Independent Peer Review Report for the Research Track Assessment of American Plaice in the Gulf of Maine and Georges Bank.

WebEx video link
18 July – 21 July 2022

Prepared for Center of Independent Experts
by
Peter Stephenson

Email: peterstephenson48@gmail.com

Executive Summary

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.).

The Research Track Assessment peer review for American plaice in the Gulf of Maine and Georges Bank was conducted using WebEx video conferencing on 18-21 July 2022. A trial of WebEx was conducted on July 16. During the Review, the presentations were perfectly readable, and the voices and images of participants were clear.

The meeting was organized very well and was open and transparent. The Chair gave all participants ample opportunity to contribute and ask questions. The Review Panel reached consensus on all of the issues raised during the three-day meeting and the Panel Report was completed by the Chair and the three CIE reviewers.

There is extensive data on landings, abundance indices, and age composition data which were well described in the Working Paper documents and were presented in detail and with confidence to the Review Panel by the assessment team. The decisions to use the only the NEFSC survey data as an index of abundance was justified to the satisfaction of the panel. The ageing protocols were excellent, and the catch data I consider reliable.

Many different configurations of the stock assessment model were explored and the model fitted the data well and produced trajectories of spawning biomass and fishing mortality with sensible confidence intervals. I was convinced that the final model configuration was reliable. The reference points chosen are in common use and the results satisfied me that the fishery is neither overfished nor experiencing overfishing.

In this report, I have made suggestions for each of the Terms of Reference, but overall I found the assessment to be scientifically sound and reliable.

Background

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 peer review for the Research Track Assessment on American Plaice (*Hippoglossoides platessoides*) was held via WebEx video link between 18 July and 21 July 2022. The links provided for the participation meeting were efficiently organized and the technical side worked very well. There were occasions when the participants left their microphones turned on causing feedback noise and other times when the participants forgot to unmute before they started talking, but these were due to lack of experience and not a cause for concern.

The Working Group produced a comprehensive Summary Report supported by seventeen Working Papers for the Review Panel to examine before the meeting (Appendix 1). There were some minor problems with the link to the Working Papers, but this was rapidly resolved.

The Individual Independent Peer Reviewer Report Requirements are shown in Appendix 2. The agenda for the meeting is shown in Appendix 3 and the list of participants in this Research Track Review Meeting, including the Review Panel are shown in Appendix 4. The members of the Review Panel had diverse backgrounds and all made major contributions to the review process with their comments and questions.

The Chair was excellent and provided ample opportunity for comment and questions from the participants, but almost exclusively these came from the Review Panel. On the basis of the discussion and deliberation,

the Review Panel found that all the Terms of Reference had been satisfactorily addressed and that the WHAM model run 29F4 was found to be sound and it was endorsed. The Review Panel members provided input into the Panel Report which was initially compiled by the Chair and added to by the CIE Reviewers.

The Working Group delivered comprehensive presentations on the data inputs and the environmental impacts on the stock, indicating how each related to the Terms of Reference. Three assessment models were used for the assessment: Age Structured Assessment Program (ASAP, Legault and Restrepo 1998), Stock Synthesis (SS, Methot and Wetzel 2013), and the Woods Hole Assessment Model (WHAM, Stock and Miller 2021).

The access to the model inputs and outputs was not straightforward. I was given access to the Google Drive where everything is archived, but the huge amount of material was not what I needed. I have made suggestions on this in TOR 8.

The Working Group confidently presented detailed descriptions of the model inputs, the methods used in the three assessment model platforms, and model outputs. The Working Group quickly responded to all questions asked, and completed additional model runs, demonstrating their competence in using the software packages. After deliberation of all the material presented, I concluded that the assessment provided by the Working Group was scientifically sound and they had provided good evidence that stock the was not overfished nor undergoing overfishing. In addition, I endorse the use of WHAM in future assessments of this species.

This report represents the independent review by Peter Stephenson in accordance with the guidelines shown in the Performance Work Statement shown in Appendix 2.

Review Process

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.*

Prior to the workshop I read all the Data Workshop reports supplied for this review. I attempted to review the input files for the assessment model Stock Synthesis and found they were not in a format that was easily readable. I was familiar with SS and my brief examination indicated that the input data matched the information in the WP reports.

I participated in a trial run of WebEx on Friday 15 July at 7 pm with Russell Brown, Steven Holmes (CIE Reviewer) and Michele Traver, the WebEx organizer. The WebEx trial went smoothly. During the three meetings from 18 to 20 July from 7 pm to just after 12 pm, I actively participated in discussion on the details of the complex dynamics of the environment and its impact on the species, the results from the three assessment platforms, and additional analyses requested by the Review Panel. Through these discussions, I was satisfied that I could make a determination on the reliability of the assessment to determine the stock status of American plaice.

The review process was far from ideal. The WebEx process was expertly organized, and the necessary links working flawlessly. The CIE reviewers, the Review Chair were encouraged to join up on WhatsApp, which we did, and this was a useful means of communication. I have comments on the WebEx process below.

1. The time difference for me was +12 hours. The staff in United States kindly set the start time at 7 am, that is 7 pm for me. I felt that all the CIE reviewers participated well in the discussion and made valuable contributions, under difficult circumstances.
2. Running the meeting with WebEx meant, I believe, the quality of the result was not as good as it could have been with face-to-face contact. It is valuable to sit down with the assessment team out of session, before the meeting and during breaks to test out ideas, before discussion in front of the whole group. Using WebEx, there in a greater chance of comments being miss-interpreted and causing disruption or offence.
3. We had ample opportunity to ask questions of particular staff and mostly they were available, and able to turn on their camera and microphone and answer the query, but it was nowhere near as useful as face-to-face communication.
4. I hope, as soon as possible, NOAA will decide to bring back travel of CIE reviewers to United States thereby making the quality of the review as high as possible.

I have addressed each of the Terms of Reference listed below with my comments.

TOR 1. Identify relevant ecosystem and climate influences on the stock. Characterize the uncertainty in the relevant sources of data and their link to stock dynamics. Consider findings, as appropriate, in addressing other TORs. Report how the findings were considered under impacted TORs.

The Gulf of Maine and Georges Bank has complex environmental drivers which are likely to affect American plaice. The Working Group reported the results of an extensive literature review of the environmental influences, indicating the area of interest in this assessment has unusually high temperature increases.

