Independent peer review report for SARC 66 --- Summer Flounder and Striped Bass Stock Assessments

Yan Jiao

Prepared for

Center for Independent Experts

January 2019

Contents

Executive Summary

The 2018 stock assessments of summer flounder and striped bass were reviewed by a CIE review panel. The review aims to determine whether or not the scientific assessments are adequate to serve as a basis for developing fishery management advice within the jurisdiction of NOAA's Greater Atlantic Regional Fisheries Office (GARFO). The SARC 66 Review took place at the Northeast Fisheries Science Center (NEFSC), Woods Hole, Massachusetts during November 27 - 30, 2018. The assessments of the stocks done by the stock assessment team were presented to the peer review panel and the validity of the data, assessment procedures, and results were discussed. The NEFSC provided all the necessary logistic support, background information, documents and further data and model configuration explorations that were requested by the review panel.

All the stock assessments for the two species have been modified with new changes in their model structures and data usage, and data are updated to 2017. All the models were either processed using the software package Age-Structured Assessment Program (ASAP; Legault and Restrepo 1998; NFT 2013a) or coded in ADMB, and likelihood approach was used to estimate parameters. The 2018 SARC 66 peer review panel found that the stock assessments addressed all the terms of references (TORs) adequately.

There were no major weaknesses in the assessment of summer flounder, but improvements could be made with regards to exploring reasons for patterns of recruitment changes over time and better incorporating the population's temporal changes in its spatial distribution with uncertainty reflected from regional and large scale surveys into the stock assessment models (Jiao et al. 2016; Jardim et al. 2018).

Alternative models to address sex-specific dynamics and length-based dynamics were also provided for the summer flounder stock assessment as background material or appendices. During the review week, these models were not presented and discussed in details. Based on the documented public comments, the life history of this species and the previous research recommendations, it is very likely that these model constructions will continue to be discussed in the future. It is more desirable in the future to compile the reports in the model description section and compare them based on a model selection framework and incorporate model selection uncertainty. Measurements for selecting models can be model goodness-of-fit, model prediction ability, model robustness and fisheries-specific measurements such as retrospective error (Jiao et al. 2012; Gelman et al. 2014; Hooten and Hobbs 2015).

The review panel recommended a one-stock statistical catch-at-age model to provide advice on striped bass fishery management but rejected the 2-stock statistical catch-at-age (2SCA) model recommended primarily by the working group. The 2SCA model for striped bass was constructed to meet the explicit

management needs but it was not ready because of problems in model construction, model fitting and model stability. The estimated productivity and biological reference points (BRPs) for each stock from this model are designed to meet the explicit management needs but are strongly influenced by the movement rate and stock composition assumption and/or estimation. The model uncertainty and key parameter estimation uncertainty need to be well addressed before a model such as 2SCA can be used. The tagging program for striped bass has lasted for a long time and across a wide range of the stock's spatial distribution. The program provides information on natural, fishing mortality, movement and population size. Uncertainties for the key parameters from the tagging models are suggested to be estimated and documented before they are used directly in the 2SCA models in the future. Future model development that integrates the SCA framework with the tagging data is encouraged strongly. Bayesian approaches, which provide more flexible frameworks for integrating the fit of alternative submodels to different sources of data, seem more appropriate.

Both the two recommended benchmark assessments for the two species are considered to represent the best scientific information available for the two species, and provide suitable bases for management advice.

1. BACKGROUND

The 2018 assessments of summer flounder (*Paralichthys dentatus*) and striped bass (*Morone saxatilis*) stocks along the Atlantic coast (Maine to North Carolina) waters were reviewed as part of the SARC 66 (Stock Assessment Review Committee No. 66) process. The assessment Review Panel met at Woods Hole, Massachusetts from Nov 27 – 30, 2008. The review panel chair was Dr. Robert Latour, and the CIE review panel included Drs. Robin Cook, John Casey and Yan Jiao.

The SARC 66 review process was coordinated by the NEFSC SAW Chairman, Dr. James Weinberg, and facilitated and supported by Drs. Russell Brown and Toni Chute (NEFSC). The stock assessment documents for summer flounder were prepared by the Summer Flounder Working Group and were presented at the meeting by Dr. Mark Terceiro (NEFSC). Stock assessment documents for the striped bass were prepared by the ASMFC Striped Bass Working Group, which is composed of members from Technical, Stock Assessment and Tagging Committee, and presentations were made by Drs. Gary Nelson, Katie Drew and Michael Celestino.

According to the CIE scope description, the SARC peer review is "the cornerstone of the Northeast Stock Assessment Workshop (SAW) process, … This review determines whether or not the scientific assessments are adequate to serve as a basis for developing fishery management advice". As a review panel member, I was provided with draft stock assessment reports and web access to relevant files and documents (Appendix 1) and participated in the Stock Assessment Review Meeting. The assessments of the stocks were presented to the review panel and the validity of the data, assessment procedures and results were discussed. Extra documents and model runs were provided during the review upon request from the SARC peer review panel.

For the two species/stocks under review, the working groups presented the existing stock assessment model used, new and alternative models that were explored, the BRPs estimated, and the fishery and population status (see Agenda in Appendix 2). There were discussions throughout the review on the quality of the data including the data standardization or synthesis, the appropriateness of the model equations and error structures, key parameters, and the estimation algorithms.

The quality of the data for each species was discussed and comments and suggestions are listed under TORs. During the review meeting, the stock assessment team was always available when required for further discussion, additional data exploration and clarification, and clarification of how the summer flounder and striped bass TORs were addressed. Recommendations from the last SARC review, SSC reports and/or from Technical Committee were reviewed to determine the extent to which they had been addressed by the working groups.

As a CIE reviewer, my duty was to evaluate the stock assessments of the two species with respect to their TORs, which are attached in Appendix 2. This report provided the findings of the independent review that was undertaken by me in accordance with the CIE Statement of Work (SOW).

2. ROLE OF INDIVIDUAL REVIEWER IN THE REVIEW ACTIVITIES

My role as a CIE independent reviewer was to conduct an impartial and independent peer review in accordance with the SOW and the predefined TORs herein.

About two weeks before the meeting, assessment documents and supporting materials were made available to the review panel via email and an http website by Dr. James Weinberg. I read all the documents that I received prior to the review.

The SARC 66 meeting followed the "tentative agenda (Appendix 2)" of the CIE review. The meeting was open and was organized constructively. On the morning of Nov 27 before the meeting, the assessment review committee met with Dr. Weinberg to discuss the meeting agenda and SARC process, reporting requirements, and meeting logistics. During the meeting, all the documents were accessible online.

Presentations were given during the review according to the agenda to provide the reviewer panel the background information on the data used in the stock assessment models, the model development, BRPs and the fishery population status. I was actively involved in the discussion during the presentations by 1) listening to the presentations carefully, making notes on the points that were not included or not clearly stated in the documents provided prior to the meeting; 2) asking questions for clarification on the data usage and model development; 3) making comments and providing possible alternative solutions to questions arising during the meeting; 4) discussing to reach agreement on each stock assessment TOR with the other review panel members.

After the peer review meeting, I summarized the findings and recommendations according to the predefined TORs. This review report is formatted according to my interpretation of the required format and content described in Appendix 2.

3. SUMMARY OF FINDINGS AND RECOMMENDATIONS FOR TORs

Below I provide the summary of findings of each TOR for each species in which the weaknesses and strengths are described, and conclusions and recommendations in accordance with the TORs.

3.1 Summer Flounder TORs

3.1.1. *Estimate catch from all sources, including landings and discards. Describe the spatial and temporal distribution of landings, discards, and fishing effort. Characterize the uncertainty in these sources of data. Compare previous recreational data to re-estimated Marine Recreational Information Program (MRIP) data (if available).*

This TOR was addressed successfully.

The data collection schemes are appropriate for estimating the quantity and size/age composition of all significant removals due to commercial and recreational fishing. Sampling intensity for length frequency has increased after 1997. The newest BIOSTAT program is used to re-estimate the commercial landings catch-at-age from 1982-2017 time series. Compared with previous commercial landings data used in the stock assessment, only minor changes were observed. Discard sampling intensity was found to be adequate after late the 1990s when the CV of discard estimation decreased. The average discard rate of the commercial fishery is about 20%, but the discard rates in 2016 and 2017 are 22% and 34% separately.

The new stock assessment adopted new re-estimated Marine Recreational Information Program (MRIP) data, which are much higher than the previously estimated recreational catches because the newly estimated effort of the shore-based and private/rental boats are six and three times the previous estimates. Recreational discarding mortality rate of 10%, has been used since Terceiro (1999) and SARC 47 (NEFSC 2013) and has been retained in all subsequent assessments. The live discards increased correspondingly given the re-estimated MRIP.

