review Webinar Consensus Statements**	**Joint SSC Sub-Panel 2016 (Includes MAFMC, SAFMC, GMFMC representatives)**
SEDAR 50-DW17	**SEDAR 50 Stock Identification – Management/Science Call Recommendations**	**Council, Science Center, and Regional Office Leadership**
SEDAR50-DW18	Blueline Tilefish Age Workshop II	Potts et al. 2016
SEDAR50-DW19	Reproductive parameters for Blueline Tilefish in Atlantic Waters from Virginia to Florida	Kolmos et al. 2017
SEDAR50-DW20	Virginia Blueline Tilefish Data Collection Summary	Cimino 2017
SEDAR50-DW21	Summary of the Blueline Tilefish meristic conversions using data from the entire US Atlantic and Gulf of Mexico	Ballew and Potts 2016
SEDAR50-DW22	SEDAR 50 Discard Mortality Ad-hoc Group Working Paper	Discard mortality ad-hoc group
SEDAR50-DW23	Estimating dispersal of blueline tilefish (*Caulolatilus microps*) eggs and larvae from drifter data	Klibansky 2017
SEDAR50-DW24	ToR #7 Ad Hoc Work Group Working Paper	ToR #7 Ad-Hoc Work Group
SEDAR50-DW25	Standardized catch rates of blueline tilefish (*Caulolatilus microps*) in the South Atlantic and Gulf of Mexico waters of the U.S. from recreational headboat logbook data	SFB-NMFS 2017
SEDAR50-DW26	Standardized catch rates of blueline tilefish (*Caulolatilus microps*) in the South Atlantic and Gulf of Mexico waters of the U.S. from commercial logbook handline data	SFB-NMFS 2017
SEDAR50-DW27	Standardized catch rates of blueline tilefish (*Caulolatilus microps*) in the South Atlantic and Gulf of Mexico waters of the U.S. from commercial logbook longline data	SFB-NMFS 2017
SEDAR50-DW28	SEDAR 50 additional management actions provided by R. Hudson	Hudson 2017

Documents Prepared for the Assessment Workshop		
SEDAR50-AW01	South Atlantic U.S. Blueline Tilefish (*Caulolatilus microps*) length composition from the recreational fisheries	SFB-NMFS 2017
SEDAR50-AW02	Commercial length composition weighting for U.S. Blueline Tilefish (*Caulolatilus microps*)	SFB-NMFS 2017
SEDAR50-AW03	Additional Commercial Fishery Statistics: Landings in Weight and Number, Mean Weights, Update to Uncertainty, and Catch and Effort Maps	SEDAR 50 Commercial WG
Documents Prepared for the Review Workshop		
SEDAR50-RW01	Information to help interpret results from the data limited toolkit for Atlantic Blueline Tilefish north and south of Cape Hatteras	Ahrens 2017
Final Assessment Reports		
SEDAR50-SAR1	Assessment of Atlantic Blueline Tilefish	To be prepared by SEDAR 50
Reference Documents		
SEDAR50-RD01	SEDAR 32 South Atlantic Blueline Tilefish Stock Assessment Report	SEDAR 32
SEDAR50-RD02	List of documents and working papers for SEDAR 32 (South Atlantic Blueline Tilefish and Gray Triggerfish) – all documents available on the SEDAR website.	SEDAR 32
SEDAR50-RD03	Managing A Marine Stock Portfolio: Stock Identification, Structure, and Management of 25 Fishery Species along the Atlantic Coast of the United States	McBride 2014
SEDAR50-RD04	Workshop to Determine Optimal Approaches for Surveying the Deep-Water Species Complex Off the Southeastern U.S. Atlantic Coast	Carmichael et al. 2015
SEDAR50-RD05	Report to Virginia Marine Resources Commission: Grant F-132-R-2 The Population Dynamics of Blueline and Golden Tilefish, Snowy and Warsaw Grouper and Wreckfish	Schmidtke et al. 2015

SEDAR50-RD06	Estimated Catch of Blueline Tilefish in the Mid-Atlantic Region: Application of the Delphi Survey Process	Allen et al. 2016
SEDAR50-RD07	MAFMC Memo: Blueline Tilefish Catch Series – Feb 23, 2016	Didden 2016
SEDAR50-RD08	Reproductive Biology of the Blueline Tilefish, Caulolatilus microps, off North Carolina and South Carolina	Ross and Merriner 1983
SEDAR50-RD09	Fish species associated with shipwreck and natural hard-bottom habitats from the middle to outer continental shelf of the Middle Atlantic Night near Norfolk Canyon	Ross et al. 2016
SEDAR50-RD10	Systematics and Biology of the Tilefishes (Perciformes: Branchiostegidae and Malacanthidae), with Descriptions of Two New Species	Dooley 1978
SEDAR50-RD11	Integrating DNA barcoding of fish eggs into ichthyoplankton monitoring programs	Lewis et al. 2015
SEDAR50-RD12	Age, growth, and reproductive biology of blueline tilefish along the southeastern coast of the United States, 1982-1999	Harris et al. 2004
SEDAR50-RD13	Description of the Circulation on the Continental Shelf	Bumpus 1973
SEDAR50-RD14	Spawning Locations for Atlantic Reef Fishes off the Southeastern U.S.	Sedberry et al. 2006
SEDAR50-RD15	Observations and a Model of the Mean Circulation over the Middle Atlantic Bight Continental Shelf	Lentz 2008
SEDAR50-RD16	Modeling larval connectivity of the Atlantic surfclams within the Middle Atlantic Bight: Model development, larval dispersal and metapopulation connectivity	Zhang et al. 2015
SEDAR50-RD17	Tilefishes of the Genus *Caulolatilus* Construct Burrows in the Sea Floor	Able et al. 1987
SEDAR50-RD18	Delineation of Tilefish, *Lopholatilus chamaeleonticeps*, Stocks Along the United States East Coast and in the Gulf of Mexico	Katz et al. 1983
SEDAR50-RD19	Chapter 22: Interdisciplinary Evaluation of Spatial Population Structure for Definition of	Cadrin et al. 2014