Regression analyses looked at how environmental factors affect growth, recruitment, and the distribution of the stock. The analysis indicated that plaice move to deeper water in winter and more shallow water in spring/summer, but decadal increasing temperatures appear to be causing a general shift of plaice into deeper water. As a cooler habitat is preferred by this species, there appears to be a contraction of the stock.

Recruitment rate (recruitment per spawner) appears to be related to temperature, with highest values corresponding to years of low temperature. However, the recruitment rate is positively related to the temperature anomaly of the North Atlantic Oscillation (NAO).

In general, increasing temperature is associated with increased growth rates, younger age at maturity, and decreased body size of older fish. Larval plaice are likely to have reduced survival with increasing temperature. The data from the NEFSC indicates that the weight at age declined in the late 1990's through to the late 2000's and becoming more stable after this (Figure 1). The reason for this decline is not clear. The Working Group indicates it unlikely to be fish condition, more likely size at age is declining.

There was discussion about whether the decline in the weight at age it could be due to increased natural mortality of older fish, or could it be a change in the availability of the plaice stock to the survey.

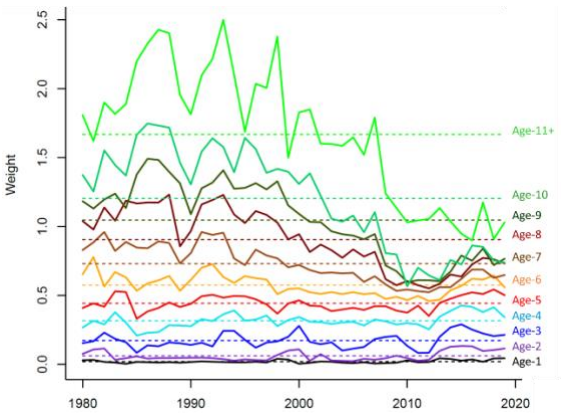

Figure 1. Weight at age of American plaice from 1980 to 2020

The population density was found to be related to bottom temperature and depth but not the Atlantic Multidecadal Oscillation (AMO). There is a seasonal North-East and South-West movement of the stock. This movement is likely to affect the availability of plaice to the surveys, especially the two State inshore surveys.

Recommendations

1. The centre of gravity of the plaice stock is changing with periods of North-East and South-West movement. Over the time series of the data, the extent of the stock has decreased. I recommend an examination of the fishery logbook data and the Federal survey data to better understand changes in the distribution of the plaice stock and how this is related to temperature and depth.

2. There was considerable discussion about year and age specific M, aiming to better capture the changes in size at age. I do not recommend this as the results would probably be difficult to interpret. I recommend that the natural mortality stay fixed, as it is in the current model formulation, with my preference being M=0.27.

3. The relationship between temperature and recruitment rate warrants further investigation to resolve the apparent contradiction of its relationship with bottom temperature and the Atlantic Multidecadal Oscillation (AMO). The relationship between water temperature and recruitment rate is likely to become increasingly important in future years. For the projections, the whole time series of recruitment is used. The working group should investigate a formulation that would take into account the likely influence of temperature on recruitment in future years.

TOR2. Estimate catch from all sources including landings and discards. Describe the spatial and temporal distribution of landings, discards, and fishing effort. Characterize the uncertainty in these sources of data.

The American plaice resource in United States is assumed to be a single stock separate from that in Canadian waters with a fairly continuous distribution of juveniles and adults in the Gulf of Maine-Georges Bank region with some regional differences. The juveniles and adults are quite sedentary, but they move to shallow spawning areas in winter and to deeper waters in summer.

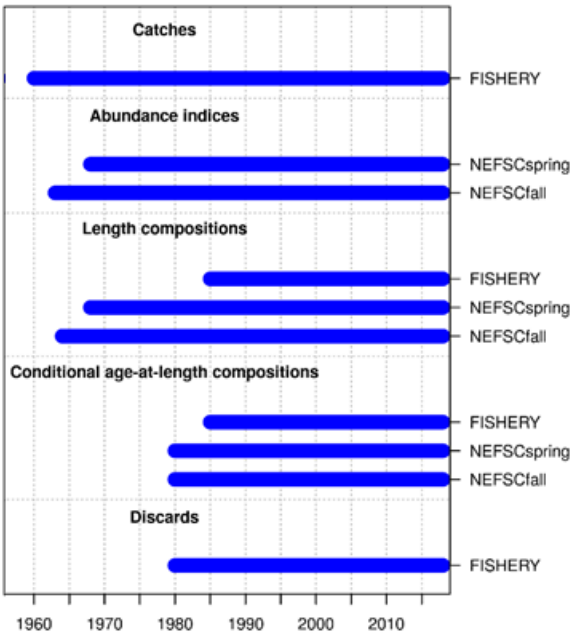

Figure 2. Time series of data used in the base case assessment model.

The time series of data from 1960-2019 are shown in Figure 2. From 1960-1994 the catch was determined from weights recorded by seafood dealers and apportioned to areas using port interviews with fishers. In 1994, a fisher's logbook system commenced which enabled spatial catch and fishing effort to be determined. The logbook catch is matched to the landings and currently 89% of the catch can be directly matched. The CV in the catch reconciliation process is estimated at less than 0.1 over the entire time series. Although there may be some unreported catch, it is believed to be very small. Misallocation of catch to another species has been identified but believed to be less than 1%, and misreporting the block where plaice were caught is also believed to be low. I am satisfied that the catches are well reported and the errors are well documented.

The catch is currently taken by the large mesh trawl fishery (about 90%), small mesh fishery, scallop, gillnet and shrimp fishery. The catch by the scallop, gillnet, and shrimp fishery have been decreasing over time due the changes in bycatch regulations and exclusion grids.

Although the stock appears to be moving to deeper water, the large mesh fishery has access to the stock in this deeper water. Ground fish catch has been recorded in the Gulf of Maine-Georges Bank region in the 1950's and also pre-1940.

The Working Group decided to use data from 1960 to 2019 for the current assessment. The landings for 2020 and 2021 were not included in this assessment but will be included in an updated assessment late in 2022. It is not ideal to assume the catch prior to 1960 was zero, but the WG decision was pragmatic and sensible given that the catches of this species prior to 1960 are believed to be small. I do not know how the catch CV's were incorporated in the assessment model.

