Center for Independent Experts Independent Peer Review of the Index Based Methods and Harvest Control Rules Research Track

Paul A H Medley

January 2021

1 Executive Summary

- The research addresses the general use of Index Based Methods (IBM) in response to the problem of retrospective patterns in age-based assessments.
- All six terms of reference defining the research project were addressed, but ToR 5 and 6 were incomplete and all could benefit from further work.
- ToR 1 – 4 were met because the project set up and implemented efficient simulations testing a wide range of index-based methods by generating metrics measuring performance.
- ToR 5 was addressed but would benefit from another research cycle based on the results from this research. This would include adjusting the scope of the scenarios considered, the IBMs tested and improving ways to evaluate performance metrics that would provide better guidelines on how to use IBMs or other approaches to deal with retrospective patterns robustly.
- For ToR 6, developing guidelines on biological reference points, it is not clear that the current approach will resolve this. It is not necessarily related to retrospective patterns and might be best addressed in a separate project.
- The research showed that IBMs do not automatically perform better than retrospective-bias-corrected stock assessment and are not necessarily more robust or more likely to achieve desired management outcomes.
- My main recommendation is to continue the work. Before embarking on more research, results so far should be used to adjust the terms of reference and scope. Critical to this might be decisions on which candidate IBMs might be taken forward and how metrics might be better used to summarize IBM performance.
- This independent report provides my findings, conclusions and recommendations, and can be read independently of the separate Review Panel's summary report.

2 Background

The National Marine Fisheries Service (NMFS) science products may require independent scientific peer reviews to ensure their credibility. In this case three qualified experts were employed to review scientific research exploring the effect of retrospective patterns on index-based methods (IBM). The experts were employed as part of the Center for Independent Experts (CIE) program and charged with conducting their peer review impartially, objectively, and without conflicts of interest. The reviewers were independent of the development of the science, and without influence from any position that the agency or constituent groups may have. Further information on the CIE program may be obtained from www.ciereviews.org.

This peer review differs from most conducted by the CIE for the NEFSC in that it is of a comprehensive research project. The research addresses appropriate responses to retrospective patterns in age-based assessments generally rather than being specific to an individual stock assessment.

Retrospective patterns have led to the rejection of several statistical catch-at-age (SCAA) assessments that were replaced by index-based methods (IBM) as a result. However, whether these have led to better management advice has not been evaluated until now.

Retrospective patterns are the result of one or more underlying model components that are assumed to be stationary in time and appropriately measured, but in reality have changed or are changing. Broadly, stock assessment tries to match changes in relative abundance against mortality. Problems arise when mortality (natural mortality or catches) or the change in relative abundance (survey catchability or selectivity) is mis-specified creating errors that propagate through time so that the cohort-based model must constantly re-adjust. The problem is particularly acute for short term projections which are important for providing management advice. While age-based models can detect such problems, they can neither detect the source of the problem nor provide a consistent way to deal with it.

The Index Based Methods and Harvest Control Rules Research was designed to address a set of six Terms of Reference specified by the Northeast Regional Coordinating Council (NRCC). The terms of reference required that the working group set up a simulation platform to apply assorted IBM and generate performance metrics to provide guidelines on choosing appropriate IBM and suitable biological reference points. This review was conducted after only 8 months work into the research project.

3 Review Activities

The peer review was conducted remotely due to the COVID-19 epidemic from December 7 to 11, 2020. The meetings were held from 8:00am to 3:00pm EST from Monday to Wednesday to review the working group's progress on each the terms of reference. The review panel consisted of Yong Chen, Robin Cook, Paul Medley, Paul Rago (chair), and the work was conducted by the Working Group chaired by Chris Legault (see

Appendix 3. Panel Membership).

The review material was provided remotely via Google Drive well before the meeting took place (Appendix 1. Bibliography) so that it could be read before the meeting. On each day, Monday 7 – Wednesday 9 December, the working group lead (Chris Legault) made a presentation explaining how they had attempted to address the ToR. There was an opportunity to ask questions by the review panel and any stakeholders present at the open meeting. Some additional minor work was requested in the form of additional graphs and calculations based on the outputs. All outputs and code were provided, so I was also able to explore some results myself, which helped my understanding of what was done. On Thursday December 10, the panel prepared a PowerPoint presentation summarizing the key findings and recommendations of the Review Panel. Finally, on December 11 the Panel met briefly to discuss the first draft of the report, overall organization of the report, and timeline for delivery to the NEFSC.

