Center for Independent Experts Independent Peer Review Report

On

Index Based Methods and Harvest Control Rules
Research Track Peer Review

Prepared by

Dr. Yong Chen

Professor of Fisheries Population Dynamics
School of Marine Sciences
University of Maine
Orono, ME 04469

January 2021

Contents

I. Executive Summary

The External Independent Peer Review of the Index Based Methods (IBMs) and Harvest Control Rules Research Track Program was held online from December 7-11, 2020. The Review aimed to evaluate the effectiveness of a developed modeling framework aimed at simulating fisheries mimicking typical groundfish fisheries characteristics with significant retrospective patterns in an age-structured stock assessment; evaluate the performance of 13 identified IBMs using the 50 metrics proposed in the study; identify key factors influencing the IBM performance; develop guidance in the application of IBMs in stock assessments showing retrospective pattern problems; and provide guidelines for setting biological reference points for the IBM-assessed stocks. The Review also needs to make recommendations to improve this topic-based Research Track Assessment.

The Index Based Methods and Harvest Control Rules Research Track is the first topic-specific Research Track Assessment. NOAA Northeast Fisheries Science Center (NEFSC) provided all the necessary logistics support, IBM Working Group (IBMWG) documentation, background information, and online access to the model description, computer codes and simulation outputs prior to and during the Review. The IBMWG Chair, Dr. Chris Legault, presented the IBMWG work regarding the Terms of References (ToRs). Dr. Legault and other IBMWG members were open to questions and suggestions in the Review and provided additional analysis and information upon request. The whole process was open and constructive, and all materials were sent to me in a timely manner. As a Center for Independent Experts (CIE) reviewer, I am charged to evaluate the 2020 Index Based Methods and Harvest Control Rules Research Track with respect to a set of pre-defined ToRs.

I would like to commend Dr. Legault's efforts for his excellent presentations and the IBMWG for providing online access to all necessary information and documentation. I was impressed by the careful consideration and justification of simulation designs, cutting edge and state-of-the-art modeling approach, breadth and depth of expertise, the amount of effort spent to compile all the data and outputs, the considerations of plausible simulation scenarios, the openness of discussion on potential issues, the willingness to consider alternative approaches and suggestions and the constructive dialogues among the Review Panel, IBMWG and other participants during the Review.

Overall, based on the materials I received, the information presented during the review and the discussion we had in the Review, I believe that the IBMWG has done an excellent job in addressing the ToRs within the allocated time frame. The work is scientifically sound and adequately addresses issues related to identifying alternative approaches when a statistical catch-at-age (SCAA) model is rejected because of a significant retrospective pattern. The IBMWG developed a management strategy evaluation (MSE) framework and relevant computer program to simulate data with desirable retrospective patterns in SCAA stock assessments. The IBMWG identified 13 IBMs and 50 performance metrics and considered many factors in developing the simulation scenarios, including life history characteristics, fishing history, recruitment dynamics, sources of retrospective pattern, and data quality, which might influence the IBM performance. The IBMWG concluded that an index-based model would not be expected to perform better than

a rho-adjusted SCAA. The rho-adjusted SCAA performance metrics were ranked high or above middle for most performance metrics. The rho-adjusted SCAA also demonstrated the best ability among all models to prevent overfishing in both short- and long-term comparisons, and was best at preventing the overfished condition for the long-term performance (but ranked in the middle for the short-term). The rho-adjusted SCAA was found resilient to the source of the retrospective pattern. However, like all simulation studies, the conclusions derived from this study may be conditional on the scenarios considered in this study. I recommend that the IBMWG continue their research effort and provide a list of research recommendations (see Section V for details) for the IBMWG to consider in addressing the concerns/comments raised during the Review.

II. Background

Age-structured stock assessment models are commonly used in the US to yield estimates of annual stock biomass and fishing mortality rates, as well as reference points to determine the stock and fishery status. The quality of a stock assessment, which is often evaluated based on various model diagnostics, is critical in determining if the assessment results can be used to provide catch advice for fisheries management. An erroneous stock assessment with large uncertainty may lead to fisheries mismanagement. Thus, it is important to conduct model diagnostics to evaluate the quality of stock assessments. One of such diagnostics is an evaluation of the retrospective pattern, which is defined as systematic inconsistency in assessment estimates with additional data used in the model fitting (Mohn 1999; Miller and Legault 2017).

Positive retrospective pattern, in which stock biomass is estimated downward and fishing mortality estimated upward as additional data is included, is common in many stock assessments in the Northeast USA. The positive retrospective pattern poses a challenge for NOAA Fisheries, New England Fisheries Management Council (NEFMC), Mid-Atlantic Fisheries Management Council (MAFMC) and Atlantic States Marine Fisheries Commission (ASMFC) in providing catch advice and determining the stock and fisheries status. This problem is particularly acute for many groundfish stocks assessed using age-structured models and managed under NOAA Fisheries and the NEFMC multispecies groundfish management plan (e.g., NEFSC 2002a, 2002b, 2005a, 2005b, 2008, 2017 and 2019).

The retrospective pattern problem is usually quantified by a measure known as Mohn's rho (Mohn 1999). For the assessment of a fish stock managed under the NEFMC, if its rho-adjusted (divide the terminal year estimate by one plus Mohn's rho) value is outside the 90% confidence interval of the terminal year estimate of spawning stock biomass or fishing mortality rate, the retrospective pattern is considered strong. In this case, the rho-adjusted values are used for determining stock status and modifying the starting stock size for projections used to provide catch advice (Brooks and Legault 2016).

Large persistent retrospective patterns may result from mismatches between model assumptions/configurations and the quality of data used in model fitting. In the Northeast US, previous studies identified three key sources that may result in the retrospective pattern problem in a stock assessment: unaccounted catch resulting from unreported discards and bycatch (and/or under-reported catch); unaccounted temporal trends in natural mortality; and mismatch between the model assumptions on temporal trends in survey catchability and actual survey catchability (Legault 2009; Miller and Legault 2017). However, in practice it is often difficult, if not impossible, to determine the source(s) of retrospective pattern in a particular stock assessment. This makes the removal of retrospective patterns by reparametrizing/reconfiguring the model in a particular stock assessment difficult.