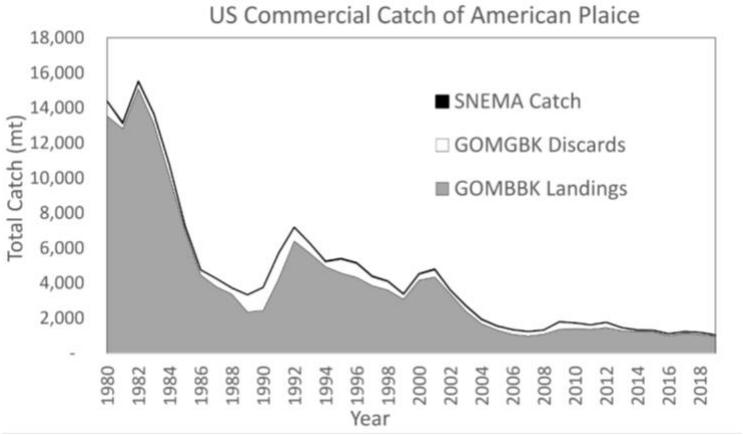

Figure 3. Landings and discards of American Plaice from 1980 to 2019

Discards

The information on the level of discards from the fleets prosecuting this species between 1980 and 2019 (Figure 3) is extensive and informative. Discards from the small mesh fishery, shrimp, gillnet, and scallop fleets have been decreasing due to the introduction of Nordmore grids in the shrimp fishery and changes in bycatch regulations. In the last 5 years, 80%-90% of the discards are from the large mesh fishery, with nearly all of the remainder from the small mesh fishery and the scallop fishery. For the last 5 years, the level of discards of American plaice has been at about 10% of the total catch of this species. The CV's are reported for the discards for all fishing sectors.

Observer coverage

The at-sea observer coverage on all sectors of the fishery was 10-20% from 2010-2019 and for the large mesh fleet the coverage was 16%-30% for the same period. After consultation with fishers and the observations of at-sea observers, the discard mortality of plaice is assumed to be 100%, which I believe is a reasonable assumption. An Electronic Monitoring Program commenced in 2018 with on-board cameras monitoring discard numbers and lengths. Currently fishers in the large mesh fishery are given the option of choosing the Electronic Monitoring Program using on-vessel cameras, or at-sea observers. It is planned that there will be 100% coverage of the large mesh fleet by the end of 2022.

Age-composition

The age and length composition data are available from 1980-2020 from factory samples and from at-sea observers. The otoliths collected by at-sea observers have not been processed.

The factories sort the catch into categories of Jumbo, Large, Medium, Small, and Pewee. The sorting is subjective, varying between factories and between individuals sorting the catch. Even though there is a price difference between the categories, all fish are available to research staff sampling the catch. The sorting of the catch does not appear to bias the sampling.

The age composition of the landings shows cohorts indicative of strong recruitments years, which are consistent with the survey data.

The age composition of the discards is determined using an age-length-key obtained from the NEFSC survey age composition. This is pragmatic and reasonable as the survey selects small plaice which would be discarded by the fleet. It would be more appropriate if the discard age composition could be derived from discards by at-sea observers, provided the otolith sample sizes are large enough to make this feasible.

The age composition for the fleet come from otoliths collected at the processors and are grouped into categories Q1, Q2, Q3, and Q4. In some years (e.g., 2018 and 2014) Q3 and Q4 were pooled due to low numbers of fish in some grading classes. The sample size of the age samples from the fleet were high (Table

1). The unofficial NAFO/ICNAF standard of sampling intensity (100 length samples per 200 mt catch) has been exceeded. In the last decade the sampling intensity has been more than twice this standard, with sampling for age composition following the same trend.

In the assessment process, the age composition data had an 11+ group for all fish over 10 years old. This is a realistic grouping as the numbers of 11+ fish is quite low. The landings at age were determined using age-length-keys with the bootstrap CV's generally <30% for ages 4-8 (majority of the catch) and up to 40% for the older fish.

I examined the ageing procedures and the excellent published manual, and I am convinced the procedures are current and best practice and include rolling reference samples (some institutions in the past have used fixed reference samples) and reconciliation of results. I was gladdened to read of the offer to lend otoliths to other institutions which is highly commendable.

Table 1. Number of otoliths ages from factory samples each year from 2014 to 2020.

Year	Number aged
2020	937
2019	1151
2018	1462
2017	2187
2016	2183
2015	1663
2014	1298

Standardized landings per unit effort (LPUE) was developed to provide a fishery dependent index of abundance. The covariates explored were fishing location, season, vessel size, depth, price, but not bottom temperature. LPUE showed a similar trend to that of the NEFSC survey index and when compared to the model estimated spawning stock biomass, there was evidence of a weak hyper stability, that is some evidence that the fleet catch rate remaining high when the stock size is declining. The standardized fleet LPUE was not included in the assessment model used for the current stock assessment.

Recommendations.

1. The otoliths collected by at-sea observers should be processed and aged if it is believed that they would provide better information on the age composition of discards than the currently used age-length-key from the NEFSA surveys.

2. It is not ideal to assume the catch prior to 1960 is zero, as done for exploratory Stock Synthesis runs. I recommend that an effort be made to determine an estimate of the catch of American plaice prior to 1960, together with any other relevant information for inclusion in the assessment process.

3. The fishery logbook program is time consuming for fishers and their efforts should be recognised. The development of standardized landings catch rate should continue, possibly including water temperature as a covariate. The LPUE should be given consideration in the assessment process, though not necessarily included in the assessment model.

4. The at-sea Observer Program and the Electronic Monitoring Program are valuable tools for the quantification of discards by the large mesh fishery. The aim for 100% coverage is commendable. I recommend that at-sea Observer Programme continue to be encouraged as this will provide the opportunity for generating discard age composition data.

5. The "best practice" procedures for fish ageing should continue. The poor agreement between the annuli count between readers is something we will probably need to live with. The annuli on American plaice have

been validated as annual growth rings, but not in this fishery. I have no problem with this, and it is pragmatic and reasonable to assume the annuli on American plaice are annual growth rings.

6. The ageing data collected from the fishery contained more old fish than that in the NEFSC survey. There was discussion about whether this was due to the spatial extent of the survey, shorter tows in the survey, or some other cause. I recommend this be further investigated.