	Fishery Management Units (excerpt from Stock Identification Methods – Second Edition)	
SEDAR50-RD20	Overview of sampling gears and standard protocols used by the Southeast Reef Fish Survey and its partners	Smart et al. 2015
SEDAR50-RD21	Age, Growth, and Mortality of Blueline Tilefish from North Carolina and South Carolina	Ross and Huntsman 1982
SEDAR50-RD22	Radiocarbon from nuclear testing applied to age validation of black drum, *Pogonias cromis*	Campana and Jones 1998
SEDAR50-RD23	A long- lived life history for a tropical, deepwater snapper (*Pristipomoides filamentosus*): bomb radiocarbon and lead-radium dating as extensions of daily increment analyses in otoliths	Andrews et al. 2012
SEDAR50-RD24	Age and growth of bluespine unicornfish (*Naso unicornis*): a half-century life-span for a keystone browser, with a novel approach to bomb radiocarbon dating in the Hawaiian Islands	Andrews et al. 2016
SEDAR50-RD25	Age, growth and reproduction of the barrelfish *Hyperoglyphe perciformis* (Mitchill) in the western North Atlantic	Filer and Sedberry 2008
SEDAR50-RD26	Age, growth, and spawning season of red bream (Beryx decadactylus) off the southeastern United States	Friess and Sedberry 2011
SEDAR50-RD27	Great longevity of speckled hind (*Epinephelus drummondhayi*), a deep-water grouper, with novel use of postbomb radiocarbon dating in the Gulf of Mexico	Andrews et al. 2013
SEDAR50-RD28	Refined bomb radiocarbon dating of two iconic fishes of the Great Barrier Reef	Andrews et al. 2015
SEDAR50-RD29	Age validation of the North Atlantic stock of wreckfish (*Polyprion americanus*), based on bomb radiocarbon (^{14}C), and new estimates of life history parameters	Lytton et al. 2016
SEDAR50-RD30	Stock Complexes for Fisheries Management in the Gulf of Mexico	Farmer et al. 2016
SEDAR50-RD31	Modelling community structure and species co-occurrence using fishery observer data	Pulver et al. 2016
SEDAR50-RD32	Descriptions of the U.S. Gulf of Mexico Reef Fish Bottom Longline and Vertical Line Fisheries Based on Observer Data	Scott-Denton et al. 2011

SEDAR50-RD33	Natural mortality estimators for information-limited fisheries	Kenchington 2014
SEDAR50-RD34	The relationship between body weight and natural mortality in juvenile and adult fish: a comparison of natural systems and aquaculture	Lorenzen 1996
SEDAR50-RD35	Mortality Rate of Fishes in the Pelagic Ecosystem	Peterson and Wroblewski 1984
SEDAR50-RD36	A Mathematical Model of Some Aspects of Fish Growth, Respiration, and Mortality	Ursin 1967
SEDAR50-RD37	MAFMC Memo: Blueline Tilefish Catch Series – Mar 14, 2016	Didden 2016
SEDAR50-RD38	Mid-Atlantic Fishery Management Council SSC Memo: Proposed BLT Subcommittee Report – March 22, 2016	Miller 2016
SEDAR50-RD39	Hierarchical analysis of multiple noisy abundance indices	Conn 2010
SEDAR50-RD40	Using demographic methods to construct Bayesian priors for the intrinsic rate of increase in the Schaefer model and implications for stock rebuilding	McAllister et al. 2001
SEDAR50-RD41	Evaluating methods for setting catch limits in data-limited fisheries	Carruthers et al. 2014
SEDAR50-RD42	Technical guidance on the use of precautionary approaches to implementing National Standard 1 of the Magnuson-Stevens Fishery Conservation and Management Act	Restrepo et al. 1998
SEDAR50-RD43	A simple method for estimating MSY from catch and resilience	Martell and Froese 2012
SEDAR50-RD44	Estimating mortality from mean length data in nonequilibrium situations, with application to the assessment of goosefish	Gedamke and Hoenig 2006

2. Review Panel Report

EXECUTIVE SUMMARY

Stock assessment scientists from NOAA's Beaufort lab provided different assessment models and results for the areas south of Cape Hatteras (SOH) and north of Cape Hatteras (NOH). The decisions made by the Data Workshop and Assessment Workshop were generally sound and robust. The RW suggested one significant change to the SOH model, as explained below. The RW did not suggest any significant changes in the NOH model.

The preferred approach of the assessment team for the SOH stock was to use an age-aggregated surplus production model (also known as a Biomass Dynamic Model (BDM)) which was implemented in ASPIC (Prager 1994). A supporting analysis was provided using what the assessment team described as an age-structured production model (ASPM). In contrast, the RW preferred the ASPM over the ASPIC because it has more appropriate population dynamics and it allowed the consequences of uncertainties in the life history parameters to be explored through alternative sensitivity analysis, and hence considered the ASPM the superior base model. The results of the SOH assessment models provide robust evidence that the stock south of Cape Hatteras is not overfished and overfishing is not occurring.

For the stock NOH, only a catch history and some length frequencies were available. Because of this limitation, the R package DLMtool was used to provide TAC range estimates. The DLM analysis of the NOH stock does not provide information about whether the stock is overfished. The medians of the frequency distributions for the three methods that provide catch recommendations based on MSY approximations (Fdem_ML, SPMSY, and YPR_ML) range from 110,000 lbs to 310,000 lbs. In comparison, the average catch for the time period 2006-2015 had a median of 474,000 lbs. Given the high uncertainty in these results, the RW concluded that these results are best interpreted qualitatively, but did agree with the assessment team that the results provide evidence that the recent landings may not be sustainable in the long term. The RW also concluded that the information on potential habitat in the NOH area is insufficient to split the stocks in that area into sustainable landing recommendations along the MAFMC/SAFMC jurisdictional boundary.

The analysts responded quickly to workshop suggestions and made further improvements to the uncertainty analysis during and after the RW meeting by adding alternative sensitivity runs. The RW offered research recommendations and provided guidance on key improvements in data or modeling approaches which should be considered when scheduling the next assessment.

2.1. Statements addressing each TOR.

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

a) Are data decisions made by the DW and AW sound and robust?

In general, the RW agreed that the data decisions made by the DW and AW are sound and robust. Spatial data is limited, including information on stock structure, and this hampers the stock assessment and the determination of status. The conclusion that blueline tilefish are genetically homogeneous in the US Mid-Atlantic bight, South Atlantic Bight, and Gulf of Mexico appears well supported, although samples are not available for all areas. However, assessing blueline tilefish as one large stock encompassing its entire range may not be appropriate because, particularly in the case of recruitment, the biological processes that determine abundance and dynamics may occur at smaller scales. Based on the assumption that eggs and larvae are transported by ocean currents, via an analysis of drifter data, the analysts showed that there could potentially be significant dispersal of eggs and larvae along the coast, and that the exchange could potentially be asymmetrical with rates of movement in one direction that differ from movements in the other direction. This suggestion, coupled with a spatial mismatch between the locations with CPUE indices and the locations of recent removals, led the analysts to model blueline tilefish in the Atlantic as two stocks separate from blueline tilefish in the Gulf of Mexico. One of these stocks occupied the area extending from the SAFMC/GMFMC boundary to Cape Hatteras (the southern stock), and the other extending north from Cape Hatteras (the northern stock). While there is considerable uncertainty about the scale over which blueline tilefish recruitment processes occur, as well as degree to which drifter data may characterize egg and larval dispersal, the RW agreed that the decisions about stock structure were practical and consistent with the available information. However, the extent to which these stocks could be considered closed, an assumption of the models used in this assessment, is not really known.

Because of significant imprecision between aging of otoliths, the aging data was determined to be unsuitable for tracking cohorts in age based stock assessment. The RW hence agreed with the decision not to use the aging data.