The existence of large retrospective patterns and the inability to remove them in the assessments led to the rejection of some age-structured stock assessments, including Atlantic mackerel (*Scomber scombrus*), Georges Bank Atlantic cod (*Gadus morhua*), Georges Bank yellowtail flounder (*Limanda ferruginea*), and witch flounder (*Glyptocephalus cynoglossus*)

(Deroba et al. 2010; Legault et al. 2014; NEFSC 2015a, 2015b). For those rejected stock assessments, their catch advices were made based on various *ad hoc* index-based approaches (McNamee et al. 2015; NEFSC 2015a, 2015b; Wiedenmann 2015). A recent study suggested that some of the approaches might be inappropriate for fish stocks in the Northeast US with a history of high exploitation rates (Wiedenmann et al. 2019). Strong positive retrospective patterns persisted for some stocks in the region (NEFSC 2019), raising concerns that more stocks may have to rely on data-limited approaches such as index-based methods in the future. Additionally, some of these stocks are in the process of rebuilding without an accepted stock assessment or a mechanism to track rebuilding progress. In 2020, index-based assessments for the red hake and Northern windowpane flounder stocks were rejected and alternative index-based methods were proposed. Taken together, these concerns raise the need to evaluate and identify appropriate index-based approaches for setting catch advice and determining stock status for stocks with age-structured assessments that are rejected for their strong retrospective patterns.

The Index-Based Methods Working Group (IBMWG) was formed to conduct a topic - based Research Track Assessment to identify and evaluate a range of index-based methods for providing catch advice and determining status for stocks with strong retrospective patterns in their age-based assessments. The Terms of Reference (ToRs) defined by the Northeast Region Coordinating Council for the Index-Based Methods (IBM) Research Track Stock Assessment are as follows:

1. *Develop methods to create data that if assessed with standard age-based approaches (e.g., VPA or ASAP) could exhibit a strong retrospective pattern.*
2. *Identify a number of index-based methods and a range of harvest control rules for use in closed-loop simulation, using index-based data resulting from ToR 1.*
3. *Identify metrics from the index-based assessment results that could be used in evaluations of trade-offs in performance among harvest control rules and index-based methods.*
4. *Evaluate the combinations of index-based methods and control rules using the metrics in ToR 3 to determine candidates for consideration by the Councils or other management authorities.*
5. *Provide guidance on specific situations that are and are not well-suited for a particular control rule or index-based method identified in ToR 4.*
6. *Create guidelines for setting biological reference points for index-based stocks.*

The IBMWG addressed these ToRs by developing a Management Strategy Evaluation (MSE) framework, using the MSE to simulate the data with desired retrospective patterns and conducting an extensive simulation study to evaluate the performance of various index-based stock assessment methods. The IBMWG focused this study on the methods that can use fishery-independent survey abundance indices available in the Northeast US. The IBMWG worked on this project remotely with weekly online meetings and a unique online distribution system and workflow management for the IBMWG participation and collaborations in computer coding, simulation work and report writing. All the simulation outputs are saved.

The Review was conducted online via webinar from 8:00 am to 3:00 pm, December 7-11, 2020 because of the Covid-19 pandemic. All sessions were open to the public. Michele Traver of the NEFSC coordinated the meeting's logistics and provided technical support. The process ran

smoothly, with little technical difficulty. The IBMWG Chair, Dr. Chris Legault, presented the IBMWG work with respect to each ToR during the Review (see the list of presentations in Appendix I). The Review Panel made several requests for additional information for clarification (e.g., Dynamic Linear Model approach) and analyses of saved simulation results (e.g., correlations between performance metrics). These additional sources of information and analyses were provided to further improve our understanding of the property of simulated data, possible impacts of various factors on the retrospective patterns and performance of selected index-based stock assessment methods. The Review was attended by the IBMWG, NOAA Fisheries scientists, NEFMC and MAFMC representatives and other stakeholders (see the List of Participants in Appendix II). All the participants were given opportunities to be engaged in the discussion (see the Review schedule in Appendix III). The MAFMC SSC Chair Dr. Paul Rago chaired the Review Panel, which consisted of him and the three CIE reviewers.

This Review is the first CIE review for the topic-based Research Track Assessment. I was provided with all necessary documentation, access to data and computer codes, simulation outputs, background information and logistics support. Documentation for the meeting was provided via Google Share Drive https://drive.google.com/drive/u/0/folders/1VqyaTfGzod5rCwuqiHhJXW4C8hvJfhBt and the NEFSC Data Portal https://apps-nefsc.fisheries.noaa.gov/saw/sasi/sasi_report_options.php. All relevant computer simulation source code and model outputs were available via a GitHub site https://github.com/cmlegault/IBMWG developed by the IBMWG Chair Dr. Chris Legault and the IBMWG. Dr. Chris Legault and the other IBMWG members were extremely responsive and open to suggestions, and provided additional information and analysis upon request. The whole review process was open and constructive.

Dr. Rago presented the key findings and preliminary recommendations on behalf of the Review Panel on Thursday December 10. The presentation included a broad outline of general conclusions and recommendations that fully reflected the perspectives of the Review Panel.

As a CIE reviewer, I am charged to evaluate the 2020 Index Based Methods and Harvest Control Rules Research Track Assessment with respect to the pre-defined ToRs. This report includes an executive summary (Section I), a background introduction (Section II), a description of my role in the review activities (Section III), my comments on each item listed in the ToRs (Section IV), a summary of my conclusion and recommendations (Section V) and references (Section VI). The final part of this report (Section VII) includes a collection of appendices including the Performance Work Statement.

III. Description of the Individual Reviewer's Role in the Review Activities

My role as a CIE independent reviewer is to conduct an impartial and independent peer review of the Index Based Methods and Harvest Control Rules Research Track Assessment with respect to the pre-defined Terms of Reference.

I received the IBMWG draft report and other relevant background papers and reports about 10 days prior to the review. I read the draft IBMWG report, background information papers and reports, and other relevant documents that were sent to me (see the list in the Appendix I). I have also searched, collected and read references relevant to the topics covered in the reports and the Performance Work Statement (PWS) prior to the Review.

I was actively involved in the discussion during the Review by (1) questioning and asking for clarification; (2) commenting on the process and work; (3) making constructive comments and suggestions for alternative approaches and additional analyses and (4) contributing to the Review Panel summary report. This report summarizes my independent findings and recommendations for the Review.

IV. Summary of Findings

My independent findings and comments on each item of the ToRs are provided under the respective subtitles of the ToRs (see below).

1. *Develop methods to create data that if assessed with standard age-based approaches (e.g., VPA or ASAP) could exhibit a strong retrospective pattern.*

This ToR was met.

The IBMWG has successfully developed the method and associated simulation algorithms and computer codes to simulate data, which can exhibit a strong retrospective pattern if assessed with standard age-based approaches. The choice to use WHAM (Woods Hole Assessment Model that is a general state-space age-structured assessment framework with options to include environmental effects on population processes), which is flexible in capturing various error sources and tailors to the biology and fishery of the Northeast US, allows for quick implementation and adjustment for future development. The simulation modeling framework is scientifically sound, efficient and flexible to simulate the dynamics of fish populations with different life history traits and fishing history and can generate pre-defined retrospective patterns that exist and are persistent in both the base period and the feedback period in this study. Thus, I conclude that this ToR has been adequately addressed. The IBMWG should be commended for their excellent job in developing the effective operating model and MSE. I would like to raise the following comments/caveats.