The spatial and temporal changes in landings and effort were explored based on Vessel Trip Report (VTR) commercial catch, observed catch and VTR recreational catch and their spatial distributions in every 5 years were plotted and compared. The summer founder fishery demonstrated spatial-temporal changes, and it is consistent with state reported landing and studies based on the survey data. The observed north-south and offshore-inshore changes in landings are likely a combination of both distributional changes and the population shrinkage/dispersion given population size changes.

The new stock assessment highlighted the differences in the new re-estimated MRIP data and its influence on stock assessments. The review panel demonstrated concerns on the largely increased recreational catch but did not look into the research on it given the reality that the revised MRIP is official and the re-estimated recreational catch are final numbers. Based on this premise, the review panel considered the fishery data sufficient to provide suitable bases for exploring a range of catch-at-age models to provide credible fishery management advice.

3.1.2. *Present the survey data available, and describe the basis for inclusion or exclusion of those data in the assessment (e.g., indices of relative or absolute abundance, recruitment, state surveys, age-length data, etc.). Investigate the utility of commercial or recreational LPUE as a measure of relative abundance. Characterize the uncertainty and any bias in these sources of data.*

This TOR was completed successfully. Further exploration on this TOR is suggested.

Seventeen indices of stock abundance including age compositions from 1982-2017 are used to calibrate population size and population structure, and they are the NEFSC winter (1992-2007), spring (1982-2008 Albatross survey shown as relative abundance, and 2009-2017 Bigelow surveys shown as absolute abundances), and fall (1982-2008 Albatross survey shown as relative abundance, and 2009-2016 Bigelow surveys shown as absolute abundances), Massachusetts spring and fall (1978-2017), Rhode Island fall and monthly fixed (1981-2017), Connecticut spring and fall (1984-2017), Delaware (1991-2017), New York (1987-2017), New Jersey (1988-2017), VIMS ChesMMAP (2002-2017), and VIMS NEAMAP spring and fall trawl surveys (2007-2017). Another age-aggregated index of URIGSO is used in the ASAP model calibration. The Bigelow indices incorporate trawl efficiency estimates at length from 'sweep-study' experiments and are expressed as absolute abundances. However, the absolute abundance indices from Bigelow spring and fall surveys are treated as indices in the model.

To calibrate recruitment at age 0, seventeen survey indices are included and they are NEFSC MARMAP and ECOMON larval surveys, YOY (Young-Of-the-Year) surveys conducted by the states of Massachusetts, Rhode Island, Connecticut, New York, New Jersey, Delaware, Maryland, Virginia and North Carolina were also used in the model calibration.

Effort was also spent investigating the utility of commercial or recreational LPUE as a measure of relative abundance by the working group of summer flounder. The working group concluded that possible bias (for both catch and effort) precludes the use of fishery CPUE as calibration indices in the model. The review panel did not spend lots of effort in discussing the utility of the fishery dependent LPUE.

Uncertainties of these indices shown as CVs are included in the initialization of the model. Not all the indices are reported with uncertainty measurements. These indices do show variations in trends over time and in age-compositions, which reflect uncertainties of summer flounder distribution variation over time and potential survey uncertainty shown as high CV in some of the indices and in some periods.

In the reported relative abundance indices, both design and model-based methods were used with quite a few of the indices reported as geometric mean with no uncertainty reported. A future study may be needed to consider a relatively organized group of methods to be applied to all the indices, and some of the methods should help to understand the spatial-temporal variations of the population and the environmental factors that may influence its catchability and spatial distribution.

Some of the surveys were conducted during more than one season and were then used as more than simple indices. Using season-specific indices from the same survey may overweight these surveys in reflecting stock trends in their corresponding survey areas. Future analysis may be needed to consider model-based approaches and treat season as a factor in the relative abundance analysis.

3.1.3. *Describe life history characteristics and the stock's spatial distribution (for both juveniles and adults), including any changes over time. Describe factors related to productivity of the stock and any ecosystem factors influencing recruitment. If possible, integrate the results into the stock assessment.*

This TOR was addressed adequately in general although further exploration and documentation on this TOR are suggested.

The ageing uncertainty was explored by comparing ageing from scales and otoliths and by otolith reading between the NCDMF and NEFSC ageing to evaluate consistency, and the study found that ageing uncertainty is low. Beginning in 2009, age sampling has changed from scale only to both scale and otolith and a recent workshop (2014) suggested that otolith ageing is recommended. After 2015, ageing sampling transitioned to otolith completely.

The mean length-at-age and weight-at-age of summer flounder were found to be decreasing in recent decades. At the same time, weight-at-length and Fulton condition factor do not show testable variation in trend, and maturity at age 0 and age 1 decreased. The changes were found not to be because of selectivity and sample size variation. The working group's best explanation is that there are some signals in the environmental data that may be affecting summer flounder growth rates.

From the stock assessment view, it is enough for the model to capture the changes in growth and the working group did a reasonable job in including the observed variation. Extra effort is suggested to explore the potential environmental factors that cause the life history variations in growth and maturity, if possible.

Sex ratio has been approaching 1:1 for larger/older fishes which has been explained as due to the decrease of F in the last decade. Recruitment estimated as age 0 fish has been lower than average for majority of the years after 2011, and the recruitment per spawner ratio (R/SSB) has been lower than pre-1996 in general. The stock assessment team conducted an analysis to explore potential factors influencing recruitment dynamics but did not find any. The low recruitment and low R/SSB ratio suggested that the productivity of the stock is likely low although potential causal ecosystem factors have not been identified so far.

The stock has shown distribution differences over time in both relative abundance and its spatial distribution pattern. A northward shift or a decrease in its southern rage of distribution in VA and MD waters has been observed in recent years. This has been linked to the climate changes.

Future exploration on the changes in productivity is suggested. Analyses can be both from finding mechanisms and recognizing patterns in the future. Ecosystem dynamics can be complicated and the mechanisms of the productivity changes can be difficult to find, but a pattern that is recognized can still be integrated into the stock assessment. Both data mining approaches and time series approaches may be considered in the future (Sun et al. 2009; Munch et al. 2018).

3.1.4. *Estimate annual fishing mortality, recruitment and stock biomass (both total and spawning stock) for the time series, and estimate their uncertainty. Include retrospective analyses (both historical and within-model) to allow a comparison with previous assessment results and projections, and to examine model fit. Examine sensitivity of model results to changes in re-estimated recreational data.*

This TOR was completed successfully. The assessment is adequate to provide a credible basis for management advice.

A set of age structured statistical catch at age models (ASAP) were constructed and used to re-construct the historical population dynamics of summer flounder between 1982 and 2017. The working group recommended base model run is from a 4-fleet (commercial, recreational, commercial discarding, and recreational discarding) model with natural mortality age-varying with mean 0.25. A likelihood estimator was used to estimate parameters with age/length compositions following multinomial distributions and catch and relative abundance indices following lognormal distributions. The ASAP framework allows flexibility in considering measurement errors in the catch, relative abundance indices and age/length compositions. The likelihood estimator allows alternative weighting of different data sources.

The recommended four-fleet ASAP model, bridges with a two-fleet ASAP model previously used in SARC 57 (NEFSC 2013), with no large differences in the results observed after re-estimated recreational catch data were updated. The retrospective error for the new model configuration with the re-estimated recreational data is lower according to the assessment results. The re-estimated recreational data increased the 1982-2017 total catch by an average of 29%. The inflated catch resulted in a higher population biomass but a similar fishing mortality estimate. Alternative model structures have been explored by the working group, and the results seemed similar in population trends but were all largely influenced by the re-estimated recreational data.

3.1.5. *State the existing stock status definitions for "overfished" and "overfishing". Then update or redefine biological reference points (BRPs; point estimates or proxies for B_{MSY}, $B_{THRESHOLD}$, F_{MSY} and MSY) and provide estimates of their uncertainty. If analytic model-based estimates are unavailable, consider recommending alternative measurable proxies for BRPs. Comment on the scientific adequacy of existing BRPs and the "new" (i.e., updated, redefined, or alternative) BRPs.*

This TOR was adequately addressed.

The newly recommended BRPs were similarly based on the approaches used in SARC 57 (NEFSC 2013). The F-based reference point is $F_{35\%}$, a proxy for F_{MSY}; the spawning stock biomass (SSB) is used as the biomass reference point and is estimated through projecting the population size at $F_{35\%}$, a proxy of SSB_{MSY}. $SSB_{THRESHOLD}$ is defined as 1/2 of the SSB_{MSY} proxy. Comparing with the reference points from SARC 57, this assessment recommended that F_{MSY} increase from 0.309 to 0.448 and that SSB_{MSY} decrease from 62,394 to 57,159mt. Pre-recruitment based reference points are often very sensitive to biological and fisheries (selectivity from the model) input. The working group explored all the life history parameters and fishing related parameters which include weight-at-age, maturity-at-age and selectivity-at-age, and found that the changes are due mainly to the effect of decreased mean weight-at-age of older fishes and a dome-shaped average fishery selectivity pattern in recent years.