All sessions were open to the public and no technical problems of note were encountered over the course of the meeting except the meeting software (Googlemeet) shared screen displays worked poorly in some browsers. As the presentations were also available on the shared drive, this did not present an insurmountable problem.

The review Panel Chair prepared the Panel Summary Report with assistance from the CIE Panelists. The NEFSC edited the report for factual inaccuracies but did not alter its content or conclusions. Additionally, each of the CIE Panelists drafted and submitted an independent reviewer's report to the Center for Independent Experts. This independent report provides my findings, conclusions and recommendations, and can be read independently of the Review Panel's summary report. I had no disagreements with the summary report, and these findings are consistent with it.

4 ToR 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.

ToR 1 was clearly met by the working group. The general set up of the model and assumptions to simulate the likely causes of retrospective patterns in stock assessments was reasonable and justified. A model was set up with the natural history characteristics of groundfish which make up most important fisheries in the region. Two of the likely main causes of retrospective pattern, mis-recorded catch and changing natural mortality, were used to generate data sets which exhibited similar retrospective patterns observed in stock assessments.

The simulations were based upon the Woods Hole Assessment Model (WHAM), which is a general state-space age-structured stock assessment model. We were not asked to review this model, but it is worth noting that state-space models provide a much better framework for modelling different types of errors. Among other things, WHAM allows recruitment and numbers-at-age to be treated as latent variables with explicit process errors. This is more flexible than treating these variables as having observation error only, and is particularly valuable in estimating errors in projections. Because this model is able to account for major attributes in population change, it is appropriate for simulating data suitable for testing index-based models.

Different scenarios were used to bracket the likely range of variation encountered in stock assessments in the region. Scenarios covered two sources of retrospective error, two fishing pressure histories, and two selectivity patterns. The modelling platform was flexible and able to apply the wide range of IBM as well as a statistical catch at age model to the simulated data. The main limitations on what was considered was the lack of time in conducting the research.

In terms of the causes of retrospectives, exclusion of survey catchability limited the range of the analyses. The causes of retrospective patterns are not altogether clear, but are the result of model misspecification (structural error) or missing catch data. The most common causes are often thought to be missing catch, changing natural mortality and/or changing the abundance index catchability (q). The scenarios covered under-reported catch and changing natural mortality only.

The scenarios used (two sources of retrospective pattern, change in fishing mortality, change in selectivity) were representative of likely variation among stocks, but not exhaustive. Another variable, the catch advice multiplier, was an addition to the IBM and probably should be an additional parameter for the IBMs tested.

This project was in a large part management of multiple simulations distributed among users. The approach, using GitHub and Google Drive, provided an excellent solution from which much could be learnt. It is difficult to see how these sorts of simulations can be conducted without a similar coordinated approach. Reusing the same random number sequence for each IBM reduced the number of simulations required to draw inferences, so the simulations were efficient. The way in which the work was conducted generally enabled the working group to address all terms of reference within a very limited time.

All code was made available through GitHub. This not only provided a way to manage the code and simulations for multiple users, but was valuable for the review as the code and results could be accessed directly. This helped a lot with understanding what had been done.

5 ToR 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.

ToR 2 was met by the working group with the caveat that not the full range of index-based methods were tested. A wide range of index-based methods (IBMs) were considered and the set chosen was a good representation of the available types. IBMs can be categorized by the data they use as well as the way they use that data. Data used by IBMs are partial compared to full stock assessment, using one or two of catch, abundance indices, age or length data. Most IBMs work by adjusting exploitation levels based either on fixed reference points or on recent trends in indices. The argument for IBMs is usually that they are more robust than full stock assessment because they either do not rely on the more dubious data or their assumptions are less critical to the results. However, none of these arguments for IBMs are necessarily true and therefore this research is useful in evaluating them.

IBMs that were not examined include IBMs that use catch-only or length data, require full history of fishery development or current stock status, or are based on surplus production models. Rejecting IBMs that require full catch-history, current stock status or rely on catch-only data was reasonable since these have requirements that cannot be met. Exclusion of IBMs that use surplus production models (SPM) is less well justified, and there may be an argument to include representative IBMs or even a biomass dynamics state space model approach as an alternative. It is not clear that length-based or SPM-based IBMs would work any better or worse, but they are still worth evaluating. Overall, the choices made were reasonable given the lack of time.