Growth model parameter estimates and the Beverton-Holt steepness parameter were derived from a meta-analysis from similar species. The RW suggested estimating the growth parameter values using the age-structured model, which provided estimates consistent with the meta-analysis. The natural mortality rate was estimated based on the relationship between natural mortality and maximum age, using an estimate of maximum age of 40 years. Age at 50% maturity was estimated from empirical data, although, as pointed out in the Data Workshop report, too few immature fish were captured to be sure the estimated value was reliable. The workshop concurred with this but suggested maturity at age be moved from age 2 to 6 given the

information from similar species. The RW agreed with the decisions that were made about life history parameter values, but recommended that several sensitivity analyses be undertaken given their high uncertainty. These analyses are discussed later in this report.

The RW expressed concern over the spike in landings in the catch history circa 1980. The ratio of blueline tilefish to other tilefish in the catch history prior to 1985 is unknown, and the ratio from time periods after 1985 were applied to the earlier time period. Commercial fishermen present at the RW stated that many of these landings were likely golden tilefish. The RW suggested that sensitivity analyses be undertaken to address this uncertainty. The recreational catch estimates in recent years are also a concern.

For CPUE, the workshop expressed concerns over the descriptive analysis in the GLM, and the documentation of the inclusion of variables could have been improved.

b) Are data uncertainties acknowledged, reported, and within normal or expected levels?

Data uncertainties were acknowledged. Genetic and drifter study results were presented to aid in assessing the extent of interregional recruitment between the Gulf of Mexico and Atlantic coast zones. The drifter analysis supporting high egg and larval drift and connectivity could be improved by the addition of important biological variables such as pelagic larval duration, egg buoyancy, vertical migration, settlement habitat, etc. As currently presented, the utility of these studies in the stock assessment process is limited and open to interpretation.

As noted above, the aging data from otolith studies could not be adequately calibrated and was not utilized by the AW. The RW agreed with this decision.

c) Are data applied appropriately within the assessment model?

The data were applied appropriately in the SOH and NOH models.

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

The input data for the SOH model was sufficient for the two models described below. The NOH model is a data-limited model that by definition utilizes fewer data series. As noted above, the earlier tilefish landings did not sufficiently differentiate between tilefish species, and the RW did express a concern with the accuracy of landings spikes earlier in the time series. Sensitivity analyses suggested by the RW showed that this is a key source of uncertainty in the assessment for the SOH stock.

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

South of Cape Hatteras

The assessment team had been expecting to use a statistical catch-age model implemented using the Beaufort Assessment Model (BAM) software (Williams and Shertzer 2015). However, with the absence of age data the preferred approach of the assessment team was to use an age-aggregated surplus production model (also known as a Biomass Dynamics Model (BDM)) which was implemented in ASPIC (Prager 1994). A supporting analysis was provided using what the assessment team described as an age-structured production model (ASPM). The ASPIC model used the catch history and explored combinations of three fishery-dependent CPUE time series, but did not use estimates of life history parameters. Instead an "intrinsic rate of growth" (r) was estimated for the population within the model, together with a carrying capacity (K), with initial depletion (B_1/K) fixed equal to 1. The model was actually parameterized so that the free parameters were F_{MSY} and MSY but this is equivalent.

The ASPM is an age-structured model which used fixed values for the majority of life history parameters which were estimated outside the model in a variety of ways (due to the absence of valid age data). The growth parameters were estimated within the model from the length frequency data and provided estimates consistent with a meta-analysis of growth model parameters for related species. The stock recruitment relationship was assumed to be a Beverton-Holt relationship with the steepness parameter from a prior developed by meta-analysis on related species. The data inputs included the catch history, CPUE indices, and a substantial number of length frequency datasets. Sensitivity runs were performed using alternative life history parameters.

It was not clear to the RW why the assessment team favored the ASPIC over the ASPM. One of the reasons cited by the assessment team was that "the ASPM was very sensitivity to life history assumptions". However, the RW believed this was also a reason for preferring the ASPM. The ASPIC hides the sensitivity of the assessment results to the poorly known life history parameters. The use of the ASPM allows the sensitivity of results to life history parameters and the robustness of conclusions to be more fully explored.

a) *Are methods scientifically sound and robust? Do the methods follow accepted scientific practices?*

The methods used were generally sound and robust. The ASPIC is often used when life history parameters are not well known but removals and biomass indices are available. The use of the ASPIC in this case does come within accepted scientific practice. However, the RW did not consider it the best choice.

The RW preferred the ASPM over the ASPIC because it has more appropriate population dynamics and it allows the consequences of uncertainties in the life history parameters to be fully explored. The ASPIC only has one type of biomass which is particularly inappropriate if vulnerable biomass (that being selected by the fishery) is very different from mature biomass (which drives egg production). The ASPIC has no lag in recruitment which is inappropriate for species which mature or recruit to the fishery later than age 1. Also, in the ASPIC used, B_{MSY} was assumed to occur at 50% *K*. This is a very high value for B_{MSY} compared to any age-structured model using a Beverton-Holt stock recruitment relationship. A ASPIC does not model biological population dynamics properly because it does not model the various separate processes of natural mortality, growth and recruitment. In practice, it can be used successfully to model the past empirical link between catches and abundance. However, this may provide only a poor prediction of future abundance and offers no clear way to explore alternative scenarios or structural uncertainties.

The use of the ASPM as the base model would also have the advantage that the considerable length frequency data available could be used. These data not only allowed the estimation of fishery selectivities but also, in the runs requested by the workshop, allowed the growth parameters to be estimated within the ASPM model. However, a model run with the length frequency data heavily weighted relative to the CPUE indices suggested that it may contain a different signal about changes in mortality rates and abundance than the CPUE indices. This issue was not fully explored at the RW, and if real, the case is not known.

b) Are assessment models configured appropriately and applied consistent with accepted scientific practices?

Generally the models were configured appropriately and applied within accepted scientific practices. However, the RW identified a number of issues which needed to be addressed.

The first issue was that the CPUE indices were fitted using only observation error (with CVs as low as 7% in some cases). Abundance indices need to have a component of "process error" which is a consequence of assumptions being violated. In particular, and especially for CPUE, it is likely that the proportionality constant (*q*) for the assumed linear relationship between CPUE and biomass actually varies from year to year. This produces an additional component of variation which is not captured by estimates of observation error. Because the CVs of the CPUE indices were not inflated to allow for process error the handline index (which had the lowest CVs) dominated the longline CPUE index in the model where they were both fitted. This led the assessment team to fit each index separately and then average the results from the two runs.

The averaging of results from two separate runs to provide a final assessment is not the best approach for the type of stock assessments being applied here. If the two runs are telling "very

different stories" then they need to be kept in separate runs (one of the runs may be providing the "truth"). If the two runs are not inconsistent then there should just be a single run with all of the data included. The base model recommended by the RW does include both CPUE time series where each time series is given equal weight (CVs = 20%).