CVs of F_{MSY} and SSB_{MSY} were both estimated to be 15%. During the workshop there were no further discussions on the uncertainty estimation of both reference points.

One thing that the review panel brought up during the review is the mean recruitment used in the projection both for BRP estimate and for management advice. The long-term average of estimated recruitment from 1982-2017 was used in the projection. The review panel wondered why the

most recent observations/estimations of weight-at-age and selectivity-at-age were used but not for recruitment, especially because the recruitment level after 2011 has been lower than previous years. Long-term recruitment average has been suggested to be used in deriving reference points and population projection in the past and in SARC 66. The low recruitment pattern persisted after the last SARC review (SARC 57). For information of BRPs, long-term biological and fisheries data seem to be more reasonable, while short-term biological and fisheries data may serve as more reasonable indicators of fisheries. Both review panel and the assessment team would stressed that when management advice is provided, the managers should consider the potential low recruitment in the recent years.

3.1.6. *Make a recommendation[a] about what stock status appears to be, based on the existing model (i.e., model from previous peer reviewed accepted assessment) and with respect to a new modeling approach(-es) developed for this peer review.*
[a]NOAA Fisheries has final responsibility for making the stock status determination for this stock based on best available scientific information.

   a. *Update the existing model with new data and make a stock status recommendation (about overfished and overfishing) with respect to the existing BRP estimates.*

   This TOR was completed successfully.

   Based on the updated existing model and the new data including the re-estimated recreational catch, the summer flounder stock was not overfished, and overfishing was not occurring in 2017 relative to the existing BRPs.

   b. *Then use the newly proposed modeling approach(-es) and make a stock status recommendation with respect to "new" BRPs and their estimates (from TOR-5).*

   This TOR was completed successfully.

   Based on the new model framework and the new data including the re-estimated recreational catch, the summer flounder stock was not overfished, and overfishing was not occurring in 2017 relative to the newly estimated BRPs.

   The long-term recruitment was used for the SSB_{MSY} estimation and population projection. The population has been fished at a level lower than F_{MSY} since 2001 but SSB decreased below SSB_{MSY} in the recent 5-6 years. This is largely because of the low recruitment in the recent seven years. Although there are no significant factors found to

cause the low recruitment, future management may need to consider the uncertainty from the recruitment dynamics, especially if it is shown as a regime shift pattern.

c. *Include descriptions of stock status based on simple indicators/metrics (e.g., age- and size-structure, temporal trends in population size or recruitment indices, etc).*

This TOR was addressed successfully.

Besides comparing fishing mortality and biomass-based reference points, the working group compared trends in age structure, survey abundance indices, weight-at-age, condition factor and sex ratios. Mixed signals were observed: Age structure is expanding and sex ratios are approaching 50:50 in recent years; there is no obvious trends in condition factor; survey abundance indices in its southern range of its distribution were decreasing; recruitment indices have been below average in general in recent years. These mixed signals from population trends, age structure, and biological characteristics indicate a likely low productivity regime.

3.1.7. *Develop approaches and apply them to conduct stock projections.*
a. *Provide numerical annual projections (5 years) and the statistical distribution (i.e., probability density function) of the catch at F_{MSY} or an F_{MSY} proxy (i.e. the overfishing level, OFL) (see Appendix to the SAW TORs). Each projection should estimate and report annual probabilities of exceeding threshold BRPs for F, and probabilities of falling below threshold BRPs for biomass. Use a sensitivity analysis approach in which a range of assumptions about the most important uncertainties in the assessment are considered (e.g., terminal year abundance, variability in recruitment).*

This TOR was accomplished successfully.

The working group conducted stock projections by stochastically projecting the stock size from its value estimated for 2017 with 100% of the 2018 ABC taken as the catch of 2018, and long-term average of recruitment, to the period 2019-2023. The other parameters used in the projection followed the estimates from the recommended assessment model except the selectivity-at-age, discarding, maturity-at-age, and weight-at-age, for which the most recent 5-year averages were used, to reflect the biological and fishing behaviour changes. The example OFL projections used F_{2019}-F_{2023} = fishing mortality threshold F_{MSY} proxy = $F_{35\%}$ and sampled from the estimated recruitment for 1982-2017. The provided projections showed zero

probabilities that $F > F_{MSY}$ and $SSB < 1/2\ SSB_{MSY}$ given the current management control rule.

Alternative sensitivity analyses were done and projections that used a short-term average recruitment were also provided. Some of the sensitivity analyses, such as projection from model uncertainty, were discussed but no detailed results were included.

Although the estimated uncertainty of BRPs, F and SSB are not used in the current management control rule, conducting a full probabilistic based assessment is still suggested for the future. Correlations among parameters should be fully considered in the projection, such as correlations among F, SSB, Recruitment and selectivity. The correlation among parameters influences the estimated BRPs and the projected population size both in scale and in the uncertainty level. It is credible to use the estimated uncertainty of the BRPs when model selection uncertainty is not very high.

b. *Comment on which projections seem most realistic. Consider the major uncertainties in the assessment as well as sensitivity of the projections to various assumptions. Identify reasonable projection parameters (recruitment, weight-at-age, retrospective adjustments, etc.) to use when setting specifications.*

This TOR was addressed adequately.

The working group recommended projection used the long-term recruitment average in the process. The review panel was concerned about the potential effect of it when making recommendation for OFL. This is also considered the major uncertainty in the projection analysis. However, the review panel also realized that there are zero probabilities that $F > F_{MSY}$ and $SSB < 1/2\ SSB_{MSY}$ when both projection and reference points using long-term recruitment and ABC is recommended based on the current control rule. In the future, an estimation of the $P(F > F_{MSY})$ and $P(SSB < 1/2\ SSB_{MSY})$ when using OFL projection based on long-term recruitment average but comparing with reference points estimated based on short term recruitment average, should provide estimates of risks when mis-specifying the productivity regimes.

c. *Describe this stock's vulnerability (see "Appendix to the SAW TORs") to becoming overfished, and how this could affect the choice of ABC.*

This TOR was addressed adequately.

The estimated $P(F > F_{MSY})$ and $P(SSB < 1/2\ SSB_{MSY})$ are both 0 when using F_{MSY} to project the OFL catches. Summer flounder mature early and are relatively long lived. The low recruitment pattern in recent years and the current stock size (lower than SSB_{MSY}) bring concerns on the short term likelihood of overfishing. Overall, both the review panel and the assessment team expressed concern that the potential for low recruitment in the following years be considered in the recommendation of ABC. In conclusion, based on the projection results and the life history of summer flounder, the stock has a low vulnerability of becoming overfished given the management control rule currently applied.

3.1.8. *Review, evaluate and report on the status of the SARC and Working Group research recommendations listed in most recent SARC reviewed assessment and review panel reports and MAFMC SSC reports. Identify new research recommendations.*

This TOR was addressed adequately.

The working group provided their comments and efforts on the many previous research recommendations. The stock assessment report combines all research recommendations in a single section. Major data and analytical needs for future assessments have been identified in the SAW 57 assessment and SARC peer review (NEFSC 2013), the SSC reports from MAFMC during 2013-2018, and by the working group for this current assessment. The working group addressed each recommendation as completed (between the last benchmark and the current assessment), ongoing or in progress at present or to be addressed in the future (previously identified). The review panel and the working group also identified new recommendations during the review period

The working group organized three workshops to evaluate alternatives models including size structure and/or sex specific models, age structured models with alternative fleets to address discards, which have been recommended in previous SARC review or SSC reports. There were no detailed model assessment results and comparison provided. However, the team leaders of these alternative models, who are also part of the working group, indicated that there are no obvious differences in the predicted stock trend and overall conclusions on fishery status and stock status. Future systematic study on model selection and performance comparison is suggested which may improve the value of the fishery and increase transparency in model recommendation.

Although model comparison and selection are not immediately necessary, it is desirable to use a systematic framework and set of metrics to select most appropriate model configuration from a series of model structures or model

runs. Measurements in selecting models can be model goodness-of-fit, model prediction ability, model robustness and fisheries-specific measurements such as retrospective error (Jiao et al. 2012; Gelman et al. 2014; Hooten and Hobbs 2015).

3.2 Striped bass TORs

3.2.1. *Investigate all fisheries independent and dependent data sets, including life history, indices of abundance, and tagging data. Discuss strengths and weaknesses of the data sources.*

This TOR was addressed adequately.

The Atlantic Striped Bass Technical Committee (TC) reviewed all the surveys on relative indices and biological data collection used in the previous benchmark stock assessment (SAW 57) as well as several new surveys. The TC used a set of evaluation criteria to determine which indices should be considered for inclusion in the assessment. The recommended model included 13 abundance indices. These indices were used to represent either one of the two stocks in a two-stock model or used to represent the overall stock trend in a one-stock assessment model.