It would be worth comparing attributes of the IBMs against suspected problems in data or population dynamics when arguing that they are more robust in any particular case. In this sense, this research project was incomplete. The lack of consideration of all causes of retrospective error, most importantly changes in survey q, implied that IBMs were not tested against all likely errors. Results suggested that the two causes of the retrospective error considered were a dominant factor in IBM performance and whether IBMs using survey data might have been more (or less) affected by retrospective caused by catchability change would be of interest.

There are several IBMs which are based on surplus production models, including state space models, which only require time series of catch and abundance indices. These have been widely used and have an advantage in generating biological reference points (see ToR 6). As someone who relies of biomass dynamics for many data limited stock assessments, I agree in general with the skepticism over their use by default, but examples included in this evaluation would be useful as they may perform as well or better under some circumstances than the tested IBM.

There is also an argument for considering length-based IBM. Although these are not a priority for the stocks being considered for this research, length-based methods may be useful in the longer term for managing many non-allocated species for which age data are unavailable.

6 ToR 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.

ToR 3 was met, but additional metrics may still prove useful in further research. The metrics calculated covered a very wide range of possible performance indicators. The Review Panel, after noting that understanding so many metrics was difficult, recommended more. It was also pointed out that stakeholders have an interest in different metrics, so reducing the number calculated was not likely to be achieved, and instead the metrics need to be evaluated through different visualizations, multivariate reduction techniques or weighting metrics based on their importance.

All metrics were made available and could be examined using provided software in R. They were presented in various graphs and tables which provided a good summary of results, even if interpretation was still difficult.

Otherwise, the research project primarily relied on two approaches to simplifying simulation model outputs. Linear models were used to look at the impact of scenario attributes on each metric separately. The effects on each metric could be then compared in tables and visually. Secondly a scoring procedure was applied based on mean values or an alternative based on a non-parametric ranking. While both these approaches were helpful, the approaches used could benefit from more work.

It is difficult to evaluate so many metrics or provide a clear justification on any one IBM without explicitly weighting the various metrics' importance. Therefore, it will be worth further considering either how to weight and combine metrics into fewer variables, or identify a few metrics that capture the important information that can be used to evaluate performance.

There are a number of multivariate techniques that might be used to condense information in the metrics. On request, a correlation matrix of the metrics was provided that indicated strong linear correlation between many of them. I carried out a simple principal components analysis of the simulation metrics (excluding the SCAA metrics) that showed that the first component captured around 50% of the variance and the first 12 components cumulatively around 90% of the variance, while variance of 6 components was zero (Figure 1). While there is no harm in reporting all metrics, this suggests that there is room for significant metric reduction without much loss of information, and this might be achieved using further multivariate analyses.

Most results were reported in terms of the standard fishery performance indicators: Catch/MSY, F/F_{MSY} and B/B_{MSY}. These, by implication, were thought to be the most important. I would generally agree with this, and derived summary indices (such as the proportion of time across simulations the population spent in each quadrant, less than or greater than the MSY reference points for biomass and fishing mortality) provide a good summary of IBM performance. While it might be considered useful to report a wide range of metrics, not all have to be used in providing advice on IBMs to use.

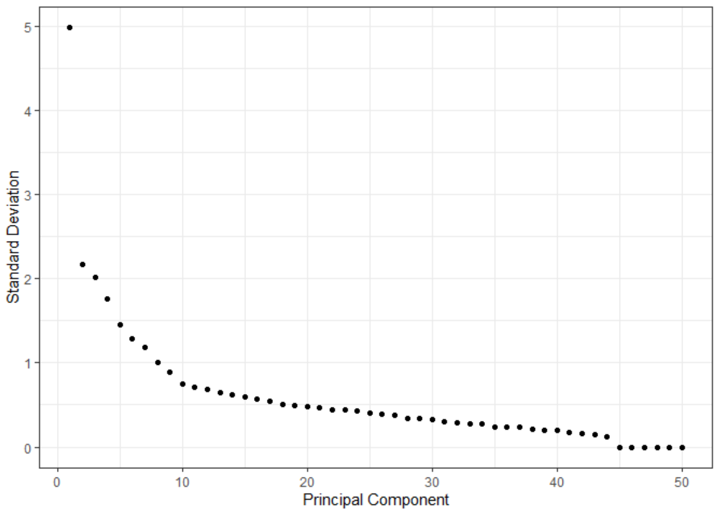

Figure 1 Standard deviations for the metrics principal components.