The IBMWG conditioned the simulation designs and scenario development to the Northeast US groundfish fisheries to more realistically simulate the retrospective patterns encountered in an actual stock assessment. The data were simulated for a specific life history trait and recruitment dynamics with considerations of a few scenarios of fishing history,

selectivity, and biased errors in catch and M. Thus, although this ToR was met, like almost all simulation-based studies, the data simulated have limitations and may not necessarily cover retrospective patterns and population dynamics shown in different fisheries. This may limit the generalization of the final recommendations. The robustness of the conclusions derived in this study may need to be evaluated with data simulated using different fisheries properties such as life history traits; recruitment dynamics; fishing history; temporal variability in selectivity and sources of retrospective patterns such as temporal changes in M and survey catchability and unaccounted catch.

The large number of simulation runs and the closed-loop hands-off simulation design make it hard to evaluate the biological realisms of individual simulation runs. This approach is also inconsistent with the actual index-based stock assessment in which the model can be tuned and adjusted based on model diagnosis. This inconsistency should be considered in the interpretation of simulation results.

The observation errors considered were limited. The CVs used by the IBMWG for catch and survey indices were comparable to those of many groundfish fisheries, but effective sample sizes used in simulating catch and survey age compositions were higher than the typical sample sizes for groundfish fisheries.

Only two sources of retrospective patterns were considered in the simulation: unaccounted catch (e.g., under-reported catch, unreported discards and/or unreported bycatch) and unaccounted temporal increase in natural mortality M (e.g., increased predations). Temporal trend in survey catchability, which has been identified as one of the major sources for retrospective patterns, was not included because the IBMWG was unsuccessful in generating data with desired Mohn's rho values. Given the issues related to the survey catchability in the Northeast groundfish fisheries and the importance of survey indices for the index-based stock assessment, it is important to consider temporal trends in survey catchability in simulating data that would lead to retrospective patterns. The independent simulation of similar levels of retrospective pattern for unaccounted catch and temporal increase in M made the relevant scenarios comparable. However, this may be different from the retrospective pattern observed in an actual fishery in which the retrospective pattern may arise as a result of some combinations of unaccounted temporal changes in M and survey catchability and unaccounted catch.

2. *Identify a number of index-based methods and a range of harvest control rules for use in closed-loop simulation, using index-based data resulting from ToR 1.*

<u>This ToR was met</u>.

The IBMWG identified 13 IBMs covering a wide range of data needs and statistical and biological assumptions. The list is comprehensive, and the justification of the choices is sound. The IBMWG also identified two HCRs for setting ABC (one is the same as the IBM modeling result, the other reduces IBM catch advice by 75%) mimicking the HCRs used in the NEFMC

for setting ABC for index-based stock assessments. I conclude that the IBMWG has adequately addressed this ToR. However, I would like to raise the following issues.

The closed-loop hands-off simulation protocol was used in the simulation, which did not allow model diagnostics for modeling tuning and adjustment commonly done in a practical index-based stock assessment. Such an inconsistency may affect the conclusions regarding the IBM performance.

For HCRs, no adjustment can be made based on fisheries status because the IBMs selected in this study cannot yield model-based biological reference points, which are often used to help determine the status of fisheries. The IBMs, which can be used to yield model-based biological reference points, were not included in this study (e.g., surplus production models).

The decision to use IBMs in providing catch advice may vary between fisheries in the Northeast US. For some fisheries, the IBMs were used because a statistical catch-at-age (SCAA) assessment model was rejected for its strong retrospective pattern. However, for some other fisheries, IBMs were used because of data quantity and/or quality issues (e.g., no reliable age composition data). In these cases, the problem of retrospective patterns may not be an issue relevant to the choice of IBMs and the results derived in this study regarding the performance of different IBMs may not be applicable to these fisheries.

3. *Identify metrics from the index-based assessment results that could be used in evaluations of trade-offs in performance among harvest control rules and index-based methods.*

This ToR was met.

The IBMWG identified 50 metrics to evaluate the performance of different IBMs and HCRs. These metrics measure different aspects of the IBM-based catch advice performance, including catch, variability in catch, biomass of target species, fishing mortality and other miscellaneous aspects (e.g., ecosystem, legislative). The metrics were identified and developed with inputs from the Council staff and were aimed to address the concerns and needs of the Council and other stakeholders. The metrics are comprehensive and provide the information to evaluate trade-offs in key fisheries and conservation statistics for different IBMs and HCRs. The metrics also provide the capacity to evaluate the IBM performance in avoiding overfishing and controlling stock biomass. The large quantity of outputs saved in the simulation study allows continuous efforts to improve understanding of IBM performance in the future. I conclude that the IBMWG has adequately addressed this ToR. I would like to raise the following issues.

Although it might be a good strategy to include and save the results for a large number of metrics in the simulation, the large amount of information derived makes the evaluation of IBM

performance challenging. Some metrics are repetitive, and studies are needed to identify the correlations between these metrics and to reduce the number of the metrics.

I suggest including the total catch and the lowest SSB for each simulation run during the feedback time period to have an overall measure of catch and avoid the risk of having unnecessarily low stock biomass. The number of simulation runs that have fishing mortality rates higher than 2.0 (i.e., the boundary value defined for F in the simulation) may also be a useful measure to be include in the metrics for identifying simulation runs with potential issues (e.g., unrealistically high F and/or low stock biomass).

4. *Evaluate the combinations of index-based methods and control rules using the metrics in ToR 3 to determine candidates for consideration by the Councils or other management authorities.*

This ToR was met.

The IBMWG conducted an extensive simulation study to evaluate the performance of different combinations of index-methods and HCRs in addressing management needs by the Council using the performance metrics developed in ToR 3. The simulation design considered the different combinations of sources of the retrospective pattern problem, fishing history, fishing selectivity and catch advice. The SCAA was also included in the evaluation. Various approaches were used to identify critical factors influencing the performance of IBMs. The IBMWG developed a sophisticated scoring protocol to facilitate and streamline the comparisons of IBMs on a large number of metrics. Some general patterns emerged in the initial analysis. I conclude that the IBMWG has adequately addressed this ToR. However, I would like to raise the following issues.

Although initial analysis identified some interesting and important results regarding the IBM performance, any generalization of the results needs to be careful about overextending claims beyond the simulation scenarios considered and analyzed in this particular study. This is especially true for the robustness of the results with respect to sources of retrospective pattern, life history traits, fishing history and recruitment dynamics.