The evaluation criteria and detailed data synthesis were not provided in both the assessment report and during the review. Considering the large amount of surveys included, no detailed discussion occurred on this topic during the review. In the future, a well-documented data inclusion/exclusion procedure is suggested to be included if possible.

The assessment team described tagging data and the natural mortality estimates derived from tag analysis, which were then directly used in the assessment model. However, uncertainty from the tag-data derived estimates, both natural mortality and stock composition, was not provided. The tagging programs and data on striped bass are pretty extensive, which is unusual compared with other species along the Atlantic coast. These tagging data have lots of potential uses in the stock assessment. A simulation study may be conducted to explore the uncertainty of the parameter estimates and evaluate the efficiency of the current sampling efforts. This kind of simulation study could also help with future sampling design and research need justification.

3.2.2. *Estimate commercial and recreational landings and discards. Characterize the uncertainty in the data and spatial distribution of the fisheries. Review new MRIP estimates of catch, effort and the calibration method, if available.*

This TOR was accomplished successfully.

The assessment team provided a newly developed two-stock model and a one-stock model as used in the SAW 2013. Commercial and recreational landing and discards were assembled and organized based on their spatial distribution and the model structure. Commercial discarding and recreational discarding were all combined into total catch numbers at age in the assessments.

The uncertainty from the MRIP estimates were provided but not for the commercial catch. The discarding composition used the catch composition as a proxy, which may be improved in the future. The re-calibrated MRIP estimates are much higher (in average 140-160% higher) than before. The review and the assessment team did not review and discuss the new MRIP estimates and calibration methods since these data have been treated as the official and final numbers.

The review panel regarded the combined catch data as representing the best current estimates of catch and appropriate for the assessment purposes.

3.2.3. *Use an age-based model to estimate annual fishing mortality, recruitment, total abundance and stock biomass (total and spawning stock) for the time series and estimate their uncertainty. Provide retrospective analysis of the model results and historical retrospective. Provide estimates of exploitation by stock component and sex, where possible, and for total stock complex.*

This TOR was accomplished successfully. The review panel and the assessment team recommended a statistical catch-at-age model with similar structure as used in the SAW 57 (2013). The model used region specific fleets (Chesapeake Bay, and Hudson and Delaware combined) and assumed one stock complex.

The assessment team first recommended a newly developed two spatial stock and three period statistical catch-at-age model (2SCA). The two stocks are Chesapeake Bay stock and the Hudson and Delaware combined stock. The model was not recommended to be used for management purposes for the following reasons: i) The initial year (1982) population structure assumption is questionable, ii) the F penalty used in earlier years is very strong, iii) the fit to the stock composition index is problematic, and iv) the tagging data after 1996 were not used because of lacking ageing information. The model construction seemed to cause the lack fit for the data in the earlier years, which might also bias the relationship between abundance indices and population size over time. The working group also indicated that the model is sensitive to the estimated stock composition and its uncertainty is not estimated but likely high. The model performance was also demonstrated to be problematic as the resulting fishing mortality, recruitment and SSB estimated for the two stocks are parallel over time.

The review panel and the assessment team then discussed the one stock one fleet statistical catch-at-age model, which is similar in construction with the SAW 2013 recommended model. The assessment results seemed heavily influenced by the new MRIP estimates and less sensitive to other data sources. There were no obvious retrospective patterns in either the F or female SSB estimates. The review panel agreed that this model was appropriate for assessing the fishery and population status, and is appropriate for management purposes.

The review panel recommended that future effort on the two spatial stock model is encouraged. The newly developed 2SCA model may integrate tagging data together directly instead of using the tag estimated mortality and movements. The tag data can also provide abundance estimates, which is another benefit of integrating them into the SCA models in the future.

3.2.4. *Use tagging data to estimate mortality and abundance, and provide suggestions for further development.*

This TOR is addressed adequately.

The assessment selected an instantaneous rate catch release (IRCR) model originally developed by Jiang et al. (2007) and recoded in ADMB. The analyses were done based on data from tagging area/agency. A likelihood estimator was used to estimate parameters. Six models with alternative assumptions on F and M given its variation across regulation periods were compared using AIC. The assessment team also compared the IRCR estimated F and N with those estimated from 2SCA and they were comparable.

The assessment team also conducted miscellaneous sensitivity analyses to explore the influence of model assumptions. The models were not sensitive to tag loss and tagging induced mortality, but F and M estimates were sensitive to the reporting rate specification.

The tagging models were built based on the tagging area and agency. Bayesian hierarchical models are suggested to be considered in the future to integrate tagging programs and data together to improve the robustness and stability of the parameter estimation (Gelman et al. 2014).

The tagging data have high potential to be integrated in the SCA model, so that estimation uncertainty on the key parameters can be integrated, and they can be validated at the same time as the catch and relative abundance indices data. A stochastic age-length relationship can be used to infer critical rate of fishes not aged.

3.2.5. *Update or redefine biological reference points (BRPs; point estimates or proxies for BMSY, SSBMSY, FMSY, MSY) for each stock component where possible and for the total stock complex. Make a stock status determination based on BRPs by stock component, where possible, and for the total stock complex.*

This TOR was generally achieved.

The assessment team did not estimate MSY related BRPs because the management of this species does not need to follow Magnuson–Stevens Fishery Conservation and Management Act. Instead, the assessment team continued to use the previously recommended relative population level in 1995, i.e., SSB_{1995} as the $SSB_{THRESHOLD}$ and $SSB_{TARGET} = 125\%$ SSB_{1995}. $F_{THRESHOLD}$ and F_{TARGET} were then estimated by projecting a series of F values until the $SSB_{THRESHOLD}$ and SSB_{TARGET} were achieved. Population size in 1995 has been treated as being a level that is relatively healthy and can provide long term sustainable yield by the working group.

At the same time, two stock-recruitment relationships were explored, and they are hockey-stick BHSR (Beverton-Holt Stock Recruitment) and empirical distribution with R from estimates between 1990-2017. The assessment team also compared the updated BRPs with F20%, F30% and F40%.

There was an extensive discussion on the spatial stock management needs and their reflection in both model development and in BRP estimation. Although the review panel understands the strong need and interest in spatial-specific fisheries management, the review panel recommends using the one stock assessment model and BRP for management purposes in the next few years, with continued research and development of the spatial-stock assessment and BRP at the same time.

According to the recommended model and scenario, the assessed stock was overfished and overfishing was happening in 2017.

3.2.6. *Provide annual projections of catch and biomass under alternative harvest scenarios. Projections should estimate and report annual probabilities of exceeding threshold BRPs for F and probabilities of falling below threshold BRPs for biomass.*

This TOR was accomplished successfully.

A range of projections were conducted with alternative F levels including the current (2017) F level, F_{TARGET} and $F_{THRESHOLD}$. The projections did estimate $P(F > F_{THRESHOLD})$ and $P(SSB < SSB_{THRESHOLD})$.

There were discussions on the selection of stock-recruitment models used in both the BRPs and in the projection analysis. The selection of the recruitment model for projection is suggested to be consistent with the choice of model used for BRP estimation. Also, the short term projected $P(F > F_{THRESHOLD})$ and $P(SSB < SSB_{THRESHOLD})$ tend to be robust to the selection of the stock-recruitment model given the current minimum catchable sizes and the estimated selectivity pattern.

3.2.7. *Review and evaluate the status of the Technical Committee research recommendations listed in the most recent SARC report. Identify new research recommendations. Recommend timing and frequency of future assessment updates and benchmark assessments.*

This TOR was addressed adequately.

The working group provided the list of the research recommendations based on four areas: Fishery-dependent priorities, fishery-independent priorities, modeling/quantitative priorities and life history and biology, and three levels of priorities: high, moderate and low. The working group also clearly identified the recommendations that have been addressed or were well in progress.

The review panel further suggested several new research recommendations, which include: Further examination of tagging data after 1995 (including developing ways of assigning ages to NY data) to look at potential time-varying emigration rates; continued work on the 2-stock SCA model and further exploration of appropriate estimation approaches for BRPs; investigate the ability of the model to detect changes in stock status with different emigration rates/exploitation patterns through a simulation study; estimate uncertainty of the stock composition and evaluate the influence of it on the stock-specific fishing and population status evaluation and BRP derivation.

4.   Acknowledgements

5. References

Gelman, A., Huang, J., and Vehtari, A. 2014. Understanding predictive information criteria for Bayesian models. Statistics and Computing. 24:997-1016.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. 2014. Bayesian Data Analysis. 3rd edition. Chapman & Hall, London, UK. 667 pp.

Hooten, M.B., and Hobbs, N.T. 2015. A guide to Bayesian model selection for ecologists. Ecological Monographs. 85: 3-28.