The ASPM runs were generally appropriate, but the base ASPM model had full maturity at age 2. This had been calculated using the length at maturity data and the externally estimated growth parameters. However, it was the general feeling of the workshop and indeed the whole meeting that full maturity at age 2 years was very unlikely for this species. A new reference model was proposed by the workshop for exploring results from the ASPM which had full age at maturity at 6 years. This is a conservative value in that the younger the age of maturity the more resilient the stock is to exploitation (according to the estimated fishing selectivities fish are not exploited until about 6 years of age).

c) Are the methods appropriate for the available data?

The methods were defensible given the available data. The final base ASPIC model used the handline and longline CPUE indices and excluded the problematic headboat time series. The ASPIC could not use the length frequencies but these were fitted in the ASPM runs. The ASPM runs used all three CPUE time series but the headboat series was split into two periods to account for changes in fishing practice (resulting in a change in selectivity).

<u>North of Cape Hatteras</u>

For the assumed stock to the north of Cape Hatteras only a catch history and some length frequencies were available. The life history estimates were borrowed from those used for south of Cape Hatteras. The R package DLMtool was used to provide TAC range estimates (Carruthers and Hordyk 2016; R Core Team 2016).

a) Are methods scientifically sound and robust? Do the methods follow accepted scientific practices?

Various data limited assessment methods exist and the DLMtool provides access to a number of such methods. It must be understood that where there are little data any stock assessment results should be treated cautiously as they are, in reality, very uncertain. With this acknowledged it is reasonable and scientifically defensible to use such a tool to provide some idea of the range of possible TACs. Six methods were used to provide alternative distributions describing possible TACs. Little is known about the relative performance of the individual methods which will be highly case specific.

b) Are assessment models configured appropriately and applied consistent with accepted scientific practices?

Appropriate data and estimated CVs were supplied to the DLMtool.

c) Are the methods appropriate for the available data?

No biomass indices are available so the use of the DLMtool is appropriate. However, two of management procedures use the Gedamke and Hoenig (2006) non-equilibrium mean length method to estimate the instantaneous total mortality rate. An assumption of this method is that all fish larger that the length at which they are fully selected to the fishery experience the same total mortality rate. The RW noted that the method would be less appropriate if the fishery selectivity was dome-shaped, an assumption at times made for gears such as hooks. The effect of violating this assumption would be to over-estimate the mortality rate, but the extent to which would depend in part on the shape of the selectivity curve. Additionally, the extent to which a mean length estimate would change with a change in total mortality would depend upon the shape of the selectivity curve for the gear used to sample the population.

TOR 3. Evaluate the assessment findings with respect to the following:

<u>South of Cape Hatteras</u>

a) Are population estimates (model output – e.g. abundance, exploitation, biomass) reliable, consistent with input data and population biological characteristics, and useful to support status inferences?

When interpreted together with their associated uncertainties, the RW concluded that the estimates of abundance, exploitation and biomass are reliable, consistent with the input data and biological characteristics of the stock, and useful to support status inferences. Uncertainties include those resulting from model selection and configuration, parameter estimation, the dated CPUE abundance indices, and the removals time series. As discussed below, the many sensitivity analyses undertaken by the assessment team indicate that the results appear most sensitive to uncertainty in the removals time series, specifically the very high landings in the early 1980's.

As discussed above (TOR 2), rather than fitting the ASPIC model with the handline and longline indices separately and averaging the results, the RW recommended a base model fitting the ASPIC model with the two indices in the same model but with equal weights. The RW agreed that the ASPM results worked well as a supporting analysis and requested several sensitivity analyses using this model. However, the results of the two models are not directly comparable because biomass is treated differently in the two models. ASPIC models biomass is in terms of what is available to the fishery, whereas the ASPM analyses estimate both total and spawner

biomass, neither of which are estimated with the ASPIC model. The RW suggested that depletion ratios (for trend) and projected yields at $F=F_{current}$ (for scale) were quantities produced by both models that might be comparable.

The ASPIC base model run, and the other model runs that do not include the headboat index, all indicate a sharp biomass decline in the early 1980's followed by increasing biomass in the 2000's to about 50% of carrying capacity (Figure 3.1). With the exception of the sensitivity run in which the peak removals in the early 1980's are reduced by 90% (Catch 0.1), biomass trends from the ASPM runs follow a similar pattern (Figure 3.2), although the extent of the biomass decline tends to be less than the ASPIC model runs. With this exception, the ASPM sensitivity runs generally support the trends produced from the ASPIC base model.

The stand out sensitivity is to the magnitude of the catch spike in the 1980s. If the spike is genuine then there must have been sufficient biomass to support the catches, and the subsequent decline in catches coupled with deterministic recruitment means that the stock must have increased to be consistent with the handline and longline indices. However, if the catch spike is removed altogether then a different story emerges. The run shows a different pattern in which biomass is further depleted and has not rebuilt to 50% of carrying capacity by 2015. It may be that the 10% run (which removes the spike completely) is a bit too extreme, but the impact of reducing that catch spike is very apparent in this sensitivity. The ASPM Catch 0.1 sensitivity was undertaken because of the uncertainty in early landings expressed at the RW, and the sensitivity of the model results to this data input confirm that it is a key source of uncertainty to consider when applying the assessment results.

Deterministic yield projections at $F=F_{current}$ from the ASPIC and ASPM reference model runs differ by about 23% (Table 3.1), suggesting that, although the biomass estimates are not directly comparable, the two models are producing estimates on roughly the same scale. Yield projections fishing at F_{MSY} differ markedly, but are not comparable given the differences in assumptions underlying the models (TOR 2).