The "bigger is better" ranking system in comparing the IBM performance over a large number of metrics may not necessarily serve the management objectives that balance the stock conservation and fishing opportunity. The large number and trade-off nature of the metrics used in this analysis made the interpretation of the comparison study challenging and less straightforward. It appears that the IBMWG did not have sufficient time to fully explore and analyze all outputs.

The IBMWG concluded that the IBMs do not outperform the SCAA with retrospective patterns being Mohn's rho adjusted. This result was rather consistent over the simulation scenarios considered in this study. However, like all other simulation studies, the generalization of this conclusion may be limited by the simulation designs considered in this study. The robustness of this conclusion and implications for recommending alternative approaches when a SCAA stock assessment is rejected because a strong retrospective pattern may need to be further evaluated.

5. *Provide guidance on specific situations that are and are not well-suited for a particular control rule or index-based method identified in ToR 4.*

This ToR was partially met.

Although the analysis conducted by the IBMWG provided important results and identified the situations where the group of models do well regarding catch and biomass-based metrics, the specific guidance regarding whether a particular IBM and/or HCRs identified in ToR 4 should be used was not fully provided. Thus, I conclude that this ToR has been partially addressed. I would like to raise the following issues.

The conclusions regarding the IBM performance were not straightforward (and sometimes were difficult to interpret) and might be conditional on the scenario designs, making difficult the development of a general guidance on specific situations not suited for an IBM. It might be unlikely that a "best" practice guideline could be developed given the complexity of issues that may not be fully covered in the simulations, the trade-off nature of performance metrics, and the potentially conflicting interests of different groups of stakeholders. The results might be tentative and stock-specific and would benefit from further work.

The IBMWG did provide a general conclusion that the IBMs would not outperform the Mohn's rho-adjusted SCAA model for the simulation scenarios considered in this study. However, more studies may be needed to evaluate the robustness of this advice with respect to the assumptions/configurations considered in this study.

Multivariate methods such as principle component analysis may be used to help reduce the dimensionality of the outputs and further identify relationships among the IBMs. Methods such as ROC (Receiver Operator Curves), decision trees and various utility functions may be useful to help develop general guidance for the use of IBMs.

6. Create guidelines for setting biological reference points for index-based stocks.

This ToR was partially met.

The IBMs considered in this study do not provide internally estimated model-based biological reference points because they lack production functions allowing examination of density-dependent tradeoffs between catch and future population size. The IBMs allowing for estimating model-based reference points (e.g., biomass dynamic models) were not included in this study. This study suggests that the use of IBMs to determine reference points might not work well. Thus, I conclude that this ToR has been partially addressed.

This ToR, as defined, might not be most appropriate, as IBMs tend not to yield internally estimated model-based biological reference points. Identifying directional changes in catch/fishing mortality without a good understanding of stock dynamics may be useful in the short run, but provides only limited advice for achieving long-term management targets. Some empirical reference points based on temporal changes in abundance indices over time might be useful for the IBMs to keep the fisheries in desirable status in comparison to historical levels. Such an approach needs to consider possible changes in the ecosystem dynamics.

V. Conclusions and Recommendations

Overall, based on the IBMWG report, background information, materials presented and additional information provided during the Review, I believe that the IBMWG has done an excellent job in addressing the ToRs. The MSE developed and used in this study is cutting edge and state-of-the-art. The results derived are scientifically sound and adequately address most of the ToRs. I realize that the IBMWG ran short of time in this Assessment to conduct a more in-depth analysis of simulation results and to evaluate and identify possible reference points, and I recommend that the IBMWG continue this research effort. I provide the following conclusions and recommendations for future research effort.

The IBM stock assessment is the first Topic-based Research Track Assessment in the Northeast Fisheries Science Center. Because of the pandemic, the work was done remotely with weekly meetings and online collaborations. Given the outputs achieved by the IBMWG, the distribution system and workflow management for workgroup participation and collaboration appears effective and productive for this project. This type of research management approach may be applicable to other research track assessments of similar nature in the future. Of course, a mixed model of in-person and online meetings and collaborations may be even more effective if permitted.

The choice of WHAM for the operating model in the MSE framework developed in this study allows for local control of the operating model and program codes. The open computer program coding system provides a general and flexible modeling framework and software for easy collaboration and customization. The decision to use the same set of 1000 random seeds across different scenarios makes the results comparable among different IBMs. Saving all model

outputs allows further analyses without rerunning the simulations. The approaches utilized in this project will serve other research assessments of a similar nature well in future.

Unlike a typical Stock-specific Research Track Assessment, this Topic-based Research Track Assessment is open-ended and requires more time and in-depth analysis to address questions and issues identified during the process. Such Topic-based Research Track Assessment calls for flexibility in redefining/modifying the ToRs with an improved understanding of the issues and questions generated during the work process. Such a program may need to have a built-in milestone for periodic evaluations by external reviewers and stakeholders for possible mismatches between what is planned in the ToRs and what is feasible and/or required; for addressing new questions arising during the project; and for possible new insights from the external experts.

More time may need to be allocated to a Topic-based Research Track Assessment, which tends to be open ended and often generate many additional questions during the process. The 8-month duration for this research track assessment appears insufficient, and time constraints precluded a full analysis and synthesis of the simulation results and evaluation of alternative simulation scenarios, IBMs, HCRs and metrics. I recommend continuation of this research effort.

The closed-loop hands-off simulation approach used was an efficient and tractable way to conduct the simulation study. The results derived from such an approach only depend on the initial conditions, making it possible to identify key factors influencing the results. This approach also can address additional aspects of IBMs for models with retrospective patterns, as well as effects of missing 2020 survey data on model performance. However, such a closed-loop hands-off simulation approach does not provide the modelers an opportunity to intervene in a simulation run (e.g., model tuning, diagnostics, and checking possible violations of biological and statistical assumptions and unrealistic catch advice). Thus, it is inconsistent with the normal practice in an actual stock assessment. I would recommend some in-depth analyses for some selected simulation runs with unexpected results or runs that hit the pre-defined boundary values (e.g., $F > 2.0$) to evaluate possible issues. I would also encourage the IBMWG to treat a few simulation runs with interesting results as if they were "real stocks" and follow the actual analysis protocol for applying model tuning and diagnostics. The modelers can then compare the results from such an in-depth analysis with those from the closed-loop simulation results to identify differences and possible implications of using the closed-loop simulation approach. Such a comparative study may help improve our understanding of the robustness of hands-off simulation design with respect to the application of conclusions to an actual stock assessment.