TOR 3. Present the survey data used in the assessment (e.g., indices of relative or absolute abundance, recruitment, state surveys, age-length data, application of catchability and calibration studies, etc.) and provide a rationale for which data are used. Describe the spatial and temporal distribution of the data. Characterize the uncertainty in these sources of data.

There are six indices of abundance described in the document list. The methods used to determine these indices is well described and the statistical methods used are sound. In the Gulf of Maine and Georges Bank there are bi-annual federal surveys conducted by the Northeast Fisheries Science Center, NEFSC spring and NEFSC fall. These bottom trawl surveys are stratified random design and generally cover the area of the fishery. A complication with the NEFSC survey was the change of survey vessel from "Albatross" to "Bigelow" in 2009. In 2008, simultaneous surveys were conducted by the two vessels to determine a calibration factor in order to generate a single time series for fall and also for spring. The WHAM model fit suggested it was more appropriate to keep the two series separate, one for pre-2009 (Albatross) and another for the rest of the time series (Bigelow). In 2020, there was no survey due to Covid-19 restrictions.

There are two state surveys, Massachusetts Department of Marine Fisheries (MADMF) and the Maine-New Hampshire (ME-NH) each conducted in the inshore area of Gulf of Maine and during both spring and fall. Although American plaice otoliths were collected, they have not been processed. Thus, the NEFSC age-length-key was used for the state surveys.

In recent years the American plaice stock is moving offshore into deeper waters. This has been observed in the survey data and has been confirmed by fishers. It is probably in response to increasing water temperature. This has made the two state surveys less useful as an index of abundance. On the basis of the WHAM model fit, the state surveys were not included in the final stock assessment model configuration.

The sample sizes for otolith collection on the NEFSC surveys was large (Table 2). The age composition data indicated there were strong year classes in 1987, 1993, 2004, and 2013. These cohorts matched well with those evident in the landings age composition.

Table 2. Number of otoliths collected in the NEFSC spring and fall surveys.

Year	NEFSC survey spring	NEFSC survey fall
2020	0	0
2019	638	639
2018	477	500
2017	329	914
2016	1237	913
2015	1148	1037
2014	868	911

A model was developed to integrate the federal NEFSC surveys and the two state surveys MADMF and ME-NH, together with the landings per unit effort (LPUE), namely Vector Autoregressive Spatio-Temporal (VAST). Depth and bottom temperature were found to be significant covariates. The model found that the

centre of gravity of the plaice stock was changing, and the stock was moving into deeper water. The integrated abundance index from the VAST analysis was not used in the base case of the assessment model due to convergence problems.

In summary, the graphs of the selected indices appear consistent, having similar trends over time and I believe they support the assessment approach.

Recommendations.

1. The VAST model deserved further investigation to see if an integrated index of abundance can be used in the assessment model. The LPUE derived from fisher's logbooks should be incorporated if possible. The spatial distribution of the NEFSC surveys and location of catches from fisher's logbooks deserves further investigation to resolve issues like differences between logbook and survey age composition.

2. It would have been useful if there had been more discussion early in the review about inclusion and exclusion of indices in the base case assessment. The Review Panel would have benefited from the assessments team's views on the reliability and relevance of the four different indices, independent of the results of the WHAM model fit.

TOR 4. Use appropriate assessment approach to estimate annual fishing mortality, recruitment and stock biomass (both total and spawning stock) for the time series, and estimate their uncertainty. Compare the time series of these estimates with those from the previously accepted assessment(s). Evaluate a suite of model fit diagnostics (e.g., residual patterns, sensitivity analyses, retrospective patterns), and (a) comment on likely causes of problematic issues, and (b), if possible and appropriate, account for those issues when providing scientific advice and evaluate the consequences of any correction(s) applied.

The Working Group is to be commended for the huge effort put into developing stock assessment models in ASAP, SS, and WHAM and determination of the most appropriate base case for each model platform, and the detailed documentation shown in the Working Papers.

ASAP has been used for many assessments and has proved reliable. Stock Synthesis is used worldwide and has proved itself to be reliable in many stock assessments. WHAM, which is a space-state model, has the advantage that process error can be incorporated using "random effects".

The model runs in the three platforms, using similar data showed generally similar results for the derived quantities required for determining stock status, that is fishing mortality and spawning biomass. There was not a stipulation that the data input be identical for each model platform. For example, length composition was incorporated in Stock Synthesis but both ASAP and WHAM do not accommodate this.

The chosen software, WHAM uses the latest model development, space-state models. The background document for this assessment model, WP14, was detailed and comprehensive. The configuration of the WHAM base model was appropriate and the reasons for selection of particular data inputs was generally sound. The model produced sound results with realistic confidence intervals. The basic diagnostics such as plots of model fits and residual plots gave satisfactory results indicating the model is an informative assessment tool. Jitter analysis, whereby starting values were varied, resulted in model convergence for the base case model. The retrospective analysis showed consistency and did not indicate any reason for concern. I commend the use of WHAM for this assessment.

Recommendations

1. The process for determining the best configuration, including where to use random effects appeared to be on the grounds of model convergence, satisfactory residual patterns, the best fit to the data, and high retrospective consistency (Appendix 5).

I recommend, in future, groups of plausible configurations for the model setup could be made at the beginning of the modelling process and these explained to the Review Panel, with explanations of the suitability of this limited number of configurations. I believe it would have been more efficient to then start at WHAM run 29F4 and discuss the reasons for choices made, like choosing certain abundance indices or the formulation of selectivity, using random effects. The discussion of a limited number of model runs would have been enough to convince me that 29B4 was suitable for determining the status of the stock.

2. The parameter values for age composition selectivity in WHAM are very poorly estimated. The large variation of ages at length is always going to cause a problem. Stock Synthesis uses length composition selectivity, and the parameter estimates are more satisfactory. My preference would be to use a simple selectivity (flat top) unless there are logical reasons for doing otherwise. Obviously, a different configuration is appropriate for the discard selectivity. More investigation is needed into a WHAM selectivity estimation.

3. In the WHAM model, I believe it would be valuable to further the work on integration of environment factors into the assessment, for example recruitment rate and temperature change. This will be especially important in future assessments, as the environment changes.