Jardim, E, Eero, M., Silva, A., Ulrich, C., Pawlowski, L., et al. 2018. Testing spatial heterogeneity with stock assessment models. PLOS ONE 13(1): e0190791. https://doi.org/10.1371/journal.pone.0190791

Jiang et al. 2007. Tag Return Models Allowing for Harvest and Catch and Release: Evidence of Environmental and Management Impacts on Striped Bass Fishing and Natural Mortality Rates. 10 p.

Jiao, Y., Smith, E., O'Reilly, R., and Orth, D. 2012. Modeling nonstationary natural mortality in catch-at-age models: an example using the Atlantic weakfish (*Cynoscion regalis*) fishery. ICES Journal of Marine Science. 69:105-118.

Jiao, Y., O'Reilly, R., Smith, E., and Orth, D. 2016. Integrating spatial synchrony/asynchrony of population distribution into stock assessment models: a spatial hierarchical Bayesian statistical catch-at-age approach. ICES Journal of Marine Science. 73: 1725-1738.

Legault, C.M., Restrepo, V.R. 1998. A flexible forward age-structured assessment program. ICCAT. Col. Vol. Sci. Pap. 49:246-253.

Munch, S.B., Giron-Nava, A., and Sugihara, G. 2018. Nonlinear dynamics and noise in fisheries recruitment: A global meta-analysis. Fish and Fisheries. 19: 964-973

NOAA Fisheries Toolbox (NFT). 2013a. Age Structured Assessment Program (ASAP) version 3.0.11. (Internet address: http://nft.nefsc.noaa.gov).

Northeast Fisheries Science Center. 2013. 57th Northeast Regional Stock Assessment Workshop (57th SAW) Assessment Report. US Department of Commerce, Northeast Fisheries Science Center Reference Doc. 13-16; 967 p. Available at http://nefsc.noaa.gov/publications/

Sun, L., Xiao, H., Yang, D., and Li, D. 2009. Forecasting fish stock recruitment by using neural network. *2009 WRI Global Congress on Intelligent Systems*, pp. 96-100.

Terceiro M. 2012. Stock Assessment of Summer Flounder for 1999. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 99-19; 178 p.

Appendix 1: Bibliography of materials provided for review

SARC Working Papers:

Working Group, Stock Assessment Workshop (SAW 66) 2018. Stock Assessment Report for Summer Flounder. Working Paper A1 SAW/SARC 66. November 27-30, 2018, NOAA Fisheries, Northeast Fisheries Science Center. Woods Hole, MA.

Working Group, Stock Assessment Workshop (SAW 66) 2018. Stock Assessment Report for Striped Bass. Working Paper B1. SAW/SARC66. November 27-30, 2018, NOAA Fisheries, Northeast Fisheries Science Center. Woods Hole, MA.

Working Group, Stock Assessment Workshop (SAW 66) 2018. Stock Assessment Summary Report for Summer Flounder. Working Paper A2. SAW/SARC 66, November 27-30, 2018, NOAA Fisheries, Northeast Fisheries Science Center, Woods Hole, MA.

Working Group, Stock Assessment Workshop (SAW 66) 2018. Stock Assessment Summary Report for Striped Bass. Working Paper B2. SAW/SARC 66, November 27-30, 2018, NOAA Fisheries, Northeast Fisheries Science Center, Woods Hole, MA.

Background material provided for summer flounder SARC review:

Deen S, Jauss V, Sullivan PJ. 2018 MS. The effect of ocean environmental conditions on the relative abundance of summer flounder (*Paralichthys dentatus*): spatio-temporal analysis and model comparison using R-INLA. Cornell University, Ithaca NY. SAW/SARC 66 Summer Flounder Working Group Working Paper. 46 p.

Gervelis BJ. 2018 MS. Summer flounder CPUE derived from Cooperative Research Study Fleet self-reported data. NOAA Fisheries NEFSC Cooperative research Branch Narragansett RI. SAW/SARC 66 Summer Flounder Working Group Working Paper. 9 p.

Langan JA, McManus MC, Schonfeld AJ, Truesdale CL, Collie JS. 2018 MS. Evaluating summer flounder (*Paralichthys dentatus*) spatial sex-segregation in a southern New England estuary. University of Rhode Island Graduate School of Oceanography. SAW/SARC 66 Summer Flounder Working Group Working Paper. 20 p.

Maunder MN. 2018a MS. Stock synthesis implementation of a sex-structured virtual population analysis applied to summer flounder. SAW/SARC 66 Summer Flounder Working Group Working Paper. 6 p.

Maunder MN. 2018b MS. Dynamic reference points for summer flounder. SAW/SARC 66 Summer Flounder Working Group Working Paper 12 p.

McNamee J. 2018 MS. Investigating alternative abundance indices for the summer flounder assessment. SAW/SARC 66 Summer Flounder Working Group Working Paper. 45 p.

Miller A, Terceiro M. 2018a MS. Spatial distribution of summer flounder captured in the commercial and recreational fisheries. SAW/SARC 66 Summer Flounder Working Group Working Paper. 18 p.

Miller A, Terceiro M. 2018b MS. Spatial distribution of summer flounder sampled by the NEFSC trawl survey. SAW/SARC 66 Summer Flounder Working Group Working Paper. 87 p.

Miller TJ, Terceiro M. 2018a MS. A state-space, sex-specific, age-structured assessment model for summer flounder. SAW/SARC 66 Summer Flounder Working Group Working Paper. 25 p.

Miller TJ, Terceiro M. 2018b MS. Even more state-space, sex-specific, age-structured assessment models for summer flounder. SAW/SARC 66 Summer Flounder Working Group Working Paper. 77 p.

Northeast Fisheries Science Center. 2013. 57th Northeast Regional Stock Assessment Workshop (57th SAW) Assessment Report. US Department of Commerce, Northeast Fisheries Science Center Reference Doc. 13-16; 967 p. Available at http://nefsc.noaa.gov/publications/

Perretti CT. 2018 MS. An analysis of summer flounder (*Paralichthys dentatus*) distribution on the Northeast U.S. shelf using a spatio-temporal model. NOAA NMFS NEFSC. SAW/SARC 66 Summer Flounder Working Group Working Paper. 47 p.

Sullivan PJ. 2018 MS. A sex-age-length fisheries stock assessment model with analysis and application to summer flounder (*Paralichthys dentatus*). SAW/SARC 66 Summer Flounder Working Group Working Paper. 38 p.

Terceiro M. 2016. Stock Assessment of Summer Flounder for 2016. US Department of Commerce, Northeast Fisheries Science Center Reference Doc. 16-15; 117 p. Available at http://nefsc.noaa.gov/publications

Terceiro M. 2017 MS. The summer flounder ASAP statistical catch at age model by sex. SAW/SARC 66 Summer Flounder Working Group Working Paper. 19 p.

***Background material provided for* striped bass SARC review*:**

ASMFC. 2003. Amendment #6 to the Interstate Fishery Management Plan for Atlantic Striped Bass. Washington (DC): ASMFC. Fisheries Management Report No. 41. 63 p.

Dorazio RM, Hattala KA, McCollough CB, Skjeveland JE. 1994. Tag recovery estimates of migration of striped bass from spawning areas of the Chesapeake Bay. Transactions of the American Fisheries Society 123:950–963.

Jiang H, Pollock KH, Brownie C, Hoenig JM, Latour RJ, Wells BK, Hightower JE. 2007. Tag Return Models Allowing for Harvest and Catch and Release: Evidence of Environmental and Management Impacts on Striped Bass Fishing and Natural Mortality Rates. North American Journal of Fisheries Management 27:387–396.

Kneebone, J., WS Hoffman, MJ Dean, DA Fox and MP. Armstrong. 2014. Movement Patterns and Stock Composition of Adult Striped Bass Tagged in

Massachusetts Coastal Waters, Transactions of the American Fisheries Society 143: 1115-1129.

Northeast Fisheries Science Center. 2013a. 57th Northeast Regional Stock Assessment Workshop (57th SAW) Assessment Summary Report. US Department of Commerce, Northeast Fisheries Science Center Reference Doc. 13-14; 39 p.

Northeast Fisheries Science Center. 2013b. 57th Northeast Regional Stock Assessment Workshop (57th SAW) Assessment Report Part B. Striped Bass Stock Assessment for 2013, Updated through 2011. US Department of Commerce, Northeast Fisheries Science Center Reference Doc. 13-16; 290 p.

Northeast Fisheries Science Center. 2013c. 57th Northeast Regional Stock Assessment Workshop (57th SAW) Assessment Report Part B. Appendices. US Department of Commerce, Northeast Fisheries Science Center Reference Doc. 13-16; 186 p.

Northeast Fisheries Science Center. 2013d. Summary Report of the 57th Northeast Regional Stock Assessment Review Committee (SARC 57). 47 p.