The MSE platform developed in this study is flexible and has the potential to provide a generalized modeling framework for simulating the stock assessment data of desirable characteristics. Such a generalized simulation framework can also be used to simulate fisheries with different life history strategies and fishing histories, compare the performance of different models, and identify possible causes of model failure and the potential impacts of data quality and quantity on the quality of stock assessment and catch advice. I would encourage the IBMWG to continue developing the MSE and make a generalized MSE framework.

This study only includes the retrospective pattern with positive Mohn's rho (i.e., the terminal year of SSB is over-estimated and the F under-estimated). Although not common in the Northeast groundfish fisheries, there are fisheries that have the retrospective pattern problem with negative Mohn's rho (i.e., terminal SSB under-estimated). Although such a retrospective pattern may not result in overfishing, it will lead to a missed fishing opportunity. It would be interesting to evaluate whether the MSE can simulate fisheries with persistent retrospective patterns of negative rho and how IBMs perform in assessing such fisheries.

Thirteen IBMs were included in this study, including those used in the Northeast US and other areas. I was impressed by the thoroughness and comprehension of the list, in particular, to the methods applicable to the groundfish fisheries in the Northeast US. Most models were well defined and studied. However, two IBMs, the Ensemble Model (EM) and Dynamic Linear Model (DLM), were not well studied in this assessment because of time constraints. Both EM and DLM have the potential to perform more robustly regarding retrospective patterns. The selection and possible weighting of the IBMs included in the EM need to be better defined and justified. The DLM approach has the inherent ability to continuously adjust the estimates of a trend with newly available data. However, because of time constraints, the DLM was incomplete for some scenarios. Further development and evaluation of both EM and DLM is recommended.

Biological realism of the simulation design was justified using typical groundfish fisheries (e.g., life history characteristics, fishing history, recruitment dynamics and data quality and quantity) in the Northeast US. Future studies may evaluate the robustness of IBM performance regarding different life history characteristics (e.g., pelagic species), fishing history, recruitment dynamics and/or data quality and quantity.

The current simulation design assumes that the IBM-based assessment is conducted every two years. This is true for most groundfish fisheries in the Northeast US. However, changing the assessment frequency (e.g., every year or every three years) may be considered in the future for some stocks. Different assessment frequencies in the feedback period may need to be considered in future studies.

The fishing mortality was capped at 2.0 in simulation runs to prevent unrealistically high fishing mortality. The number of simulation runs in which F is higher than 2 in a scenario may represent the poor performance of an IBM and/or unrealistic simulation settings. Thus, I suggest that the number of simulation runs with F reaching the cap of F =2.0 be used as one of the metrics for the IBM performance.

The IBM performance was evaluated through short-term (6 years) and long-term (20 years) metrics that measured different characteristics of IBM performance. The short-term evaluation results are greatly influenced by initial conditions, and the long-term results may suggest equilibrium behavior. The long-term metrics may not be indicative of future stock status. Separate analyses are needed for the short-term and long-term metrics in evaluating IBM performance.

I appreciate the IBMWG's efforts to include an exhaustive list of performance metrics that measure five aspects of IBM performance in fisheries management. Although it is good to have the results for all 50 metrics saved in the simulation for future analysis and to address the interests/concerns of different stakeholders, the inclusion of all 50 metrics in the analysis and subsequent interpretation becomes challenging. Many metrics are highly correlated and some multivariate statistics methods (e.g., principal component analysis) can be used to reduce the dimensionality of the 50 metrics in the IBM performance analysis and comparison. The following four measures are often used in MSE to compare the performance of assessment/management methods during the management time period: (1) total catch to measure overall yield; (2) variation in catch to measure the stability of a fishery; (3) the lowest stock biomass to prevent short-term high fishing mortality and avoid overfished stock status; and (4) the stock biomass at the end of the management period. Stakeholder consultation may be needed to reduce the number of performance metrics.

The current MSE formulation does not yield length composition data, excluding the length-based data-limited assessment methods being included and evaluated. Many length-based data-limited methods may be potential candidates for the groundfish stock assessments that are rejected because of retrospective pattern problems with statistical age-structured assessment models. Future studies may modify the MSE to allow for the output of length composition data so that they can be used to evaluate length-based methods.

Index-based models have fewer biological and statistical assumptions, fewer model parameters, and require less information compared to the SCAA models. Unlike the SCAA models, the IBMs have no direct way of responding to the source of the retrospective pattern. Instead, they rely on an ability to respond more flexibly to changing conditions, additional information to fix population scale (e.g., ES methods), and/or robust age composition information to estimate biomass and total mortality (e.g., CC methods). I suggest listing statistical and/or biological assumptions and requirements of data quality and quantity for each index-based method used in the study for a better understanding of the limitations and possible implications of using a particular method.

The IBMWG developed an interesting "bigger is better" ranking system to compare IBM performance. However, there may not be a simple interpretation of "better" for different performance metrics. For example, higher biomass may not always be good (possibly missing fishing opportunity) and lower catch (possibly avoiding overfished stock when biomass is low) is not always bad. Thus, there is no simple way to clearly identify the "best" IBM for any given scenario. However, this study can help identify "bad" IBMs that lead to overfishing and "overfished" stocks. An "odds ratio" calculated by the proportion of simulation runs falling in each quadrant may be used to compute the odds of overfishing given that the stock is below Bmsy and conversely the odds of under harvesting when the stock is above Bmsy. A decision tree approach, which can build on the exploratory results of the ANOVAs and cluster analysis, may be used to advise the utility of the IBMs given our understanding of various sources of uncertainty in the stock assessment.

The IBMs differ in their sensitivity to different sources of retrospective patterns. DLM, PlanB, ES-Frecent, Islope and two catch curve methods tended to be more robust to retrospective

pattern sources. However, the mechanisms for different performances are unknown. More studies are needed to investigate why the differences occur among IBMs.

The IBMWG concludes that an index-based model will not be expected to perform better than a rho-adjusted SCAA. The rho-adjusted SCAA performance metrics were ranked high or above middle for most performance metrics, and the rho-adjusted SCAA demonstrated the best ability among all models to prevent overfishing in both short- and long-term comparisons, and was best at preventing the overfished condition in the long-term performance (but ranked in the middle for the short-term). The rho-adjusted SCAA was found resilient to the source of the retrospective pattern. While the performance of some IBMs was highly dependent on the source of the retrospective pattern (underreported catch vs increased natural mortality), the SCAA performed well across both sources of retrospective pattern included in the study. However, because retrospective pattern was simulated only for a single source at a time and no temporal trend in survey catchability was considered in simulating retrospective patterns, we do not know if the SCAA is robust to the retrospective pattern when it arises from the presence of two or three sources. A simulation study might be useful to address this issue.