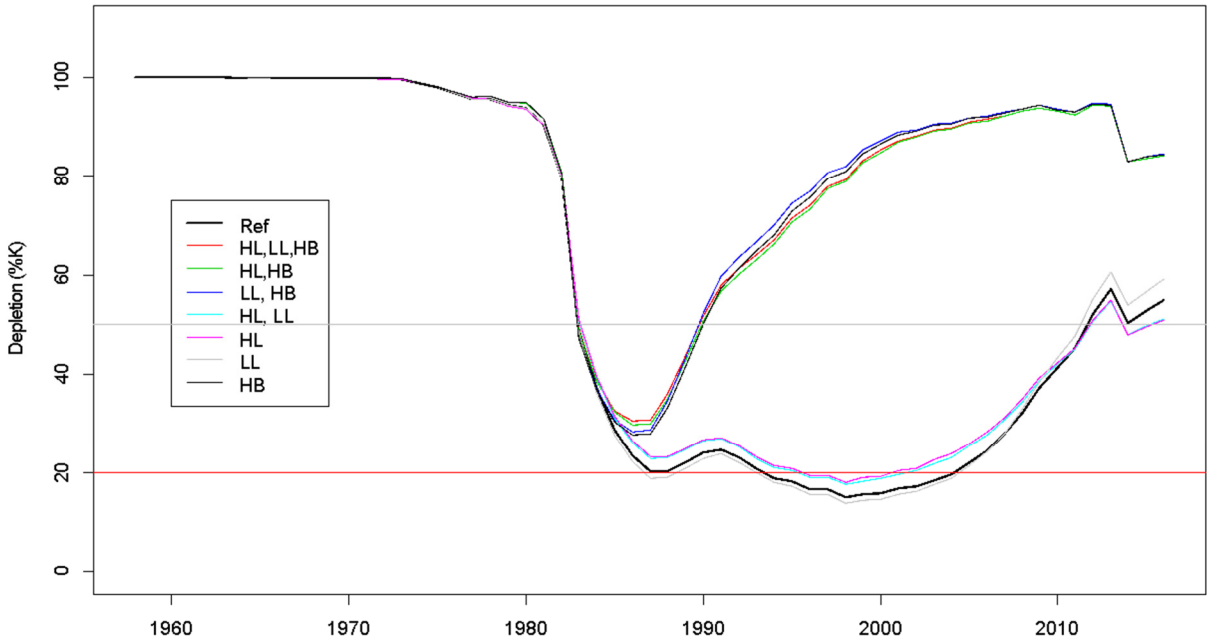

Figure 3.1. Comparison of depletion (% of carrying capacity) for the base ASPIC surplus production model (heavy dark line) and sensitivity analyses undertaken with the model (colored lines) with respect to fitting to various combinations of the handline (HL), longline (LL) and headboat (HB) indices.

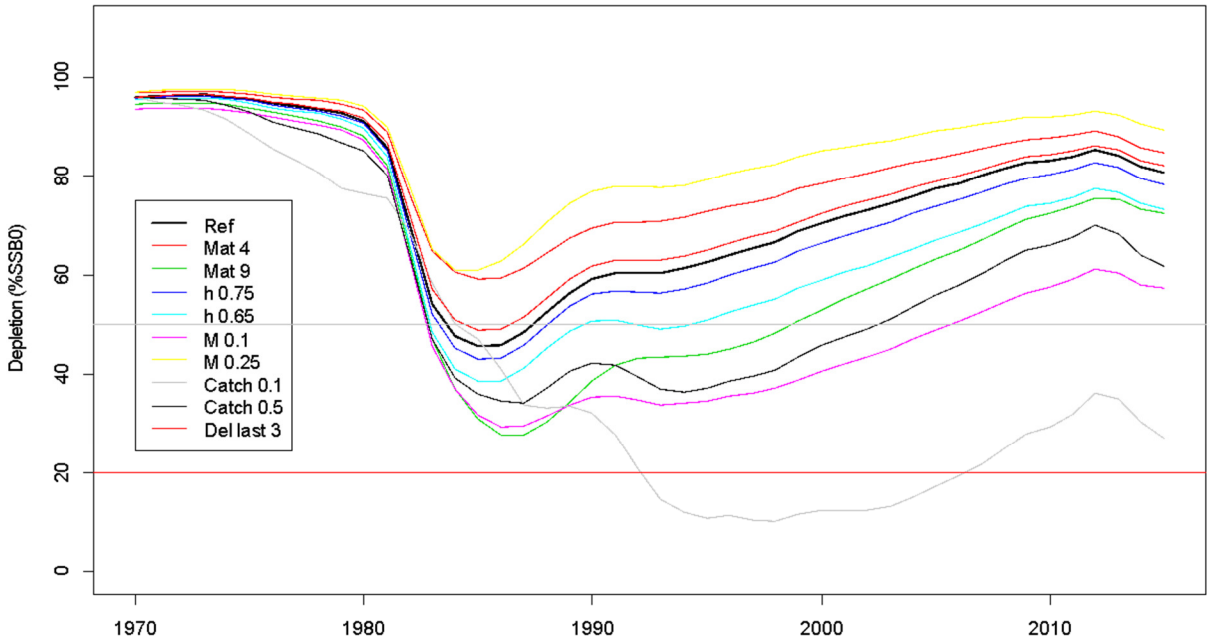

Figure 3.2. Comparison of depletion (% of unfished spawner biomass) for the reference ASPM analysis (heavy dark line) with sensitivity analyses undertaken with the model (colored lines). Sensitivity analyses include: older ages-at-maturity (ages 4 and 9); lower steepness values (0.75, 0.65); different values for the instantaneous natural mortality rate (0.1, 0.25); reducing the peak landings in the early 1980's to 50% and 10% of those used in the reference model; and removing the last three years in the indices.

Table 3.1. Comparison of deterministic yield projections at $F=F_{current}$ from the ASPIC and ASPM reference model runs. Adapted from SEDAR 50 Addendum Tables A5 and A23.

Year	Yield (1000 lbs)						
	ASPIC			ASPM			
	$F=F_{MSY}$	$F=F_{current}$	$F=F_{target}$	$F=F_{MSY}$	$F=F_{current}$	$F=F_{30\%}$	$F=F_{40\%}$
2017	236	218	180	900	166	795	569
2018	233	218	184	766	165	689	521
2019	231	218	187	674	164	617	484
2020	229	217	190	610	163	564	455

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

The results of this assessment provide evidence that the stock south of Cape Hatteras is not overfished. Information in support of this conclusion includes estimated *B/MSST* ratios above 1.3 from both the ASPIC model for the base model recommended by the RW, and sensitivity analyses pertaining to which CPUE series are included in the model (Table 3.2). Additionally, for the base model, projection results with $F=F_{current}$ indicate a probability of *B>MSST* of 0.97 in 2016 (SEDAR 50 Addendum Table A5). Analyses undertaken using the ASPM provide further support for this conclusion: *B/MSST* ratios from the ASPM sensitivity analyses are above 2.2 (Table 3.3).

This conclusion is subject to some uncertainty resulting from uncertainty in the removals time series (the peak in the early 1980's) and questions about the derivation of the CPUE time series. Additionally, the CPUE series do not provide information on relative abundance since 2007. Recent status determinations are based on productivity as estimated to the end of the CPUE series, projected forward under the assumption that productivity has not changed.

Table 3.2. Estimates of status determination criteria for the South of Cape Hatteras stock from ASPIC for the base model run (bold) and sensitivity analyses undertaken with the model with respect to fitting to various combinations of the commercial handline (HL), longline (LL) and headboat (HB) indices. The numerator in F/F_{MSY} is the geometric mean F from the last three years of the assessment (2013-2015) and B is the biomass in the terminal year of the assessment (2015). Adapted from SEDAR 50 Addendum Table A3.