Shepherd GR, Nelson GA, Rago PJ, Richards RA, Boreman J, Goodyear CP. 2018. A Chronicle of Striped Bass Population Restoration and Conservation in the Northwest Atlantic, 1979-2016. NOAA Tech Memo NMFS NE-246; 51 p.

Additional Documents:

Cover letter for the "Public testimony for the summer flounder independent expert review". Dated: Nov. 15, 2018. From: Save the Summer Flounder Fishery Fund, Greg Hueth, Chairman.

"Public testimony for the summer flounder independent expert review". From: Save the Summer Flounder Fishery Fund. Nov. 2018.

Appendix 2: Statement of Work
Performance Work Statement (PWS)
National Oceanic and Atmospheric Administration (NOAA)
National Marine Fisheries Service (NMFS)
Center for Independent Experts (CIE) Program
External Independent Peer Review

66th Stock Assessment Workshop/Stock Assessment Review Committee (SAW/SARC)
Benchmark stock assessment for Summer flounder and Striped bass

Background

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

Scientific peer review is defined as the organized review process where one or more qualified experts review scientific information to ensure quality and credibility. These expert(s) must conduct their peer review impartially, objectively, and without conflicts of interest. Each reviewer must also be independent from the development of the science, without influence from any position that the agency or constituent groups may have. Furthermore, the Office of Management and Budget (OMB), authorized by the Information Quality Act, requires all federal agencies to conduct peer reviews of highly influential and controversial science before dissemination, and that peer reviewers must be deemed qualified based on the OMB Peer Review Bulletin standards[1]. Further information on the Center for Independent Experts (CIE) program may be obtained from www.ciereviews.org.

Scope

The Northeast Regional Stock Assessment Review Committee (SARC) meeting is a formal, multiple-day meeting of stock assessment experts who serve as a panel to peer-review tabled stock assessments and models. The SARC peer review is the cornerstone of the Northeast Stock Assessment Workshop (SAW) process, which includes assessment development, and report preparation (which is done by SAW Working Groups or Atlantic States Marine Fisheries Commission (ASMFC) technical committees), assessment peer review (by the SARC), public presentations, and document publication. This review determines whether or not the scientific assessments are adequate to serve as a basis for

---

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

developing fishery management advice. Results provide the scientific basis for fisheries within the jurisdiction of NOAA's Greater Atlantic Regional Fisheries Office (GARFO).

The purpose of this meeting will be to provide an external peer review of a benchmark stock assessment **Summer flounder and Striped bass**. The requirements for the peer review follow. This Statement of Work (PWS) also includes: **Appendix 1**: TORs for the stock assessment, which are the responsibility of the analysts; **Appendix 2:** a draft meeting agenda; **Appendix 3:** Individual Independent Review Report Requirements; and **Appendix 4:** SARC Summary Report Requirements.

Requirements

NMFS requires three reviewers under this contract (i.e. subject to CIE standards for reviewers) to participate in the panel review. The SARC chair, who is in addition to the three reviewers, will be provided by either the New England or Mid-Atlantic Fishery Management Council's Science and Statistical Committee; although the SARC chair will be participating in this review, the chair's participation (i.e. labor and travel) is not covered by this contract.

Each reviewer will write an individual review report in accordance with the PWS, OMB Guidelines, and the TORs below. All TORs must be addressed in each reviewer's report. No more than one of the reviewers selected for this review is permitted to have served on a SARC panel that reviewed this same species in the past. The reviewers shall have working knowledge and recent experience in the application of modern fishery stock assessment models. Expertise should include forward projecting statistical catch-at-age (SCAA) models. Reviewers should also have experience in evaluating measures of model fit, identification, uncertainty, and forecasting. Reviewers should have experience in development of Biological Reference Points (BRPs) that includes an appreciation for the varying quality and quantity of data available to support estimation of BRPs. For summer flounder, knowledge of flatfish biology and population dynamics would be useful. For striped bass, knowledge of anadromous species and SCAA models with spatial considerations would be useful.

Tasks for Reviewers
- Review the background materials and reports prior to the review meeting
- Attend and participate in the panel review meeting
    - The meeting will consist of presentations by NOAA and other scientists, stock assessment authors and others to facilitate the review, to provide any additional information required by the reviewers, and to answer any questions from reviewers
- Reviewers shall conduct an independent peer review in accordance with the requirements specified in this PWS and TORs, in adherence with the required formatting and content guidelines; reviewers are not required to reach a consensus.
- Each reviewer shall assist the SARC Chair with contributions to the SARC Summary Report

- Deliver individual Independent Review Reports to the Government according to the specified milestone dates
- This report should explain whether each stock assessment Term of Reference of the SAW was or was not completed successfully during the SARC meeting, using the criteria specified below in the "Tasks for SARC panel."
- If any existing Biological Reference Points (BRP) or their proxies are considered inappropriate, the Independent Report should include recommendations and justification for suitable alternatives. If such alternatives cannot be identified, then the report should indicate that the existing BRPs are the best available at this time.
- During the meeting, additional questions that were not in the Terms of Reference but that are directly related to the assessments may be raised. Comments on these questions should be included in a separate section at the end of the Independent Report produced by each reviewer.
- The Independent Report can also be used to provide greater detail than the SARC Summary Report on specific stock assessment Terms of Reference or on additional questions raised during the meeting.

Tasks for SARC panel
- During the SARC meeting, the panel is to determine whether each stock assessment Term of Reference (TOR) of the SAW was or was not completed successfully. To make this determination, panelists should consider whether the work provides a scientifically credible basis for developing fishery management advice. Criteria to consider include: whether the data were adequate and used properly, the analyses and models were carried out correctly, and the conclusions are correct/reasonable. If alternative assessment models and model assumptions are presented, evaluate their strengths and weaknesses and then recommend which, if any, scientific approach should be adopted. Where possible, the SARC chair shall identify or facilitate agreement among the reviewers for each stock assessment TOR of the SAW.
- If the panel rejects any of the current BRP or BRP proxies (for B_{MSY} and F_{MSY} and MSY), the panel should explain why those particular BRPs or proxies are not suitable, <u>and</u> the panel should recommend suitable alternatives. If such alternatives cannot be identified, then the panel should indicate that the existing BRPs or BRP proxies are the best available at this time.
- Each reviewer shall complete the tasks in accordance with the PWS and Schedule of Milestones and Deliverables below.

Tasks for SARC chair and reviewers combined:
Review both the Assessment Report and the draft Assessment Summary Report. The draft Assessment Summary Report is reviewed and edited to assure that it is consistent with the outcome of the peer review, particularly statements about stock status recommendations and descriptions of assessment uncertainty.

The SARC Chair, with the assistance from the reviewers, will write the SARC Summary Report. Each reviewer and the chair will discuss whether they hold similar views on each stock assessment Term of Reference and whether their opinions can be summarized into a single conclusion for all or only for some of the Terms of Reference of the SAW. For terms where a similar view can be reached, the SARC Summary Report will contain a summary of such opinions. In cases where multiple and/or differing views exist on a given Term of Reference, the SARC Summary Report will note that there is no agreement and will specify - in a summary manner – what the different opinions are and the reason(s) for the difference in opinions.

The chair's objective during this SARC Summary Report development process will be to identify or facilitate the finding of an agreement rather than forcing the panel to reach an agreement. The chair will take the lead in editing and completing this report. The chair may express the chair's opinion on each Term of Reference of the SAW, either as part of the group opinion, or as a separate minority opinion. The SARC Summary Report will not be submitted, reviewed, or approved by the Contractor.

If any existing Biological Reference Points (BRP) or BRP proxies are considered inappropriate, the SARC Summary Report should include recommendations and justification for suitable alternatives. If such alternatives cannot be identified, then the report should indicate that the existing BRP proxies are the best available at this time.

Foreign National Security Clearance
When reviewers participate during a panel review meeting at a government facility, the NMFS Project Contact is responsible for obtaining the Foreign National Security Clearance approval for reviewers who are non-US citizens. For this reason, the reviewers shall provide requested information (e.g., first and last name, contact information, gender, birth date, country of birth, country of citizenship, country of permanent residence, country of current residence, dual citizenship (yes, no), passport number, country of passport, travel dates.) to the NEFSC SAW Chair for the purpose of their security clearance, and this information shall be submitted at least 30 days before the peer review in accordance with the NOAA Deemed Export Technology Control Program NAO 207-12 regulations available at the Deemed Exports NAO website:
http://deemedexports.noaa.gov/ and
http://deemedexports.noaa.gov/compliance_access_control_procedures/noaa-foreign-national-registration-system.html. The contractor is required to use all appropriate methods to safeguard Personally Identifiable Information (PII).

Place of Performance
The place of performance shall be at the contractor's facilities, and at the Northeast Fisheries Science Center in Woods Hole, Massachusetts.