The IBMWG tested the efficacy of Mohn's rho adjustment by rescaling the biomass estimate by the multiplier 1/(1+Mohn's rho). It is important to note that there are other ways to adjust the retrospective pattern and that the approach used in this study differs from the current practices in the Northeast. In the Northeast stock assessment, the retrospective adjustment is applied only when the rho-adjusted estimates of biomass and fishing mortality are outside the joint 90% confidence regions of B and F in the terminal year. The impacts of such a difference between this study and current practice may need to be evaluated for future use of Mohn's rho-adjustment for the SCAA-based stock assessment.

ToR 6, as defined, might not be most appropriate, because IBMs included in this study do not yield internally estimated biological reference points. Identifying directional changes in catch/fishing mortality without a good understanding of stock dynamics might be useful in the short run, but may only provide limited advice in achieving long-term management targets. Some empirical reference points based on temporal changes in abundance indices over time might be useful for the IBMs to keep fisheries in desirable status in comparison to historical levels. This approach, however, needs to consider possible changes in the ecosystems.

VI. References

Deroba, J., Shepherd, G., Gregoire, F., Nieland, J., Rago, P. 2010. Stock assessment of Atlantic mackerel in the Northwest Atlantic for 2010. Transboundary Resources Assessment Committee, Reference Document 2010/01. 59 pp.

Legault, C.M., Alade, L., Gross, W.E., Stone, H.H. 2014. Stock Assessment of Georges Bank Yellowtail Flounder for 2014. TRAC Ref. Doc. 2014/01. 214 p. Available from http://www.nefsc.noaa.gov/saw/trac/

Legault CM, Chair. 2009. Report of the Retrospective Working Group, January 14-16, 2008, Woods Hole, Massachusetts. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 09-01; 30 p. Available from: National Marine Fisheries Service, 166 Water Street, Woods Hole, MA 02543-1026, or online at http://www.nefsc.noaa.gov/nefsc/publications/

McNamee, J., Fay, G., Cadrin, S. 2015. Data Limited Techniques for Tier 4 Stocks: An alternative approach to setting harvest control rules using closed loop simulations for management strategy evaluation. Final report to the Mid Atlantic Fishery Management Council. Available: https://static1.squarespace.com/static/511cdc7fe4b00307a2628ac6/t/55a661a5e4b060ebc9d03cf0/1436967333432/DLanalysis_bsb_final.pdf

Miller, T. and C. Legault. 2017. Statistical behavior of retrospective patterns and their effects on estimation of stock and harvest status. Fisheries Research 186:109-120. DOI: 10.1016/j.fishres.2016.08.002

Mohn, R. 1999. The retrospective problem in sequential population analysis: An investigation using cod fishery and simulated data. ICES J. Mar. Sci. 56: 473-488.

Northeast Fisheries Science Center (NEFSC). 2002a. Assessment of 20 Northeast groundfish stocks through 2001: a report of the Groundfish Assessment Review Meeting (GARM), Northeast Fisheries Science Center, Woods Hole, Massachusetts, October 8-11, 2002. Northeast

Fish. Sci. Cent. Ref. Doc. 02-16. Available from National Marine Fisheries Service, 166 Water Street, Woods Hole, MA 02543-1026, or online at http://www.nefsc.noaa.gov/nefsc/publications/

Northeast Fisheries Science Center (NEFSC). 2002b. Re-evaluation of biological reference points for New England groundfish. Northeast Fish. Sci. Cent. Ref. Doc. 02-04; 395 p. Available from: National Marine Fisheries Service, 166 Water Street, Woods Hole, MA 02543-1026 or online at http://www.nefsc.noaa.gov/nefsc/publications/

Northeast Fisheries Science Center (NEFSC). 2005. Assessment of 19 Northeast groundfish stocks through 2004. 2005 Groundfish Assessment Review Meeting (2005 GARM), Northeast Fisheries Science Center, Woods Hole, Massachusetts, 15-19 August 2005. U.S. Dep. Commer., Northeast Fish. Sci. Cent. Ref. Doc. 05-13; 499 pp. Available from: National Marine Fisheries Service, 166 Water Street, Woods Hole, MA 02543-1026, or online at http://www.nefsc.noaa.gov/nefsc/publications/

Northeast Fisheries Science Center (NEFSC). 2008. Assessment of 19 Northeast Groundfish Stocks through 2007: Report of the 3rd Groundfish Assessment Review Meeting (GARM III), Northeast Fisheries Science Center, Woods Hole, Massachusetts, August 4-8, 2008. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 08-15; 884 pp. Available from: National Marine Fisheries Service, 166 Water Street, Woods Hole, MA 02543-1026, or online at http://www.nefsc.noaa.gov/nefsc/publications/

Northeast Fisheries Science Center (NEFSC). 2015a. Stock Assessment Update of 20 Northeast Groundfish Stocks Through 2014. US Dept Commer, Northeast Fish Sci Cent Ref Doc.15; 238 p. Available from National Marine Fisheries Service, 166 Water Street, Woods Hole, MA 02543-1026, or online at http://www.nefsc.noaa.gov/nefsc/publications/

Northeast Fisheries Science Center. (NEFSC) 2015b. 60th Northeast Regional Stock Assessment Workshop (60th SAW) Assessment Report. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 15-08; 870 p. Available from: National Marine Fisheries Service, 166 Water Street, Woods Hole, MA 02543-1026, or online at http://www.nefsc.noaa.gov/publications/

Northeast Fisheries Science Center (NEFSC). 2017. Operational Assessment of 19 Northeast Groundfish Stocks, Updated Through 2016. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 17-17; 259 p. Available from: National Marine Fisheries Service, 166 Water Street, Woods Hole, MA 02543-1026, or online at http://www.nefsc.noaa.gov/publications/

Northeast Fisheries Science Center (NEFSC). 2019. Operational Assessment of 14 Northeast Groundfish Stocks, Updated Through 2018. US Dept Commer, Northeast Fish Sci Cent Ref. 205 57p. Available from: National Marine Fisheries Service, 166 Water Street, Woods Hole, MA 02543-1026, or online at http://www.nefsc.noaa.gov/publications/

Wiedenmann, J. 2015. Application of data-poor harvest control rules to Atlantic mackerel. Final report to the Mid-Atlantic Fishery Management Council.

Wiedenmann, J., C.M. Free, and O.P. Jensen. 2019 . Evaluating the performance of data-limited methods for setting catch targets through application to data-rich stocks: A case study using Northeast U.S. fish stocks. Fisheries Research. 209: 129–142.