Run	RunName	F/F_{MSY}	B/B_{MSY}	$B/MSST$
80	**Base (HLLL CV0.2)**	**0.92**	**1.05**	**1.40**
51	HLLLHb	0.41	1.67	2.23
52	HLHb	0.41	1.67	2.23
53	LLHb	0.39	1.68	2.24
54	HLLL	1.06	0.99	1.32
55	HL	1.07	0.99	1.32
56	LL	0.81	1.13	1.51
57	Hb	0.40	1.68	2.23

Table 3.3. Estimates of status determination criteria for the South of Cape Hatteras stock from ASPM for the reference model run and selected sensitivity analyses undertaken with the model. Sensitivity analyses include: older ages-at-maturity (ages 4 and 9); lower steepness values (0.75, 0.65); different values for the instantaneous natural mortality rate (0.1, 0.25); reducing the peak landings in the early 1980's to 50% and 10% of those used in the reference model; and removing the last three years in the indices. Adapted from SEDAR 50 Addendum Table A21.

Description	F/F_{MSY}	B/B_{MSY}	$B/MSST$
Reference	0.12	4.96	6.61
A100is4	0.07	4.87	6.49
A100is9	0.21	4.79	6.39
Steep0.75	0.17	3.76	5.02
Steep0.65	0.26	2.87	3.83
M0.1	0.34	2.82	3.76
M0.25	0.06	6.25	8.34

0.1HLRemo80to85	0.92	1.66	2.21
0.5HLRemo80to85	0.29	3.76	5.02
IndRemLast3yr	0.11	5.06	6.75

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

The results of this assessment provide evidence that the stock south of Cape Hatteras is not undergoing overfishing. Information in support of this conclusion includes the estimated F/F_{MSY} ratios less than one from the base ASPIC model run recommended by the RW (Table 3.2), and from the ASPM analyses (Table 3.3). This conclusion is less certain than the conclusion about whether the stock is in an overfished state. For the ASPIC reference model, bootstrap runs indicate a significant portion of the probability density for F/F_{MSY} that is above one (SEDAR 50 Addendum Figure A5). Additionally, the conclusion is sensitive to decisions about which CPUE indices to include in the model. Two runs including the commercial handline CPUE index provided estimates of F/F_{MSY} ratios slightly greater than one (Table 3.2).

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

Most fisheries models that incorporate population dynamics are based on the assumption that the population is closed, and stocks are often delineated such that this assumption is met. Recruitment dynamics for Atlantic blueline tilefish are poorly understood, and the Assessment Workshop Report discusses, with high uncertainty, how egg and larval drift could result in dispersal among the three putative stocks, resulting in stocks for which recruitment is partially driven by immigration and emigration. In the absence of information about the extent to which immigration and emigration contribute to recruitment processes, the RW concurred that using a stock recruitment relationship (or logistic growth curve for ASPIC) to evaluate productivity and future stock conditions is practical at this time.

ASPIC and ASPM model population dynamics differently; the former using a logistic growth curve that can be used to evaluate productivity. The assumption of the symmetrical shape provides estimates of productivity that differ from what would be expected from an age structured model. The four ASPIC model runs that did not include the headboat index, produced estimates of F_{MSY} that were relatively consistent (SEDAR 50 Addendum Table A3; range: 0.23 to 0.33), implying that that r_{max} is also estimated consistently with these models. Similarly, the estimates of K from the four models were also consistent. The RW accepted that these were useful for evaluating productivity and future stock conditions, noting also that the functional relationships are consistent with those used in the assessment model. However, ASPIC does not

discriminate among all types of production including fish growth as well as well as recruitment, so this is not strong evidence of a closed population or simple stock recruitment relationship. Furthermore, BMSY estimates from the ASPM appeared too low to be considered as target reference points, so the RW does not believe a reliable stock recruitment relationship has been demonstrated for this stock. It is likely that the stock structure will need to be modelled more accurately before an accurate stock recruitment relationship could be determined.

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

The RW concurred that the quantitative indicators of status determination are appropriate for management use for this stock. The status determinations were robust to many sensitivity analyses undertaken using both ASPIC and ASPM. With respect to current biomass status and projected yields, the base model run from ASPIC provides advice that is lower than some sensitivity model runs and most ASPM model runs, but not all of these are equally plausible.

North of Cape Hatteras

a) Are population estimates (model output – e.g. abundance, exploitation, biomass) reliable, consistent with input data and population biological characteristics, and useful to support status inferences?

The RW accepted that the DLM analyses undertaken for northern blueline tilefish are appropriate given the limited amount of data available for this stock. The assessment team filtered the many methods available in the DLM toolbox to select methods that would provide TAC recommendations. These included three methods that use MSY approximations, and five scenarios based on average catch. Although these analyses do not provide time series of abundance, exploitation, or biomass, they do provide output that can be used to compare average catch management procedures with management procedures based on MSY approximations.

These analyses are consistent with the input data and biological characteristics and subject to uncertainty, provide a basis to support status inferences with respect to whether the stock is undergoing overfishing, but not whether it is in an overfished state.

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

The DLM analyses undertaken for the blueline tilefish stock north of Cape Hatteras do not provide information about whether the stock is overfished.

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

The medians of the frequency distributions for the three methods that provide catch recommendations based on MSY approximations, Fdem_ML, SPMSY, and YPR_ML, range from 110,000 lbs to 310,000 lbs. In comparison, the average catch for the late time period (AvC.late: 2006-2015) had a median of 474,000 lbs. although the value would be sensitive to the years used in its calculation. All recommendations are highly uncertain and the frequency distribution for the AvC.late overlaps with the distributions from the methods that use MSY approximations (SEDAR 50 AW Report Figure 61).

Given the high uncertainty in these results, the RW concluded that these results are best interpreted qualitatively, but did agree with the assessment team that the results provide evidence that the recent landings may not be sustainable in the long term.

Analyses of annual removals by habitat area suggest that recent removals per unit area in the area just north of Cape Hatteras are high relative to annual removals per unit area for the southern stock. The implications of this analysis are highly uncertain because relative habitat quality in the two areas is not known, there is subjectivity in the selection of areas for habitat standardization, and the abundance of this stock even further north is not well known. The RW agreed with the assessment team that the analysis appears to suggest that such high removals over such a small area may be cause for concern: specifically that they could lead to localized depletion. However, the effects on the entire stock north of Cape Hatteras are unknown.

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

Given the very limited amount of information available for the stock north of Cape Hatteras, the DLM analysis was not set up to provide an evaluation of productivity and future stock conditions. Although an estimate of the steepness of the stock recruitment relationship was provided to the model, because the estimate was obtained from a meta-analysis of species with similar life-histories, its representativeness for this stock is unknown. Additionally, the unfished equilibrium recruitment is not known. For these reasons, the RW concurred that an informative stock recruitment relationship is not available.

The DLM toolbox does provide methods for projecting future stock conditions, but these methods are intended primarily for the evaluation of management procedures. Given the wide relative frequency distributions for the TAC's provided by the methods that include MSY approximations, the RW agreed with the assessment team's decision not to undertake these analyses at this time.

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

Quantitative estimates of status determination criteria for the North of Cape Hatteras stock were not provided by the DLM analyses. The RW concurred that this was appropriate given the limited data available for this stock.