4. There was considerable discussion on the choice of natural mortality, M. In the previous assessment, M=0.2 was used, as there was a very good estimate of M for American plaice in Canadian waters. It is known that in the Gulf of Maine and Georges Bank, the growth rate of this species is greater, and they mature at a younger age than in Canada. This information caused a re-examination of the estimate, based on maximum age and biological characteristics. The working group decided on a new estimate of M=0.27, which was rounded to one decimal place. When the WHAM model was run with values of 0.3 and also 0.2, the model diagnostics were better with the former and the derived values of spawning stock biomass and fishing mortality had a somewhat different trajectory, but the current status of the stock was not very different. At the requests of the reviewers, the WHAM model was re-run with M=0.27, and as expected the spawning stock biomass and fishing mortality changed very little. I believe the change to the larger value of M is sensible and justified, but I do not agree with the reasoning to round it to one decimal place and my preference is to use M=0.27.

I do not believe that using age and time varying M is a good idea. I think it is a sound idea to keep it simple and I am not convinced it is worth the effort of exploring these complex formulations whose impacts on the parameter estimates and derived outputs could be difficult to interpret.

5. In my experience, the choice of effective sample size can be important, as it can affect the weighting of the different data sets. In the Review Panel meeting, the effective sample size in the different model platforms received scant attention. In ASAP, an effective sample size of 30 was chosen. In the Stock Synthesis model, the McAlister Ianelli calculation was employed but I do not know the result of this formulation. The calculation of effective sample size used in Stock Synthesis is not appropriate for WHAM due to the incorporation of process error in the random effects. An effective sample size of 50 was used in the WHAM model. In future assessments, I think it would be beneficial to have discussion on the reasons for the choice of effective sample size in the WHAM assessment model and the impact on the model fit of varying the effective sample size.

6. The Working Paper 17 describes the Stock Synthesis model with the focus on base case run 14. The brief of the Working Group generates a good model configuration for the Stock Synthesis version, not to make it the same as WHAM run 29B4. A major difference is that Stock Syntheses uses length composition data directly and estimates selectivity at length, which makes a lot of sense as the fishing gear selects by length. The parameter estimates, plots of derived quantities and the confidence intervals were well described, and the spawning biomass and fishing mortality matched the results from the WHAM assessment quite well. The length selectivity gave sensible results with realistic confidence intervals.

I appreciate that the assessment team does not want to spend a lot of time on Stock Synthesis when it is not the preferred platform, however I recommend, though not a high priority, that time be allocated to develop a run in Stock Synthesis which closely matches the base case WHAM model.

7. I recommend that two WP documents, in Notebook or Excel, be produced showing the data, starter, and control file for the Stock Synthesis model, and also input data for the WHAM model. This would enable a reviewer to easily compare the model inputs and CVs with the data and CVs in the WP documents.

8. The WHAM model allows for incorporation of random effects. A model using random effects is generally more efficient than a fixed effects model, but random effects should only be incorporated after careful consideration in the model configuration phase of the assessment process. At the review meeting, there was no discussion about whether the incorporation of random effects in various sections of the model was appropriate, for example why it would be expected that the unobserved effects and the explanatory variables are uncorrelated. There was no mention of testing for this correlation using a Wu-Hausman statistic.

I recommend that, although use of random effects often leads to efficient estimation, it should not be used without careful consideration. A simple model is easier to understand and may lead to more consistent results in some circumstances.

9. The WHAM model run 29F4 has been endorsed by the Review Panel and the two years of data, 2020 and 2021 will be incorporated before the Management Track peer review meeting late in 2022.

For future assessments, I recommend the use of Supervised Learning. In Supervised Learning, taking American plaice as an example, the data is divided into a training set, say 1960 to 2014 (it should contain a reasonable number of years of the Bigelow survey data) and the test set, say 2015-2021. Now decide on the design considerations for the WHAM model and hyper parameters (like convergence criteria, fixed CV's, fixed M=0.27, steepness of S-R relationship h=0.61, etc.). The model is fitted to the training data in WHAM using these design considerations. The WHAM model "one-step-ahead" feature gives estimates of the derived parameters and their confidence limits.

The data in the test set is now added one year at a time and model refitted. The model fit is assessed with the usual diagnostics and the output of derived parameters compared. This fitting to the training and test data could be repeated for a limited set of predetermined design configurations (like selection of abundance indices, selectivity estimation, age composition likelihood) to decide on the optimum design configuration. For the final model, the training data and the test data are then be combined, with this optimum configuration.

In summary, the WHAM assessment model is a "state of the art" fisheries assessment tool which is flexible and can allow for many configuration options including environmental covariates. I recommend its use in this and future American plaice assessments.

TOR 5. Update or redefine status determination criteria (SDC; point estimates or proxies for BMSY, $B_{THRESHOLD}$, F_{MSY} and MSY reference points) and provide estimates of those criteria and their uncertainty, along with a description of the sources of uncertainty. If analytic model-based estimates are unavailable, consider recommending alternative measurable proxies for reference points. Compare estimates of current stock size and fishing mortality to existing, and any redefined, SDCs.

In this assessment, no stock-recruitment relationship was available and the commonly used proxies $F_{SPR40\%}$ and $SSB_{FSPR40\%}$ are used as proxies for F_{MSY} and B_{MSY}. In the base case WHAM assessment, calculation of these proxies used the whole time series of recruitment, the last 5 years of estimates of selectivity, and the observed weight at age, to calculate $SSB_{F40\%}$. I believe this procedure is sound and the estimates $F_{40\%}$ =0.43 and $SSB_{F40\%}$ =18,000 mt are reliable.

The model output graphs indicate the SSB has been increasing and fishing mortality decreasing. The confidence intervals appear realistic, and I support the conclusion that the American plaice stock is not overfished, and overfishing is not occurring. Given the stock status, the lack of a stock-recruitment relationship is not a concern.

Recommendation

The changes in bottom temperature in recent years and the likely changes in recruitment rate could be a problem in the future. If the stock size declines, the use of the current SDCs will need to be re-evaluated.

TOR 6. Define appropriate methods for producing projections; provide justification for assumptions of fishery selectivity, weights at age, maturity, and recruitment; and comment on the reliability of resulting projections considering the effects of uncertainty and sensitivity to projection assumptions.

The base case WHAM run produced short-term projections to 2022 using the model uncertainties in the model parameters, the 2020 abundance-at-age, and recruitment from the whole time series with associated process variance, the last 5 years of selectivity, and the constant weight at age, and constant maturity at age.