7 ToR 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.

ToR 4 was met by the working group. The project produced a target of 1000 simulations for each combination of IBM and historical scenario. This produced a large amount of complex output which was difficult to understand. The shiny-app developed to allow users to explore various measures of IBM performance was useful, but further support is likely required to understand the differences among IBMs because there are so many different combinations that can be examined.

Comparisons between simulations were made using linear models. Other multivariate techniques were not extensively used, although some similarity measures were produced between IBMs and the scenario attributes. The project made some progress in analyzing and visualizing the results, but further work in this area would be useful in simplifying the results, improving interpretation.

Further multivariate analyses might be used to reduce the number of IBMs being tested. As confirmation of the general working group results, a simple cluster analysis based on the combined metrics could be used to identify IBMs which produce similar outcomes on average (Figure 2). This suggests that there is an interaction between the types of data being used and the reference points (local slope vs. fixed values), but also that there may be scope to group similar IBMs and recommend choices from these groups, dependent on data and suspected model problems rather than any specific IBM. This could be used to reduce the number of simulations and the complexity of the results.

IBMs were not tuned during their application. In practice, such tuning might be carried out, but for the purposes here, to identify appropriate IBMs for use when a stock assessment with retrospective pattern, not adding this complication was justified. It does however mean that IBM evaluation could not be as rigorous as possible. While full tuning during the projection would be difficult, either alternative IBM parameters, or fitting parameters once at the end of the historical time series where that is appropriate, could be proposed as future work.

The simulation projection length and how this was dealt with seemed reasonable. Results can be sensitive to simulation time. A standard bio-economic approach, using discount rates, can lead to distortions of their own and high computing costs for little improvement in results. There is no reason to suppose that time boundaries chosen would cause problems, but it is worth bearing in mind that if the stock is overfished at the end of the simulation, it will incur future unaccounted-for costs.

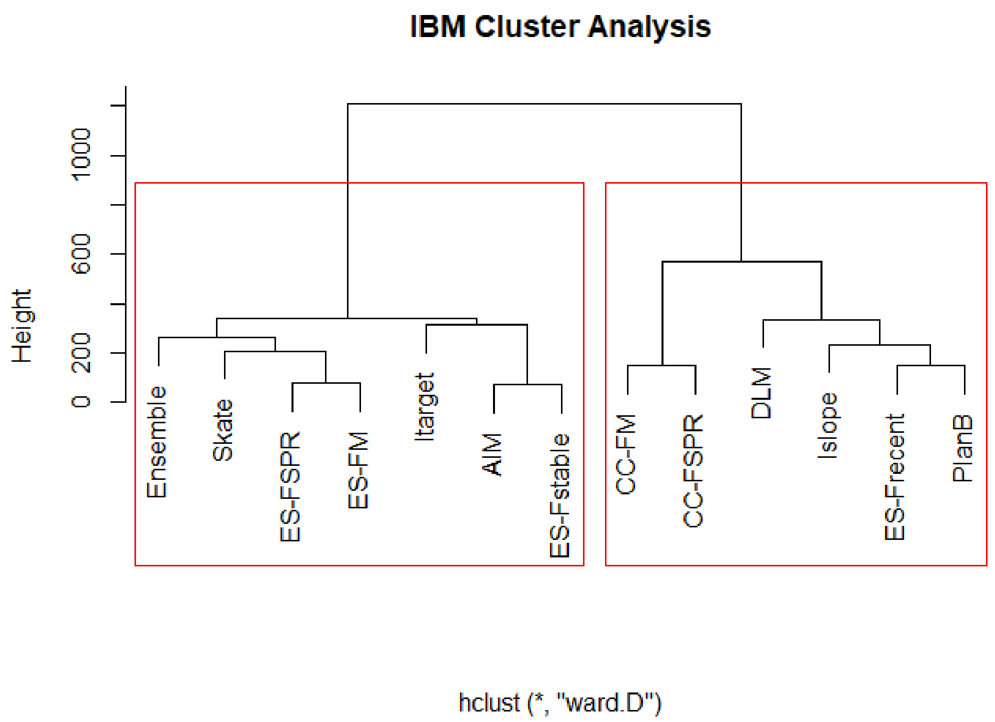

Figure 2 Cluster analysis of IBM using PCA 1-10 and default Wards method with a simple Euclidean distance measure.