TOR 4. Evaluate the stock projections, addressing the following:

South of Cape Hatteras

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

Model projections used both maximum likelihood and bootstrap parameter estimates with fixed fishing mortality covering the range of plausible alternatives. The estimated F in 2015 was applied in 2016, and thereafter 2017-2021 the different scenario fishing mortalities were applied. For each projection, information was provided on the stock status and yields for the projection period 2016-2021. The projected yield was clearly compared to current yield, providing suitable information for setting total allowable catches.

The approach used for projections was appropriate for the models and available data and represents good practice. The use of fishing mortality and range of fishing mortalities tested was appropriate and is standard practice. Uncertainties were captured using the bootstrap simulations.

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

The methods applied by the analytical team were appropriate for the models and outputs. The structure of the projection models was the same as that of the respective assessment model.

For the ASPIC model, the projection 2017-2021 was conducted for fixed fishing mortalities: $F = F_{MSY}$, $F=F_{2015}$ and $F=F_{target}$ (75% F_{MSY}). The biomass and yield was reported from the maximum likelihood estimates and the 1000 bootstrap estimates, consisting of the median biomass and the probability that biomass is greater than B_{MSY} and MSST. Similarly, for the ASPM assessment, the projection 2017-2021 of fixed fishing mortalities was applied, but in this case fixed fishing mortalities applied were $F = F_{MSY}$, $F=F_{2015}$, $F=F_{30\%}$ and $F=F_{40\%}$. No ASPM bootstrap estimates were available for the review because this was not the analytical teams preferred model. However it is expected that all suitable indicators will be in the final report, including probability SSB is above the MSY proxy (recommended $F_{40\%}$ level) and the MSST.

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

The projection results are informative and reasonably robust. The results are described in relation to the most important indicators and the analytical team has demonstrated that management advice should be robust to the key uncertainties.

The short term projections are useful for indicating probable future conditions for the fishery, but precision of future estimates will decrease rapidly. The RW agreed with the analytical team that projecting any further than 5 years could be misleading given the limited data and uncertainties.

For the ASPM assessment, the RW recommended not to use the estimated F_{MSY} as a target, as it cannot be estimated reliably and the estimate is not precautionary. $F_{30\%}$ and $F_{40\%}$ are projected as well and it was agreed that these would make more suitable targets for this assessment. These spawner-biomass-per-recruit reference points would also be more robust to over-estimation of the stock's productivity if a significant amount of recruitment originated from larval transport from outside the stock's boundary (they do not depend on a stock recruitment relationship).

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

The key uncertainties are the structural uncertainties covered by the sensitivity analyses and the observation errors covered by the bootstrap estimates. For the ASPIC model, the bootstrap runs were used for the projections to generate the probability estimates for stock status and supply appropriate confidence intervals for indicators. For ASPM, no bootstrap was carried out for the review meeting, so the RW did not see the ASPM projections with confidence intervals or probability estimates for stock status, only projections based on the maximum likelihood estimates. Bootstrap projections are needed for the ASPM assessment and are expected to be carried out. It was assumed that the median bootstrap estimates would be close to the maximum likelihood estimates for the indicators of interest. The bootstraps should provide adequate estimates of uncertainty caused by the observation error.

Alternative projections have been run with the key sensitivity analyses. The RW identified an alternative catch history with a lower proportion of the tilefish catch 1981-85 allocated to blueline tilefish as the key structural uncertainty. These additional runs, with the bootstraps, should reflect the key uncertainties in the projection results.

The between model uncertainty is very large as the projected yields from the ASPM reference run at $F_{40\%}$ are more than five times higher than those for the sensitivity that removed the 1980s catch spike (Table 4.1). The projected yields for the ASPIC model at F_{targ} are less than half of those from the ASPM reference run (Table 4.1), indicating that results from these models are highly uncertain. If the magnitude of the spike in the 1980s catch is genuine then much higher catches than have recently been removed can be sustainably taken.

Table 4.1: Deterministic projected yields (t) using $F_{current}$ for 2016 and the indicated fishing mortality for 2017–2020. Estimates are given for the ASPIC base model, the ASPM reference run, and the ASPM that removed the spike in the early 1980s catches.

	ASPIC F_{targ}	ASPM ref $F_{40\%}$	ASPM no-spike $F_{40\%}$
2016	219	168	139
2017	180	569	73
2018	184	521	78
2019	187	484	82
2020	190	455	85

North of Cape Hatteras

No projections were run for the fishery North of Cape Hatteras. Given the available data and methods, it was not appropriate to carry out projections for this fishery. The method "DLMtools" implements simulations to test management procedures and does not provide an estimate of stock status. Therefore, it was not appropriate to project the population status forward. Instead, simulations applied fixed catches to various stock models to provide a benchmark for current catches and help determine whether they were sustainable.

TOR 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 all sources of uncertainty in the population, data sources, and assessment methods

b) Are the implications of uncertainty in technical conclusions clearly stated?

The stock structure is a significant source of uncertainty. Evidence was presented that suggested there were strong linkages through recruitment between the Gulf of Mexico and along the whole Atlantic coast. While the RW understood and supported the decision to assess three areas separately, this source of uncertainty has not been addressed in the stock assessment outputs. An important RW recommendation for future assessments is to develop several models with different spatial links to run at least as alternative sensitivity analyses.

South of Cape Hatteras

Uncertainty was evaluated using two different models, sensitivity analyses and bootstrapping. The implications of the uncertainties are clearly presented and stated in the text and figures. Both the sensitivities and bootstraps were used to report uncertainty in parameter estimates and derived quantities and indicators, so that the consequences of the uncertainty could be evaluated. No bootstrapping was carried out for the ASPM assessment for the RW. The RW recommended this model for management advice, and requested bootstrapping be carried out for the final assessment report. It is expected that the ASPM median bootstrap estimates for important estimated quantities will be close to the maximum likelihood estimate.

Sensitivity analyses considered a wide range of assumptions and model structures. For the ASPIC model, the sensitivities covered which abundance indices were included in the model. The ASPM tested the effect of various life history parameters, including using values from golden tilefish, the previous assessment and setting M and age-at-maturity to alternative plausible values.

Bootstraps are based on sampling theory and simulate alternative data sets using the available data. They are primarily used to assess the effect of observation error. There are many different ways to implement bootstraps. They are a collection of *ad hoc* procedures and many lack a complete underlying theory to support their use, but they are a very flexible tool to characterise uncertainty, particularly when used alongside sensitivity analysis. However, more modern techniques using likelihoods and priors (e.g. MCMC) are generally preferred to cover all sources of uncertainty, including observation and process error.

Additional sensitivity runs were requested by the RW and completed by the stock assessment team. These covered some additional tests for the robustness of the assessment. Notably, reduce the length of the abundance indices, higher, but plausible, age at 100% maturity, lower M, lower SR steepness and an alternative catch history allocating a lower proportion of the tilefish catch 1981-85 to blueline tilefish. The RW identified the alternative catch history as a key structural uncertainty.