The projection used four scenarios: F at F40% OFL, 75%F40%=ABC, F2019=status quo, and F =0. The projections had consistency and the associated confidence limits appeared realistic.

Recommendation

1. The future increases in water temperature and the impact on recruitment rate in the projections, needs to be evaluated and a different configuration for recruitment may be required for the projections in future assessments.

2. When the base case WHAM model is re-run in 2022, the implications of the missing survey data for 2020 need investigation and reporting.

TOR 7. Review, evaluate, and report on the status of research recommendations from the last assessment peer review, including recommendations provided by the prior assessment working group, peer review panel, and SSC. Identify new recommendations for future research, data collection, and assessment methodology. If any ecosystem influences from TOR 1 could not be considered quantitatively under that or other TORs, describe next steps for development, testing, and review of quantitative relationships and how they could best inform assessments. Prioritize research recommendations.

In the Summary Report, there is a comprehensive compilation of the previous research recommendations. The only outstanding recommendation appears to be the age-reading of the samples in the state inshore surveys. This, I believe still remains a low priority.

Below are the new Research Working Group recommendations, in order of priority. The Review Panel endorsed these, and I concur.

1. Continue to monitor shifts in distributions of plaice, particularly depth and environmental covariates on catchability.
2. Exploration of spatiotemporal integration of federal and state surveys should continue.
3. Investments are needed to streamline the estimation of commercial catch and promote reproducibility of estimates.
4. Consider deriving discards from electronic monitoring when an integrated catch monitoring system is developed.
5. As the Gulf of Maine scallop fishery expands, it should be included in discard estimation.
6. Archived otolith samples should be processed (state surveys, at-sea observers, 1975-1979).
7. The relationship between recruitment and ocean temperature should continue to be monitored.
8. Time-varying natural mortality, possibly with environmental covariate should be explored.
9. Methods should be developed to compare models with and without environmental covariates.
10. If the proposed assessment approach does not meet the standards of peer review, an alternative model should be developed to integrate information from catch, age composition and indices.

The Review Panel suggested further recommendations. I have not reproduced them, but rather given my view on recommendations for future research, some of which appear similar to those of the Review Panel.

1. It is important that the abundance index derived from fishers' logbook data be further developed and included in the assessment process, but not necessarily the assessment model. The considerable work put in by fishers needs to be acknowledged in a visible way.

2. The VAST index should be further explored, including the landing abundance index. Further attempts should be made to incorporate it into the assessment model.

3. I endorse the continued development of Electronic Monitoring, but I suggest that the at-sea Observer Programme continues to operate as well because this provides the opportunity for collection of length data and more importantly otolith samples for the whole size range of the catch, including discards.

5. I believe a simple formulation for M, constant by age and year, has proved satisfactory in many assessments. I do not believe that a more complex formulation will help resolve the problems with selectivity or changes in weight at age. Again, I believe this should be decided up front rather than changing it to fix problems in model fit.

TOR 8. Develop a backup assessment approach to providing scientific advice to managers if the proposed assessment approach does not pass peer review or the approved approach is rejected in a future management track assessment.

The Working Group carefully considered a backup plan in case the proposed assessment, using base case WHAM, be rejected during the Peer Review or in Management Track Assessment. They concluded that index-based methods, like catch curve analysis, would not be satisfactory.

There was discussion about the use of a Stock Synthesis model as a backup assessment. The assessment team gave a convincing argument that in the time between the current assessment being rejected and the time when advice for Management was required, there would be insufficient time to get a Stock Syntheses model configured and fully tested.

The ASAP run 43 has a similar data setup to WHAM and it would be feasible to get it set up with the latest data and fully tested in a short time.

I endorse the Working Group proposal that ASAP should be used as the backup assessment model.

TOR 9. Identify and consider any additional stock specific analyses or investigations that are critical for this assessment and warrant peer review, and develop additional TOR(s)* to address as needed.

Recommendation.

The analyses presented in the background documents were comprehensive and I have no further analyses of the input data. My recommendation on a training set of data for the WHAM in a new analysis would be very time consuming, but still worthy of consideration.

Appendix 1. Summary Report and Working Papers prepared for the Data Workshop

Report	Report of the American Plaice Research Track Working Group. July 2022.
WP 1	Size distribution analysis of American plaice, by Tyler Pavlowich[1]. August 2021.
WP 2	Overview if American Plaice ageing in the Northwest Atlantic. NOAA Fisheries Service. Josh Dayton[1] and Eric Robillard[1]. 2022.
WP 3	Updating Parameters for Length and Weight Relationships of American Plaice. Ashley Silver[7], Tyler Pavlowich[1], and Larry Alade[1]. September 2021.
WP 4	Maturity Analyses of American Plaice in the Georges Bank and Gulf of Maine region. Shakira Goffe[7], Daniel Hennen[1], and Larry Alade[1]. September 2021.
WP 5	Fishing Industry Knowledge of American plaice. A working paper submitted to the American plaice research track stock assessment, 2022. Tyler Pavlowich[1], David Richardson[1], John Manderson[9], Greg DeCelles[8].
WP 6	Exploration of Fishery Data to Evaluate Catch Rates of American Plaice. Max Grezlik[4], Lucy McGinnis[4], Keith Hankowsky[4], Gavin Fay, Steve Cadrin[4], and Alex Hansell[1]. November 2021.
WP 7	Catch Rate Standardization of American Plaice Trawl Fishery. Keith Hankowsky[4], Max Grezlik[4], Lucy McGinnis[4], Gavin Fay, Steve Cadrin[4], and Alex Hansell[1]. February 2022.
WP 8	Fitting a geostatistical model sdmTMB to standardize the catch rates of American Plaice (*Hippoglossoides platessoides*) from the Gulf of Maine and Georges Bank. Andrew Jones[1], Tyler Pavlowich[1], David Richardson[1], Anna Mercer[1].
WP 9	Fishery Dependent Data Indices of Abundance (LPUE or CPUE) for American Plaice. Mark Terceiro[1]. November 2021.
WP 10	Audit Model Electronic Monitoring Data: American Plaice. Cate O'Keefe[5] and Mel Sanderson[6].
WP11	American Plaice Research Track Stock Assessment Working Group. Seasonal Variation in Size-at-Age of American Plaice from Survey Data. Steve Cadrin[4]. November 2021.
WP 12	Spatio-temporal dynamics of American plaice (*Hippoglossoides platessoides*) in US waters of the northwest Atlantic. Alexander Hansell[1], Larry Alade[1], Andrew Allyn[2], Lauran Brewster[3], Steve Cadrin[4], Lisa Kerr[2].
WP13	Relative efficiency of a chain sweep and the rock hopper sweep used for the NEFSC bottom trawl survey and biomass estimates for American plaice, by Tim Miller[1], David Richardson[1], Andrew Jones[1], and Phil Politis[1]. December 2021).
WP14	Ecosystem and Climate Influences, by Jamie Behan[8], Lisa Kerr[2], Amanda Hart[8], Alex Hansell[1], Tyler Paklovitch[1] and Steve Cadrin[4]. November 2021.
WP 15	Approximation of Natural Mortality Rate for American Plaice in US Waters Based on Life History Traits. Steve Cadrin[4]. February 2022.
WP 16	Environmental Influences on American Plaice Stock Dynamics. Jamie Behan[8] and Lisa Kerr[2].
WP 17	American Plaice Assessment Model Developed in Stock Synthesis. Dan Hennen[1] and Alex Hansell[1]. April 2022.
WP 18	A state-space assessment of American plaice using the Woods Hole Assessment Model (WHAM). Amanda Hart[2], Lisa Kerr[2], and Tim Miller[1]. June 2022.
WP 19	Fishery Data. Larry Alade[1].
WP 20	Survey data. Larry Alade[1].
WP21	Projections. Larry Alade[1].