8 ToR 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.

ToR 5 was addressed, but definitive guidance on a full range of scenarios was not achieved given time limitations. Guidance on choosing IBMs appropriate to particular situations was provided, but this did not cover all situations. For example, catch-curve IBM produced the lowest catches and were therefore the most precautionary, although the level of precaution could be adjusted through choice of IBM parameters.

However, it was found that a major factor affecting IBMs was the source of retrospective pattern. In reality, this source is not known, which makes choosing an appropriate IBM difficult. The working group recognized that further work might be required to complete this task.

The criteria for choosing an IBM were based on monotonically increasing variables ("more is better"), which was recognized as problematic. It is possible to apply constraints (e.g., $F>F_{MSY}$ is "bad"), so the results are still informative. However, this does not deal with the fundamental problem in fisheries science, which is managing the risk of overfishing against the need to maximize catch. This suggests that further consideration of performance measures will be necessary, for example, as suggested during the review meeting, of measures of overfishing/overfished risks for each IBM.

There will be a need to engage stakeholders (mainly managers) in this process, so visualization is important. Significant progress had been made in this regard and much of the output was accessible with a little help. It would still be useful to have workshops with managers to discuss metrics and measures of fishery performance. These do not necessarily have to come up with definitive solutions, but they do help clarify the problems for everyone involved in the decision-making process and can help guide in how to report the necessary trade-offs and risks in managing these resources.

It is not clear to me that any universal guidance, in the form of a decision tree for example, will be possible. It may never to be possible to simulate a full range of circumstances that lead to simplified results in which there can be much confidence. Even these evaluations have shown that the implications of retrospective patterns can be complex.

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

There was an attempt to address this ToR by the working group. This led to some general conclusions, but no guidelines, not least because it is unclear which IBM might be adopted in any particular case, and this determines whether biological reference points might be available.

It is not clear that this ToR can be met with a simulation research approach as used here. The ToR Simulations may show particular IBMs are robust (you make the right decision even if your information is poor), but they do not really address how to improve the stock assessment or thereby derive reference points. There has been a long period of research on retrospective patterns, but issues remain unresolved (ICES 2020). Therefore, it is difficult to see how biological reference points might be reliably set if the stock assessment is unable to provide a reliable estimate of biological reference points.

Whether biological reference points are available from the IBM depends upon how the IBM has been formulated. For example, age-based catch curves can use growth, length-weight and maturity models to calculate standard SPR reference points for the estimated fishing mortality, but an IBM based on abundance trends has no such reference points.

Surplus production models (SPM) are good candidates for estimated reference points, but were not tested. They can, for example, be tuned to the SCAA to produce a production curve that is maximized at the same exploitable biomass level, so could be made consistent with the original SCAA stock assessment. Whether this would be helpful in proposing reference points to determine stock status is unclear, however.

It may be necessary to use probability-based criteria for reference points because exact estimation using an IBM approach is unlikely to be reliable. An example would be using Bayesian state space biomass dynamics models to define reference points for survey indices. In this case, a simple stock assessment model might be used to define the value of the median survey index at MSY. However, this is a different procedure to the ones being tested through simulation and therefore outside the project's scope.

The preliminary results from these simulations suggest that it may not be possible to propose any general guidelines to address this issue. Therefore, if such guidelines are still sought, they may need to be addressed in a separate project which is not linked to dealing with retrospective patterns.

10 Conclusions

The simulation system and workflow were well set up and efficient, allowing the working group to address all Terms of Reference to some extent. However, the research will need more time to complete all tasks, including some of the recommendations from the review.

The research showed that IBMs do not automatically perform better than retrospective-bias-corrected stock assessment and are not necessarily more robust or more likely to achieve desired management outcomes. At the very least, this implies that any choice of IBM in place of a stock assessment with retrospective problems needs specific justification and cannot be assumed to resolve the issue.

IBMs have often been assumed to be more robust than full stock assessment models. However, in contrast to IBMs, age structured models are able to identify when a retrospective issue occurs, which indicates a problem either in the data or model. IBMs have limited ability to identify such errors, which can lead to underestimating uncertainty.