Period of Performance

The period of performance shall be from the time of award through January 31, 2019. Each reviewer's duties shall not exceed **16** days to complete all required tasks.

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

No later than Oct. 26, 2018	Contractor selects and confirms reviewers
No later than Nov. 13, 2018	NMFS Project Contact will provide reviewers the pre-review documents
Nov. 27-30, 2018	Each reviewer participates and conducts an independent peer review during the panel review meeting in Woods Hole, MA
Nov. 30, 2018	SARC Chair and reviewers work at drafting reports during meeting at Woods Hole, MA, USA
Dec. 14, 2018	Reviewers submit draft independent peer review reports to the contractor's technical team for review
Dec. 14, 2018	Draft of SARC Summary Report, reviewed by all reviewers, due to the SARC Chair *
Dec. 21, 2018	SARC Chair sends Final SARC Summary Report, approved by reviewers, to NMFS Project contact (i.e., SAW Chairman)
Jan. 2, 2019	Contractor submits independent peer review reports to Government
Jan. 9, 2019	The COR and/or technical POC distributes the final reports to the NMFS Project Contact

* The SARC Summary Report will not be submitted to, reviewed, or approved by the Contractor.

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

Travel
All travel expenses shall be reimbursable in accordance with Federal Travel Regulations (http://www.gsa.gov/portal/content/104790). International travel is authorized for this contract. Travel is not to exceed $12,000.

Restricted or Limited Use of Data

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

NMFS Project Contact

Dr. James Weinberg, NEFSC SAW Chair
Northeast Fisheries Science Center
166 Water Street, Woods Hole, MA 02543
James.Weinberg@noaa.gov   Phone: 508-495-2352

Appendix 2.1. Stock Assessment Terms of Reference for SAW/SARC-66

*T*he *SARC Review Panel shall assess whether or not the SAW Working Group has reasonably and satisfactorily completed the following actions.*

The stock assessments for SAW/SARC66 require new calibrated catch and effort data from the Marine Recreational Information Program (MRIP). For these assessments to happen, the assessment scientists need the new MRIP data in a form ready for analysis by July 1, 2018.

A. Summer flounder

1. Estimate catch from all sources, including landings and discards. Describe the spatial and temporal distribution of landings, discards, and fishing effort. Characterize the uncertainty in these sources of data. Compare previous recreational data to re-estimated Marine Recreational Information Program (MRIP) data (if available).

2. Present the survey data available, and describe the basis for inclusion or exclusion of those data in the assessment (e.g., indices of relative or absolute abundance, recruitment, state surveys, age-length data, etc.). Investigate the utility of commercial or recreational LPUE as a measure of relative abundance. Characterize the uncertainty and any bias in these sources of data.

3. Describe life history characteristics and the stock's spatial distribution (for both juveniles and adults), including any changes over time. Describe factors related to productivity of the stock and any ecosystem factors influencing recruitment. If possible, integrate the results into the stock assessment.

4. Estimate annual fishing mortality, recruitment and stock biomass (both total and spawning stock) for the time series, and estimate their uncertainty. Include retrospective analyses (both historical and within-model) to allow a comparison with previous assessment results and projections, and to examine model fit. Examine sensitivity of model results to changes in re-estimated recreational data.

5. State the existing stock status definitions for "overfished" and "overfishing". Then update or redefine biological reference points (BRPs; point estimates or proxies for B_{MSY}, $B_{THRESHOLD}$, F_{MSY} and MSY) and provide estimates of their uncertainty. If analytic model-based estimates are unavailable, consider recommending alternative measurable proxies for BRPs. Comment on the scientific adequacy of existing BRPs and the "new" (i.e., updated, redefined, or alternative) BRPs.

6. Make a recommendation[a] about what stock status appears to be, based on the existing model (i.e., model from previous peer reviewed accepted assessment) and with respect to a new modeling approach(-es) developed for this peer review.

a. Update the existing model with new data and make a stock status recommendation (about overfished and overfishing) with respect to the existing BRP estimates.

b. Then use the newly proposed modeling approach(-es) and make a stock status recommendation with respect to "new" BRPs and their estimates (from TOR-5).

c. Include descriptions of stock status based on simple indicators/metrics (e.g., age- and size-structure, temporal trends in population size or recruitment indices, etc).

7. Develop approaches and apply them to conduct stock projections.

a. Provide numerical annual projections (5 years) and the statistical distribution (i.e., probability density function) of the catch at F_{MSY} or an F_{MSY} proxy (i.e. the overfishing level, OFL) (see Appendix to the SAW TORs). Each projection should estimate and report annual probabilities of exceeding threshold BRPs for F, and probabilities of falling below threshold BRPs for biomass. Use a sensitivity analysis approach in which a range of assumptions about the most important uncertainties in the assessment are considered (e.g., terminal year abundance, variability in recruitment).

b. Comment on which projections seem most realistic. Consider the major uncertainties in the assessment as well as sensitivity of the projections to various assumptions. Identify reasonable projection parameters (recruitment, weight-at-age, retrospective adjustments, etc.) to use when setting specifications.

c. Describe this stock's vulnerability (see "Appendix to the SAW TORs") to becoming overfished, and how this could affect the choice of ABC.

8. Review, evaluate and report on the status of the SARC and Working Group research recommendations listed in most recent SARC reviewed assessment and review panel reports and MAFMC SSC reports. Identify new research recommendations.

[a]NOAA Fisheries has final responsibility for making the stock status determination for this stock based on best available scientific information.

B. Striped bass

1. Investigate all fisheries independent and dependent data sets, including life history, indices of abundance, and tagging data. Discuss strengths and weaknesses of the data sources.

2. Estimate commercial and recreational landings and discards. Characterize the uncertainty in the data and spatial distribution of the fisheries. Review new MRIP estimates of catch, effort and the calibration method, if available.

3. Use an age-based model to estimate annual fishing mortality, recruitment, total abundance and stock biomass (total and spawning stock) for the time series and estimate their uncertainty. Provide retrospective analysis of the model results and historical retrospective. Provide estimates of exploitation by stock component and sex, where possible, and for total stock complex.

4. Use tagging data to estimate mortality and abundance, and provide suggestions for further development.

5. Update or redefine biological reference points (BRPs; point estimates or proxies for BMSY, SSBMSY, FMSY, MSY) for each stock component where possible and for the total stock complex. Make a stock status determination based on BRPs by stock component, where possible, and for the total stock complex.

6. Provide annual projections of catch and biomass under alternative harvest scenarios. Projections should estimate and report annual probabilities of exceeding threshold BRPs for F and probabilities of falling below threshold BRPs for biomass.

7. Review and evaluate the status of the Technical Committee research recommendations listed in the most recent SARC report. Identify new research recommendations. Recommend timing and frequency of future assessment updates and benchmark assessments.

SAW Assessment TORs:

Clarification of Terms
used in the Stock Assessment Terms of Reference

Guidance to SAW Working Group about "Number of Models to include in the Assessment Report":

In general, for any TOR in which one or more models are explored by the Working Group, give a detailed presentation of the "best" model, including inputs, outputs, diagnostics of model adequacy, and sensitivity analyses that evaluate robustness of model results to the assumptions. In less detail, describe other models that were evaluated by the Working Group and explain their strengths, weaknesses and results in relation to the "best" model. If selection of a "best" model is not possible, present alternative models in detail, and summarize the relative utility each model, including a comparison of results. It should be highlighted whether any models represent a minority opinion.

On "Acceptable Biological Catch" (DOC Nat. Stand. Guidelines. Fed. Reg., v. 74, no. 11, 1-16-2009):

Acceptable biological catch (ABC) is a level of a stock or stock complex's annual catch that accounts for the scientific uncertainty in the estimate of Overfishing Limit (OFL) and any other scientific uncertainty…" *(p. 3208) [In other words, OFL ≥ ABC.]*

ABC for overfished stocks. For overfished stocks and stock complexes, a rebuilding ABC must be set to reflect the annual catch that is consistent with the schedule of fishing mortality rates in the rebuilding plan. *(p. 3209)*

NMFS expects that in most cases ABC will be reduced from OFL to reduce the probability that overfishing might occur in a year. (p. 3180)

ABC refers to a level of ''catch'' that is ''acceptable'' given the ''biological'' characteristics of the stock or stock complex. As such, Optimal Yield (OY) does not equate with ABC. The specification of OY is required to consider a variety of factors, including social and economic factors, and the protection of marine ecosystems, which are not part of the ABC concept. (p. 3189)

On "Vulnerability" (DOC Natl. Stand. Guidelines. Fed. Reg., v. 74, no. 11, 1-16-2009):

"Vulnerability. A stock's vulnerability is a combination of its productivity, which depends upon its life history characteristics, and its susceptibility to the fishery. Productivity refers to the capacity of the stock to produce Maximum Sustainable Yield (MSY) and to recover if the population is depleted, and susceptibility is the potential

for the stock to be impacted by the fishery, which includes direct captures, as well as indirect impacts to the fishery (e.g., loss of habitat quality)." (p. 3205)

Participation among members of a Stock Assessment Working Group:

Anyone participating in SAW meetings that will be running or presenting results from an assessment model is expected to supply the source code, a compiled executable, an input file with the proposed configuration, and a detailed model description in advance of the model meeting. Source code for NOAA Toolbox programs is available on request. These measures allow transparency and a fair evaluation of differences that emerge between models.