Affiliation 1. Northeast Fishery Science Center, NOAA, Woods Hold, MA, USA.
2. Gulf of Maine Research Institute, 350 Commercial Street, Portland, ME, USA.
3. Florida Atlantic University, Fort Pierce FL, USA.
4. School for Marine Science and Technology, University of Massachusetts Dartmouth, New Bedford, MA, USA.
5. Fishery Applications Consulting.
6. Cape Cod Commercial Fishermen's Alliance.
7. University of Maryland Eastern Shore.
8. Global Risk Management Institute Inc., University of Toronto.
9. Open Ocean Research, New Jersey, United States.
10. Orsted North America, Boston 11. Fisheries and Oceans Canada

Appendix 2. Individual Independent Peer Reviewer Report Requirements

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

American Plaice Research Track Virtual Peer Review

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].

Scope

The Research Track Peer Review 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 research track peer review is the cornerstone of the Northeast Region Coordinating Council stock assessment process, which includes assessment development, and report preparation (which is done by Working Groups or Atlantic States Marine Fisheries Commission (ASMFC) technical committees), assessment peer review (by the peer review panel), public presentations, and document publication. The results of this peer review will be incorporated into future management track assessments, which serve as the basis for developing fishery management recommendations.

The purpose of this meeting will be to provide an external peer review of the American plaice stock. The requirements for the peer review follow. This Performance Work Statement (PWS) also includes: **PWS Appendix 1**: TORs for the research track, which are the responsibility of the analysts; **PWS Appendix 2:** a draft meeting agenda; **PWS Appendix 3:** Individual Independent Review Report Requirements; and **PWS 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

[1] https://www.whitehouse.gov/wp-content/uploads/legacy_drupal_files/omb/memoranda/2005/m05-03.pdf

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. Modifications to the PWS and ToRs cannot be made during the peer review, and any PWS or ToRs modifications prior to the peer review shall be approved by the Contracting Officer's Representative (COR) and the CIE contractor. All TORs must be addressed in each reviewer's report. The reviewers shall have working knowledge and recent experience in the use and application of index-based, age-based, and state-space 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 virtually 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 Peer Review Panel (co)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 peer review meeting, using the criteria specified below in the "Tasks for Peer Review 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 peer review 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 Peer Review Panel 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 Peer Review Panel chair and reviewers combined:
Review the Report of American plaice Research Track Working Group.

The Peer Review Panel 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 peer review meeting. For terms where a similar view can be reached, the Peer Reviewer Summary Report will contain a summary of such 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 their 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 WebEx video conferencing.

Period of Performance
The period of performance shall be from the time of award through September 2022. 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.

Within 2 weeks of award	Contractor selects and confirms reviewers
Approximately 2 weeks later	Contractor provides the pre-review documents to the reviewers
July 18-21, 2022	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

PWS Appendix 1. Generic Research Track Terms of Reference

1. Identify relevant ecosystem and climate influences on the stock. Characterize the uncertainty in the relevant sources of data and their link to stock dynamics. Consider findings, as appropriate, in addressing other TORs. Report how the findings were considered under impacted TORs.

2. Estimate catch from all sources including landings and discards. Describe the spatial and temporal distribution of landings, discards, and fishing effort. Characterize the uncertainty in these sources of data.

3. Present the survey data used in the assessment (e.g., indices of relative or absolute abundance, recruitment, state surveys, age-length data, application of catchability and calibration studies, etc.) and provide a rationale for which data are used. Describe the spatial and temporal distribution of the data. Characterize the uncertainty in these sources of data.

4. Use appropriate assessment approach to estimate annual fishing mortality, recruitment and stock biomass (both total and spawning stock) for the time series, and estimate their uncertainty. Compare the time series of these estimates with those from the previously accepted assessment(s). Evaluate a suite of model fit diagnostics (e.g., residual patterns, sensitivity analyses, retrospective patterns), and (a) comment on likely causes of problematic issues, and (b), if possible and appropriate, account for those issues when providing scientific advice and evaluate the consequences of any correction(s) applied.

5. Update or redefine status determination criteria (SDC; point estimates or proxies for BMSY, BTHRESHOLD, FMSY and MSY reference points) and provide estimates of those criteria and their uncertainty, along with a description of the sources of uncertainty. If analytic model-based estimates are unavailable, consider recommending alternative measurable proxies for reference points. Compare estimates of current stock size and fishing mortality to existing, and any redefined, SDCs.

6. Define appropriate methods for producing projections; provide justification for assumptions of fishery selectivity, weights at age, maturity, and recruitment; and comment on the reliability of resulting projections considering the effects of uncertainty and sensitivity to projection assumptions.

7. Review, evaluate, and report on the status of research recommendations from the last assessment peer review, including recommendations provided by the prior assessment working group, peer review panel, and SSC. Identify new recommendations for future research, data collection, and assessment methodology. If any ecosystem influences from TOR 2 could not be considered quantitatively under that or other TORs, describe next steps for development, testing, and review of quantitative relationships and how they could best inform assessments. Prioritize research recommendations.