The last ToR requiring guidelines for setting biological reference points may be impossible to meet for all IBMs and is perhaps misplaced in this ToR set. It is not clear that, given a significant problem exists in a model or data, it will be possible to determine reliably whether overfishing is occurring or a stock is overfished. This may require, among other things, adjustment of reference points from absolute measures to a probability or risk-based measure, making it not just a scientific problem.

It is worth emphasizing the limitations of this research so far. There is a strong dependence of results on the underlying model assumptions, and the assumptions adopted by this model have not been exhaustive. The model does not depart from the standard age-structured population approach and was not intended to cover other assumptions (such as density dependent growth or mortality, or ecosystem effects). In addition, for simulations of this and more complex types, often the research may end up being a study of non-linear model dynamics without necessarily a strong link to the real world. It is therefore worth remembering that any result which is strongly dependent on an assumption that does not hold in the real world, has little value for the purpose of this study. Given this, it may be worth re-evaluating the ToRs so that they are more circumscribed to actual case studies rather than attempting to provide very general results.

11 Recommendations

My main recommendation is to continue the work. While remarkable progress has been made, the project seems unfinished. Before embarking on more research, results so far should be used to adjust the terms of reference and scope of the research. Critical to this might be decisions on which candidate IBMs might be taken forward and how the metrics might be better used to summarize IBM performance.

If more work of this type is to be carried out generally, it may be useful to review the suitability of available virtual research environments (VRE) software and cloud or server-based computing to conduct this sort of project. The ideas behind the approach developed for these simulations have been developed more formally in VRE. Regional fisheries management organizations have used these to carry out some stock assessment work because inter-sessional co-operation is otherwise difficult and expensive (see for example https://bluebridge.d4science.org/home). These provide cloud-based computing and shared environments, implementing GitHub-style control over changes to the shared files.

Cloud based computing might also have made this task easier, but would be more costly. Costs could potentially be reduced by cloud-based computing if the number of desktop computers and costs of computer maintenance were otherwise reduced, but this would need to be incorporated into long term budget planning. My experience running simulations on an

Oracle platform in RStudio worked well and was not overly expensive (compared to some alternatives I tried). A cloud-based solution would allow multiple simulations to be run quickly, but would incur a cost based primarily on CPU time.

In addition, automating the workflow, through an R package for example, could be helpful if simulations were to be used as a tool for standard stock assessments. As was used here, there are already "tidyverse" packages that are designed to help with this sort of procedure. The procedures developed for this research could be encapsulated in functions making their use easier in future.

In terms of research scope, I would recommend including changes in survey q and/or selectivity changes as sources of retrospective pattern. I understand why these were not included in the original scenarios, but given the importance of the retrospective source in these tests and that some IBMs rely on the survey data, it would make sense to include this issue in some form.

I would recommend expanding the IBMs tested to include representative surplus production model-based and length-based methods. In contrast, I would remove the catch advice multiplier scenario as it did not help much with the evaluation.

It would also be useful to apply some sort of tuning to some of the IBMs where appropriate to ensure they are offering their best performance. While full tuning during the projection is not recommended, adjusting the IBM parameters, or fitting parameters once at the end of the historical time series where that is appropriate, could be proposed as future work. For example, rather than include catch curve with F=M, the $SPR_{F=recent}$, $SPR_{20\%}$ and $SPR_{30\%}$ could be used as these should be directly comparable.

I would also consider adding a TAC variation constraint into the IBM rather than have it as a metric, which would also serve to simplify the outputs a little. This could be added as a parameter to all current IBM restricting the change in TAC to 5%, 10% or 20% for example.

The Dynamic Linear Model (DLM) approach appeared to have great potential. Methods that attempt to separate process and observation error are likely to have better performance overall. However, alongside the current highly efficient filtering method, a more general approach would be worth exploring where estimation could make use of other probability density functions than the Normal. This could be done in Stan (mc-stan.org), for example, which copes reasonably well with time series. This would not be as efficient and could have significant fitting problems, but would free up the model structure, so it could be seen whether current DLM assumptions make much difference.