Appendix 2.2. Draft Review Meeting Agenda

{Final Meeting agenda to be provided at time of award}

66th Stock Assessment Workshop/Stock Assessment Review Committee (SAW/SARC)
Benchmark stock assessment for A. Summer flounder and B. Striped bass

November 27-30, 2018

Stephen H. Clark Conference Room – Northeast Fisheries Science Center
Woods Hole, Massachusetts

DRAFT AGENDA* (version: June 14, 2018)

TOPIC   PRESENTER(S)   SARC LEADER   RAPPORTEUR

Tuesday, Nov. 27

10 – 10:45 AM
  Welcome/Description of Review Process **James Weinberg**, SAW Chair
  Introductions/Agenda   **TBD,** SARC Chair
  Conduct of Meeting

10:45 – 12:45 PM   Assessment Presentation (A. Summer flounder)
                                         Mark Terceiro   TBD

12:45 – 1:45 PM   Lunch

1:45 – 3:45 PM   Assesssment Presentation (A. Summer flounder)
                                         Mark Terceiro   TBD

3:45 – 4 PM   Break

4 – 5:45 PM   SARC Discussion w/ Presenters (A. Summer flounder)
                                         TBD, SARC Chair   **TBD**

5:45 – 6 PM   Public Comments

TOPIC	PRESENTER(S)	SARC LEADER	RAPPORTEUR

Wednesday, Nov. 28

8:30 – 10:30 AM   Assessment Presentation (B. Striped bass)
                                    Katie Drew   **TBD**

10:30 – 10:45 AM   Break

10:45 – 12:30 PM   Assessment Presentation (B. Striped bass)
                                    Katie Drew   **TBD**

12:30 – 1:30 PM   Lunch

1:30 – 3:30 PM   SARC Discussion w/presenters (B. Striped bass)
                                    TBD, SARC Chair   **TBD**

3:30 – 3:45 PM   Public Comments

3:45 -4 PM   Break

4 – 6 PM   Revisit with Presenters (A. Summer flounder)
                                    TBD, SARC Chair   **TBD**

7 PM   **(**Social Gathering**)**

TOPIC	PRESENTER(S)	SARC LEADER	RAPPORTEUR

Thursday, Nov. 29

8:30 – 10:30   Revisit with Presenters (B. Striped bass)
                            TBD, SARC Chair   **TBD**

10:30 – 10:45   Break

10:45 – 12:15   Review/Edit Assessment Summary Report (A. Summer flounder)
                            TBD, SARC Chair   **TBD**

12:15 – 1:15 PM   Lunch

1:15 – 2:45 PM   (cont.) Edit Assessment Summary Report (A. Summer flounder)
                            TBD, SARC Chair   **TBD**

2:45 – 3 PM   Break

3 – 6 PM   Review/edit Assessment Summary Report (B. Striped bass)
                            TBD, SARC Chair   **TBD**

Friday, Nov. 30

9:00 AM – 5:00 PM   SARC Report writing

*All times are approximate, and may be changed at the discretion of the SARC chair. The meeting is open to the public; however, during the Report Writing sessions we ask that the public refrain from engaging in discussion with the SARC.

Appendix 2.3. Individual Independent Peer Review Report Requirements

1. The independent peer review report shall be prefaced with an Executive Summary providing a concise summary of whether they accept or reject the work that they reviewed, with an explanation of their decision (strengths, weaknesses of the analyses, etc.).

2. The report must contain a background section, description of the individual reviewers' roles in the review activities, summary of findings for each TOR in which the weaknesses and strengths are described, and conclusions and recommendations in accordance with the TORs. The independent report shall be an independent peer review, and shall not simply repeat the contents of the SARC Summary Report.

   a. Reviewers should describe in their own words the review activities completed during the panel review meeting, including a concise summary of whether they accept or reject the work that they reviewed, and explain their decisions (strengths, weaknesses of the analyses, etc.), conclusions, and recommendations.

   b. Reviewers should discuss their independent views on each TOR even if these were consistent with those of other panelists, but especially where there were divergent views.

   c. Reviewers should elaborate on any points raised in the SARC Summary Report that they believe might require further clarification.

   d. The report may include recommendations on how to improve future assessments.

3. The report shall include the following appendices:

   Appendix 1: Bibliography of materials provided for review
   Appendix 2: A copy of this Statement of Work
   Appendix 3: Panel membership or other pertinent information from the panel review meeting.

Appendix 2.4. SARC Summary Report Requirements

1. The main body of the report shall consist of an introduction prepared by the SARC chair that will include the background and a review of activities and comments on the appropriateness of the process in reaching the goals of the SARC. Following the introduction, for each assessment reviewed, the report should address whether or not each Term of Reference of the SAW Working Group was completed successfully. For each Term of Reference, the SARC Summary Report should state why that Term of Reference was or was not completed successfully.

   To make this determination, the SARC chair and reviewers should consider whether or not the work provides a scientifically credible basis for developing fishery management advice. If the reviewers and SARC chair do not reach an agreement on a Term of Reference, the report should explain why. It is permissible to express majority as well as minority opinions.

   The report may include recommendations on how to improve future assessments.

2. If any existing Biological Reference Points (BRPs) or BRP proxies are considered inappropriate, include recommendations and justification for alternatives. If such alternatives cannot be identified, then indicate that the existing BRPs or BRP proxies are the best available at this time.

3. The report shall also include the bibliography of all materials provided during the SAW, and relevant papers cited in the SARC Summary Report, along with a copy of the CIE Statement of Work.

   The report shall also include as a separate appendix the assessment Terms of Reference used for the SAW, including any changes to the Terms of Reference or specific topics/issues directly related to the assessments and requiring Panel advice.

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

SARC and CIE Reviewers
Robert Latour
Robin Cook
John Casey
Yan Jiao

Presenters from Summer Flounder Stock Assessment Team
Mark Terceiro

Presenters from Striped Bass Stock Assessment Team
Katie Drew
Gary Nelson
Michael Celestino

Other participants and their affiliation and contacts (provided by NEFSC)

Name	Affiliation	Email
Charles Adams	NEFSC	charles.adams@noaa.gov
Miriam Ameworwor	URI/UCC	mameworwor@gmail.com
Max Appelman	ASMFC	mappelman@asmfc.org
Greg Ardini	NEFSC	gregory.ardini@noaa.gov
Rick Bellavance	RIPCBA	rickbellavance@gmail.com
Jason Boucher	DE Fish and Wildlife	jason.boucher@state.de.us
Russell Brown	NEFSC	russell.brown@noaa.gov
Jeff Brust	NJ DFW	jeffrey.brust@dep.nj.gov
Steve Cadrin	SMAST	scadrin@umassd.edu
Toni Chute	NEFSC	toni.chute@noaa.gov
Jessica Coakley	MAFMC	jcoakley@mafmc.org
Tiffany Cunningham	MADMF	tiffany.vidal@mass.gov
Kiley Dancy	MAFMC	kdancy@mafmc.org
Dvora Hart	NEFSC	deborah.hart@noaa.gov
Evans Kwasi Arizi	URI	evansarizi@uri.edu
Najih Lazar	URI-GSO	nlazar@uri.edu
Chris Legault	NEFSC	chris.legault@noaa.gov
Nicole Lengyel	RI DMF	nicole.lengyel@dem.ri.gov
Brian Linton	NEFSC	brian.linton@noaa.gov
Alicia Miller	NEFSC	alicia.miller@noaa.gov
Paul Nitschke	NEFSC	pnitschke@noaa.gov
Patrick Paquette	MA Striped Bass Association	BasicPatrick@aol.com
Charles Perretti	NEFSC	charles.perretti@noaa.gov
Anne Richards	NEFSC	anne.richards@noaa.gov
Kirby Rootes-Murdy	ASMFC	krootes-murdy@asmfc.org
Gary Shepherd	NEFSC	gary.shepherd@noaa.gov
Katherine Sosebee	NEFSC	ksosebee@noaa.gov
Patrick Sullivan	Cornell University	pjs31@cornell.edu
Michele Traver	NEFSC	michele.traver@noaa.gov
Jim Weinberg	NEFSC	james.weinberg@noaa.gov
Tony Wood	NEFSC	anthony.wood@noaa.gov