VII. Appendices

VII-1. Bibliography of materials provided for review

- Legault, C. et al. Draft Report of the Index Based Methods Working Group (59 pages)
- Legault, C. et al. Draft Report Tables (30 pages)
- Legault, C et al. Draft Report Figures (39 pages)
- Legault, C et al. Draft Report Appendices 1-6 (472 pages)
- NEFMC Groundfish Fisheries New Regulations Summary from January 1, 2001 to December 31, 2018 (41 pages)
- NEFMC 2020. Northeast Multispecies Fisheries Management Plan, Framework Adjustment 59 (132 pages)
- Description of New England and Mid-Atlantic Region Stock Assessment Process (16 pages)
- ICES. 2020. Workshop on Catch Forecast from Biased Assessments (WKFORBIAS; outputs from 2019 meeting). ICES Scientific Reports. 2:28. 38 pp. http://doi.org/10.17895/ices.pub.599

Presentations given during the review
Introduction to the Index Based Methods Research Track

- TOR 1 Make Data
- TOR 2 Pick IBMs
- TOR 3 Select Metrics
- TOR 4 Crank Sims
- TOR 5 Advise
- TOR 6 Ref Points
- Homework Day 1
- Homework Day 2
- A Bayesian State-Space Approach to Improve Biomass Projections for Managing New England Groundfish

Appendix VII-2. List of Participants

Index Based Methods Peer Review Meeting
December 7-11, 2020
Attendance

Acronyms
NEFSC - Northeast Fisheries Science Center
SMAST - University of Massachusetts, School of Marine Science and Technology
NEFMC - Northeast Fisheries Management Council
MAFMC - Mid Atlantic Fisheries Management Council
MADMF - Massachusetts Division of Marine Fisheries
RIDEM - Rhode Island Department of Environmental Management
GARFO - Greater Atlantic Regional Fisheries Office

Paul Rago - Chair
Paul Medley - CIE Review Panel
Robin Cook - CIE Review Panel
Yong Chen - CIE Review Panel

Andrew Jones - NEFSC
Brandon Mufflley - MAFMC
Brian Linton - NEFSC
Brian Stock - NEFSC
Burton Shank - NEFSC
Charles Adams - NEFSC
Charles Perretti - NEFSC
Chris Kellogg - NEFMC
Chris Legault - NEFSC
Chris Tholke - NEFSC
Corinne Truesdale - RIDEM
David Richardson - NEFSC
Deb Lambert - NOAA Fisheries HQ
Gavin Fay - SMAST
Jamie Cournane - NEFMC

Jennifer Couture - NEFMC
John Wiedenmann - Rutgers University
Jon Deroba - NEFSC
Karen E Greene - NOAA Fisheries HQ
Kathy Sosebee - NEFSC
Kelly Whitmore - MADMF
Kiersten Curti - NEFSC
Larry Alade - NEFSC
Liz Brooks - NEFSC
Liz Sullivan - GARFO
Mackenzie Mazur - Gulf of Maine Research Institute
Mark Grant - GARFO
Mark Terceiro - NEFSC
Mike Simpkins - NEFSC
Michele Traver - NEFSC
Paul Nitschke - NEFSC
Quang Huynh - University of British Columbia
Robin Frede - NEFMC
Russ Brown - NEFSC
Steve Cadrin - SMAST
Susan Wigley - NEFSC
Tim Miller - NEFSC
Tom Nies - NEFMC
Toni Chute - NEFSC
Tony Wood - NEFSC
Tyler Pavlowich - NEFSC

Appendix VII-3. Performance Work Statement

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

Index Based Methods and Harvest Control Rules
Research Track Peer Review

Dec. 7 -11, 2020

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

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

assessments are adequate to serve as a basis for 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 index based stock assessment methods and harvest control rules. The requirements for the peer review follow. This Performance Work Statement (PWS) also includes: **Appendix 1**: TORs for the research track, which are the responsibility of the analysts; **Appendix 2:** a draft meeting agenda; **Appendix 3:** Individual Independent Review Report Requirements; and **Appendix 4:** Peer Reviewer 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 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 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 use and application of both index-based and age-based stock assessment models, including familiarity with retrospective patterns and how catch advice is provided from stock assessment models. In addition, knowledge and experience with simulation analyses is required

Tasks for Reviewers

- Review the background materials and reports prior to the review meeting
    - Two weeks before the peer review, the Assessment Process Lead will electronically disseminate all necessary background information and reports to the CIE reviewers for the peer review.
- 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 Peer Reviewer Summary Report
- Deliver individual Independent Reviewer Reports to the Government according to the specified milestone dates

- This report should explain whether each research track Term of Reference 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 and research topics 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 Peer Reviewer Summary Report on specific stock assessment Terms of Reference or on additional questions raised during the meeting.

Tasks for Review panel

- During the SARC meeting, the panel is to determine whether each research track Term of Reference (TOR) 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 research track TOR.
- 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, and 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 the Report of the Index Based Methods and Harvest Control Rules Working Group.

The SARC Chair, with the assistance from the reviewers, will write the Peer Reviewer Summary Report. Each reviewer and the chair will discuss whether they hold similar views on each research track 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 Peer Reviewer 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 Peer Reviewer 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 Peer Reviewer 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 research track Term of Reference, either as part of the group opinion, or as a separate minority opinion. The Peer Reviewer Summary Report will not be submitted, reviewed, or approved by the Contractor.

Place of Performance
The place of performance shall be held remotely, via Google Meets video conferencing.

Period of Performance
The period of performance shall be from **01 November 2020 through 31 January 2021**. Each reviewer's duties shall not exceed **14** days to complete all required tasks.

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

Schedule	**Deliverables and Milestones**
Within 2 weeks of award	Contractor selects and confirms reviewers
Approximately 2 weeks later	Contractor provides the pre-review documents to the reviewers
December 7-11, 2020	Panel review meeting
Approximately 2 weeks later	Contractor receives draft reports
Within 2 weeks of receiving draft reports	Contractor submits final reports to the Government

* The Peer Reviewer 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

No travel is necessary, as this meeting is being held remotely.

Restricted or Limited Use of Data

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

NMFS Project Contact
Michele Traver, NEFSC Assessment Process Lead
Northeast Fisheries Science Center
166 Water Street, Woods Hole, MA 02543
Michele.Traver@noaa.gov
Phone: 508-495-2195

Appendix 1. Index Based Methods and Harvest Control Rules Research Track Terms of Reference and Background

1. Develop methods to create data that if assessed with standard age-based approaches (e.g., VPA or ASAP) could exhibit a strong retrospective pattern.
2. Identify a number of index-based methods and a range of harvest control rules for use in closed-loop simulation, using index-based data resulting from ToR 1.
3. Identify metrics from the index-based assessment results that could be used in evaluations of trade-offs in performance among harvest control rules and index-based methods.
4. Evaluate the combinations of index-based methods and control rules using the metrics in ToR 3 to determine candidates for consideration by the Councils or other management authorities.
5. Provide guidance on specific situations that are and are not well-suited for a particular control rule or index-based method identified in ToR 4.
6. Create guidelines for setting biological reference points for index-based stocks.