In terms of simplifying the results, metrics might be summarized or the dimensions reduced using more multivariate analyses. Various multivariate methods that might be useful in simplifying metrics include factor analysis, to identify underlying latent variables which can be used to group and separate IBM candidates on performance. If new variables are constructed from originals, such as in PCA, their meaning can become obscure, however. To cope with this, they might be used to justify the selection of key representative original variables or develop new variables with a clear meaning, such as an "overfishing score". For example, multivariate analyses could be used to demonstrate whether the B/B_{MSY}, F/F_{MSY} and Catch/MSY metrics are sufficient in identifying an appropriate IBM or some other combination of derived variables are needed achieve this.

Another approach would be to explicitly separate constraints from the variable to be optimized. Only one variable can be optimized. So, in this case, the aim might be to maximize catch subject to an overfishing constraint. So, each candidate IBM parameters could be tuned to meet the constraint (e.g., so that only 1 in 20 years across simulations results in overfishing) or rejected if they cannot meet the constraint. A better IBM would then, otherwise, have the higher average catch. It should be considered adequate to achieve metaheuristic optimization over a wide set of random trials rather than some true optimum.

Another important way to process the large number of metrics would be to convert them explicitly to a single utility variable. There are several practical ways to do this, but most fall within the multi-criteria decision-making area of research. These might be considered here to help make the choice among IBMs.

All measures of outcome preference can be mapped to a single theoretical utility measure, which would allow direct assessment of preference among IBMs. Unfortunately, measuring utility is difficult; it is also subjective and may vary over time. Nevertheless, it may be worth considering a practical measure directly as another way to understand IBM performance and make recommendations.

If a more direct measure is desired, I have found the Analytical Hierarchical Process[1] (AHP) a useful technique because its foundations are simple to understand, which is important for purposes of transparency as well as application. AHP consists of structuring the problem in a hierarchy and using pairwise comparisons among elements in the hierarchy to construct scoring weights, so each hierarchical component can be scored using simple additive weighting. This allows a large number of attributes, in this case metrics, to be incorporated into a single score. For example, for the 50 metrics identified in this study, calculation of the weights would require 1225 comparisons, whereas grouping the metrics into 10 groups of 5, this would be reduced to 145 comparisons. This is practical, particularly considering a lot of comparisons could be done very quickly or would be redundant. However, it is still a lot of work and costly to do with stakeholders or managers. It is also risky, considering that managers may not collectively accept the metric weights and so not trust the final result. Different methods can give very different results (e.g., Zanakis et al. 1998) and it is therefore not clear that they are consistently measuring utility.

However, even if the scoring results are not accepted as a valid utility score, this process may still be worth working through. One of the advantages of AHP is it takes a complex preference problem and tries to break it down into simpler components which are more easily understood. This is much simpler and more practical than conjoint analysis, for example. The process of structuring the problem can be helpful in reaching a consensus solution (the choice of an appropriate IBM) even if there is no agreement that it is optimal.

Another approach to utility which may be more practical is to consider "regret" rather than "utility" directly. Considering regret may use current objectives to provide a reference for the relative utility measure. This is easier because the objectives can be considered separately. By way of example, a simple loss measure (R) could compare the catch limit set by the IBM compared to the desired catch in each year:

$$R = -\sum \frac{|C_y - C_{opt}|}{C_{opt}}$$

The desired catch could be calculated each year because knowledge is perfect within the simulation and any positive or negative departures from this optimum treated as a loss. The

[1] https://en.wikipedia.org/wiki/Analytic_hierarchy_process. AHP was originally proposed as a way to score sustainability in the Marine Stewardship Council assessments, although inevitably the methodology has moved a long way from this approach.

preferred IBM would minimize this loss over an appropriate time period, although catches higher than the optimum may make it harder to achieve the optimum in later years, so the period will need to be of sufficient length to take that into account. The definition of the desired catch (C_{opt}) could be based on MSY or some other desired exploitation rate, dependent on management objectives (NEFMC 2016).

Finally, given the difficulties in constructing scenarios which adequately cover the full range of possibilities that can give general guidance accepted by a review panel, the project could consider applying a series of case studies where the simulations match as far as possible particular stocks and fisheries. It seems likely that whatever scenarios are tested, they may never be enough, and new simulations with different assumptions could undermine any general findings. This research could be used to set up an approach (in the form of, for example, an R package) that can be tuned for particular stock assessments. Once the WHAM model is tuned to represent the observed retrospective pattern, covering the main likely sources of the pattern, then approaches can be tested on simulated projections and recommendations made in that case. This would make recommendations easier to justify as the range of scenarios bracketing the main uncertainty would be small. The current research indicates that the most important would be the possible retrospective sources (change in M, survey q or unrecorded catch).