The RW believes that the sensitivity analyses capture the most important uncertainties in the assessment and provide important information on the likely plausible range for indicators such as spawning stock biomass and fishing mortality.

North of Cape Hatteras

The DLM tool package is a MSE tool for testing management plans in data poor and hence highly uncertain, fisheries. It simulates population trajectories under different harvest regimes. In each case, key population parameters are drawn from uniform probability distributions covering what is believed to be their possible range. The process gives guidance on setting TACs based on

the decision rule and population assumptions, and explicitly addresses the uncertainty. Possible TACs for each rule were presented as densities, but uncertainty over resulting stock status was not evaluated. All consequences for the uncertainty cannot be evaluated and therefore the analytical team was unable to recommend any particular method. Future work could apply the different catch rules to the same population model and data scenarios so that a better comparison might be made on their performance with respect to uncertainty.

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.*

Research recommendations were provided by the Data and Assessments workshops and were reviewed at the Review Workshop. The two main areas where further research would help improve the assessment for both stocks are development of fishery independent indices and resolution of the age determination issues. A third area of research pertains to improving our understanding of biological processes such as reproduction and recruitment dynamics.

NMFS should continue the development of fishery dependent and independent indices for these stocks. The development of a continuous, random, stratified fishery-independent survey implemented for blueline tilefish throughout its range is potentially the single most important recommendation for this stock. The survey could also be used to fill other knowledge gaps (e.g. incorporate a hook selectivity study).

Reliable age reading or growth curve development is also urgently needed for blueline tilefish. If possible, the age reading issues of blueline tilefish otoliths should to be resolved. Other age validation techniques or methods to derive reasonable growth curves should be investigated (e.g., Pb\Ra ratio, or tagging studies). Resolution of the age determination issues would allow considerable past information to be incorporated into the assessment.

Beyond those priorities, further understanding of egg and larval dispersal through biophysical modeling and genetic analysis may be useful. Studies should be conducted on the identification of blueline tilefish larvae and also on the location, duration, and dispersal mechanisms of the egg and larval stages. Eggs and larvae can effectively be identified using genetic techniques, and there are apparently many samples that have been collected but have not yet been genotyped. A good place to start would be genotyping more of the available samples. This would also benefit the science on other species.

Increased observer coverage for both commercial and recreational fisheries in the South Atlantic would improve discard estimation and provide estimates of discard sizes and weights. Implementing electronic monitoring of bycatch would improve discard estimation accuracy and provide size and weight composition of discards. In the absence of a fishery-independent index, additional information on the targeting behavior of fishermen, in particular the depth or geographic locations fished within a given trip as well as more refined information on fishing effort is needed.

Investigation of alternative methods in proportioning unclassified tilefish to golden, blueline, or other species could be explored. For recreational fisheries, research into and implement rare-event data collection procedures (e.g. mandatory reporting, logbooks, reef fish stamp to determine universe.) could also benefit the assessment. ,

An increase in sample size on the catch compositions from both commercial and recreational fisheries would be useful.

Estimates of immediate and delayed discard mortality through tag-recapture, acoustic tagging, or other methods in both commercial and recreational fisheries are needed. Special interest was expressed into developing mortality estimates when using descender devices to aid recompression, since these devices may have the potential to substantially lower mortality rates

Special effort could be put into the data collection in North and South of Cape Hatteras to better understand the spatial distribution of fishing effort, size composition of catch, and catch rate.

Further study on the maturity and reproductive biology would be of value to the assessment. Reproductive biology studies of blueline tilefish should be expanded to include the full distributional range of the species, specifically targeting samples from the west and east coasts of Florida and the Mid-Atlantic region. These data are needed to assess possible shifts in spawning season. Sampling of young fish is needed to improve the maturity ogive to improve reproductive parameters estimates.

Particularly if the age determination issues remain unresolved, development and use of methods to provide growth parameter and natural mortality estimates (may serve as prior elicitation in the future) would benefit the assessment. The focus should be on acceptable approaches for parameter values and error distributions (e.g. meta analyses, use of related species, use of species with comparable life history strategies, etc.)

With respect to assessment models, consideration should be given to developing Bayesian, age-structured models to integrate prior knowledge from the meta-analysis on the life history processes.

• *Provide recommendations on possible ways to improve the SEDAR process.*

The current SEDAR process looks well designed. The process provides for a thorough review and evaluation of the available data, provides thorough consideration and review of analytical approaches and modeling results, provides very good guidance on the information expected to result from the process, and provides very good documentation of the process including decisions made throughout the assessment. The process is highly transparent, particularly because documents produced for review remain unedited after the review. The pre-Review Workshop teleconference is a good component that can help get the Review Workshop meeting to quick start by providing the analytical team advance notice of areas that the RW is likely to question. Distributing presentations in advance of the workshop is useful. As a minor recommendation would be to ensure there is time for at least two rounds of review of the Review Workshop report in the event that there are significant additions to the report or addendum material provided to RW after the meeting is adjourned. This would help to ensure that the RW has the opportunity to reach consensus on all aspects of the report.

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

A high quality stock assessment requires careful data preparation as well as the use of appropriate modelling tools.

The length frequency data were not carefully analyzed and may not have been appropriately stratified and scaled. An analysis of the variability of fish length within each fishery should be undertaken before the next stock assessment so that appropriately scaled length frequencies can be produced for the years within each fishery where there are adequate data.

The CPUE standardizations were not well documented and more work may have been done than was described. However, there is clearly the need for more detailed analysis. The catch and effort data should be fully investigated and explored before a standardization is attempted. Such a descriptive analysis provides the foundation for a standardization. Explanatory variables need to be carefully chosen and should include effort variables. Hook hours may not be the best unit of effort as bait is not necessarily effective beyond 30 minutes. Also, interactions and/or nested effects need to be considered. For example, seasonal effects may differ by subarea. Area-year interactions especially need to be considered as different trends in different subareas will require an exploration of the effect of different weightings (of the trends across subareas to produce an overall trend).

Hopefully, age data will be available for the next stock assessment and a statistical catch-at-age model can then be used. Should there be a stock for which no age data are available it would still best to use an age-structured model and fit whatever data are available. Any sensitivities to poorly known life history parameters should be explored in an age-structured model rather than hidden by using the ASPIC.

Capturing the uncertainty in stock assessment results using bootstrap procedures is adequate but not ideal. There are many ways to bootstrap any particular problem which means that the approach is necessarily *ad hoc*. It is better to use a formal likelihood approach with asymptotic approximations to confidence/credibility intervals or to adopt a formal Bayesian approach.

2.2. References

Gedamke, T. and Hoenig, J.M., 2006. Estimating mortality from mean length data in nonequilibrium situations, with application to the assessment of goosefish. *Transactions of the American Fisheries Society*, 135(2), pp.476-487.

Prager, M. H. 1994. A suite of extensions to a nonequilibrium surplus-production model. *Fishery Bulletin*, 92, pp.374–389.