Background

There are two reasons stock are assessed with index-based approaches. Either the data are not available to support an age-based assessment, e.g., ocean pout, or the age-based assessment was rejected and replaced by an index-based approach, e.g., Georges Bank yellowtail flounder. In recent years, the number of index-based assessments due to the latter reason has increased. This research track is focused on how to deal with this situation because the presence of a strong retrospective pattern is an indication of an inconsistency in the data and model that prevents standard simulation testing approaches to be used.

The Councils are charged with setting harvest control rules for each stock. The work conducted during this research track is meant to inform this decision by testing a range of harvest control rules against simulated data that would generate strong retrospective patterns in an age-based assessment.

Many of the index-based approaches currently used do not have the ability to generate biological reference points because they do not have an underlying population dynamics model. The creation of reference points for such situations requires expert knowledge about the fish and fishery. The guidelines created to address ToR 6 cannot be formulaic because of this dependency. Instead, the guidelines can be considered more of a checklist of items to consider when setting the biological reference points for a particular stock. The National Standard 1 technical guidance working group (subgroup 1) will provide some of the information to support this effort.

Simulation will be the approach used to address the ToR. If time permits, historical data may be used to see how the catch advice resulting from any recommended harvest control rules compares to what was used, particularly for situations where retrospective adjustments were made to analytical models in the past. The most recent data for any stock will not be used to prevent the creation of a "new" assessment that could require action by a Council.

Index-based approaches can be more impacted by missing survey data than age-based assessments, in some situations. This research track is not intended to examine the challenges associated with missing or partial survey data, or any other logistical issues associated with the generation of an index to be used.

SAW Research Track TORs:

General Clarification of Terms that may be used in the Research Track Terms of Reference

Guidance to SAW Research Track Working Group about "Number of Models to include in the Peer Reviewer 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 Research Track 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. Draft Review Meeting Agenda

Index Based Methods and Harvest Control Rules
Research Track Assessment Peer Review Meeting

December 7 – 11, 2020

Google Meet link: TBD
Phone: TBD

DRAFT AGENDA

**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.*

Monday, December 7th, 2020

Time	Topic	Presenter(s)	Rapporteur
1:00 – 1:30pm	Welcome/Description of Review Process Introductions/Agenda/Conduct of Meeting	Michele Traver, Assessment Process Lead TBD, Chair	
1:30 – 3:00pm	TOR #1	Chris Legault, IBMWG Chair	TBD
3:00 – 3:15pm	Break		
3:15 – 4:15pm	TOR #1 cont.	Chris Legault, IBMWG Chair	TBD
4:15 – 4:45pm	Discussion/Review/Summary	Review Panel	TBD
4:45 – 5:00pm	Public Comment	Public	TBD
5:00pm	Adjourn		

Tuesday, December 8th, 2020

Time	Topic	Presenter(s)	Rapporteur
8:30 – 8:45am	Welcome/Logistics	Michele Traver, Assessment Process Lead TBD, Chair	
8:45 – 10:15am	TOR #2	Chris Legault, IBMWG Chair	TBD
10:15 – 10:30am	Break		
10:30 – 11:30am	TOR #2 cont.	Chris Legault, IBMWG Chair	TBD

11:30 – 12:00pm	Discussion/Review/Summary	Review Panel	
12:00 – 12:15pm	Public Comment	Public	
12:15 – 1:15pm	Lunch		
1:15 – 3:00pm	TOR #3	Chris Legault, IBMWG Chair	TBD
3:00 – 3:15pm	Break		
3:15 - 4:15pm	TOR #3 cont.	Chris Legault, IBMWG Chair	TBD
4:15 – 4:45pm	Discussion/Review/Summary	Review Panel	TBD
4:45 – 5:00pm	Public Comment	Public	TBD
5:45pm	Adjourn		

Wednesday, December 9th, 2020

Time	Topic	Presenter(s)	Rapporteur
8:30 – 8:45am	Welcome/Logistics	Michele Traver, Assessment Process Lead TBD, Chair	
8:45 – 10:15am	TOR #4	Chris Legault, IBMWG Chair	TBD
10:45 – 10:30am	Break		
10:30 – 11:30am	TOR #4 cont.	Chris Legault, IBMWG Chair	TBD
11:30 – 12:00pm	Discussion/Review/Summary	Panel	TBD
12:00 – 12:15pm	Public Comment	Public	TBD
12:15 – 1:15pm	Lunch		
1:15 – 3:00pm	TOR #5	Chris Legault, IBMWG Chair	TBD
3:00 – 3:15pm	Break		
3:15 – 4:15pm	TOR #5 cont.	Chris Legault, IBMWG Chair	TBD
4:15 – 4:45pm	Discussion/Review/Summary	Review Panel	TBD
4:45 - 5:00pm	Public Comment	Public	TBD
5:00pm	Adjourn		

Thursday, December 10th, 2020

Time	Topic	Presenter(s)	Rapporteur
8:30 – 8:45am	Welcome/Logistics	Michele Traver, Assessment Process Lead TBD, Chair	
8:45 – 10:15am	TOR #6	Chris Legault,	TBD

		IBMWG Chair	
10:45 – 10:30am	Break		
10:30 – 11:30am	TOR #6 cont.	Chris Legault, IBMWG Chair	TBD
11:30 – 12:00pm	Discussion/Review/Summary	Panel	TBD
12:00 – 12:15pm	Public Comment	Public	TBD
12:15 – 1:15pm	Lunch		
1:15 – 2:15pm	Discussion of Key Points	Review Panel	TBD
2:15 – 5:00pm	Report Writing	Review Panel	
5:00pm	Adjourn		

Friday, December 11th, 2020

Time	Topic	Presenter(s)	Rapporteur
8:30 – 5:00pm	Report Writing	Review Panel	

Appendix 3. Individual Independent Peer Reviewer Report Requirements

1. The independent Peer Reviewer 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 Peer Reviewer 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 Peer Reviewer 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 Performance Work Statement
   Appendix 3: Panel membership or other pertinent information from the panel review meeting.

Appendix 4. Peer Reviewer 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 /research topic reviewed, the report should address whether or not each Term of Reference of the Research Track Working Group was completed successfully. For each Term of Reference, the Peer Reviewer 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 Peer Reviewer Summary Report, along with a copy of the CIE Performance Work Statement.

   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.