Appendix 1. Bibliography

Documentation for the meeting was provided via Google Share Drive[2] and the NEFSC Data Portal[3]. All relevant source code and summary model outputs were available via GitHub[4].

The following report documents. The very extensive tables and figures were presented in separate documents (Appendix 6) and helpfully in a spreadsheet.

Original Documents (from Google Drive)

Readme.txt

1_Report Text.pdf

2_Report Tables.pdf

3_Report Figures.pdf

4_Appendices_1-5.pdf

5_Appendix_6.pdf

5_Appendix_6_part1.pdf

5_Appendix_6_part2.pdf

6_Report_Tables_Excel.xlsx

Review Meeting Presentations

All documents generated for the meeting were also providing during and after the meeting concluded.

70_Introduction.pptx

71_TOR_1_Make_Data.pptx

72_TOR_2_Pick_IBMs.pptx

73_TOR_3_Select_Metrics.pptx

74_TOR_4_Crank_Sims.pptx

75_TOR_5_Advise.pptx

76_TOR_6_Ref_Points.pptx

81_Homework_Day_1.pptx

83_Homework_Day_2.pptx

Score summary ranks.xlsx

Background Information

Groundfish Assessment Regulations Summary December 2018

Groundfish FW59 FINAL Affected Environment Excerpt

Stock assessment process June 2020

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.5997 Langan_IBMWG_Review.pdf

NEFMC 2016. Risk Policy Road Map. Report from the Risk Policy Working Group. June 2016. New England Fishery Management Council

Zanakis, S.H., Solomon, A. Wishart, N. Dublish, S. (1998) Multi-attribute decision making: A simulation comparison of select methods. European Journal of Operational Research 107(3) 507-529,https://doi.org/10.1016/S0377-2217(97)00147-1.

[2] https://drive.google.com/drive/u/0/folders/1VqyaTfGzod5rCwuqiHhJXW4C8hvJfhBt

[3] https://apps-nefsc.fisheries.noaa.gov/saw/sasi/sasi_report_options.php. (Assessment Year: 2020, Species Name: Index Based Methods, Stock Area: Georges Bank, Review Type: Research Track, Information Type: All)

[4] https://github.com/cmlegault/IBMWG

Appendix 2. 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[5]. 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 developing fishery management advice. Results provide the scientific basis for fisheries within the jurisdiction of NOAA's Greater Atlantic Regional Fisheries Office (GARFO).

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

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 the time of award through 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	**Milestones and Deliverables**
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. <u>Draft</u> 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, WG Chair	TBD
3:00 – 3:15pm	Break		
3:15 – 4:15pm	TOR #1 cont.	Chris Legault, WG 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, WG Chair	TBD
10:15 – 10:30am	Break		
10:30 – 11:30am	TOR #2 cont.	Chris Legault, WG 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, WG Chair	TBD
3:00 – 3:15pm	Break		
3:15 - 4:15pm	TOR #3 cont.	Chris Legault, WG 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, WG Chair	TBD
10:45 – 10:30am	Break		
10:30 – 11:30am	TOR #4 cont.	Chris Legault, WG 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, WG Chair	TBD
3:00 – 3:15pm	Break		
3:15 – 4:15pm	TOR #5 cont.	Chris Legault, WG 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, WG Chair	TBD
10:45 – 10:30am	Break		

10:30 – 11:30am	TOR #6 cont.	Chris Legault, WG 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.

Appendix 3. Panel Membership

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

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

Working Group (attendees)
Chris Legault - NEFSC (Chair)
Andrew Jones - NEFSC
Brandon Muffley - MAFMC
Jamie Cournane - NEFMC
John Wiedenmann - Rutgers University
Jon Deroba - NEFSC
Gavin Fay - SMAST
Tim Miller - NEFSC
Liz Brooks – NEFSC (honorary member)
Joe Langan - University of Rhode Island (honorary member)

Meeting Co-ordination and Management
Russell Brown, Chief, Population Dynamics Branch
Michele Traver, Assessment Process Lead