8. Develop a backup assessment approach to providing scientific advice to managers if the proposed assessment approach does not pass peer review or the approved approach is rejected in a future management track assessment.

Research Track TORs:

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

Guidance to Peer Review Panels 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 peer review 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.

PWS Appendix 2. Draft Review Meeting Agenda

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

American plaice Research Track Assessment Peer Review Meeting

July 18-22, 2022

WebEx link: TBD

DRAFT AGENDA* (v. 5/3/2022)

**All times are approximate, and may be changed at the discretion of the Peer Review Panel 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 Peer Review Panel.*

Monday, July 18, 2022

Time	Topic	Presenter(s)	Notes
9 a.m. - 9:30 a.m.	Welcome/Logistics Introductions/Agenda/ Conduct of Meeting	Michele Traver, Assessment Process Lead Russ Brown, PopDy Branch Chief Yong Chen, Panel Chair	
9:30 a.m. - 10:30 a.m.	TOR #1		
10:30 a.m. - 10:45 a.m.	Break		
10:45 a.m. - 11:45 a.m.	TOR #2		
11:45 a.m. - 12:15 p.m.	Discussion/Summary	Review Panel	
12:15 p.m. - 12:30 p.m.	Public Comment	Public	
12:30 p.m. - 1:30 p.m.	Lunch		
1:30 p.m. - 3 p.m.	TOR #3		
3 p.m. - 3:15 p.m.	Break		
3:15 p.m. - 4:15 p.m.	TOR #4		
4:15 p.m. - 4:45 p.m.	Discussion/Summary	Review Panel	
4:45 p.m. - 5 p.m.	Public Comment	Public	
5 p.m.	Adjourn		

Tuesday, July 19, 2022

Time	Topic	Presenter(s)	Notes
9 a.m. - 9:15 a.m.	Welcome/Logistics	Michele Traver, Assessment Process Lead Yong Chen, Panel Chair	
9:15 a.m. - 10:30 a.m.	TOR #5		
10:30 a.m. - 10:45 a.m.	Break		
10:45 a.m. - 11:45 a.m.	TOR #6		
11:45 a.m. - 12:15 p.m.	Discussion/Summary	Review Panel	
12:15 p.m. - 12:30 p.m.	Public Comment	Public	
12:30 p.m. - 1:30 p.m.	Lunch		
1:30 p.m. - 3 p.m.	TOR #7		
3 p.m. - 3:15 p.m.	Break		
3:15 p.m. - 4:15 p.m.	TOR #8		
4:15 p.m. - 4:45 p.m.	Discussion/Summary	Review Panel	
4:45 p.m. - 5 p.m.	Public Comment	Public	
5 p.m.	Adjourn		

Wednesday, July 20, 2022

Time	Topic	Presenter(s)	Notes
9 a.m. - 9:15 a.m.	Welcome/Logistics	Michele Traver, Assessment Process Lead Yong Chen, Panel Chair	
9:15 a.m. - 10:30 a.m.	TOR #5		
10:30 a.m. - 10:45 a.m.	Break		
10:45 a.m. - 11:45 a.m.	TOR #		
11:45 a.m. - 12:15 p.m.	Discussion/Summary	Review Panel	
12:15 p.m. - 12:30 p.m.	Public Comment	Public	
12:30 p.m. - 1:30 p.m.	Lunch		
1:30 p.m. - 3 p.m.	TOR #		

3 p.m. - 3:15 p.m.	Break		
3:15 p.m. - 4:15 p.m.	TOR #		BRPs, Projections and EGB Reference Points
4:15 p.m. - 4:45 p.m.	Discussion/Summary	Review Panel	
4:45 p.m. - 5 p.m.	Public Comment	Public	
5 p.m.	Adjourn		

Thursday July 21, 2022

Time	**Topic**	**Presenter(s)**	**Notes**
9 a.m. - 5 p.m.	Report Writing	Review Panel	

PWS 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.

   b. 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.

PWS Appendix 4. Peer Reviewer Summary Report Requirements

1. The main body of the report shall consist of an introduction prepared by the Research Track Peer Review Panel 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 peer review meeting. 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 peer review panel chair and reviewers should consider whether or not the work provides a scientifically credible basis for developing fishery management advice. If the reviewers and peer review panel 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 peer review meeting, 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 peer review meeting, including any changes to the Terms of Reference or specific topics/issues directly related to the assessments and requiring Panel advice.

Appendix 3. Agenda for the American plaice research track stock assessment peer review meeting.

American Plaice Research Track Assessment Peer Review Meeting July 18-21, 2022

WebEx link: https://noaanmfs-meets.webex.com/noaanmfs-meets/j.php?MTID=mf97d5121d96d26f36e88243f0dd9e013

Meeting number: 2763 669 5649 **Meeting password**: mP4vVXESd74 **Join by phone:** +1-415-527-5035 US Toll

AGENDA

**All times are approximate, and may be changed at the discretion of the Peer Review Panel 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 Peer Review Panel.*

Monday, July 18, 2022

Time	Topic	Presenter(s)	Notes
7 a.m. - 7:15 a.m.	Welcome/Logistics Introductions/Agenda/ Conduct of Meeting	Michele Traver Russ Brown, Yong Chen	
7:15 a.m. - 7:45 a.m.	Introduction and Overview	Steve Cadrin	
7:45 a.m. - 8:45 a.m.	TOR #1	Lisa Kerr and Jamie Behan	Environmental Effects
8:45 a.m. - 9 a.m.	Break		
9 a.m. - 11 a.m.	TOR #2	Steve Cadrin	Fishery Data
11 a.m. - 11:15 a.m.	Break		
11:15 a.m. - 12:15 p.m.	TOR #3	Paul Nitschke and Alex Hansell	Survey Data
12:15 p.m. - 12:30 p.m.	Summary/Discussion	Review Panel	
12:30 p.m. - 12:45 p.m.	Public Comment	Public	
12:45 p.m.	Adjourn		

Tuesday, July 19, 2022

Time	Topic	Presenter(s)	Notes
7 a.m. - 7:05 a.m.	Welcome/Logistics	Michele Traver, Yong Chen	
7:05 a.m. - 8 a.m.	TOR #3 cont.	Paul Nitschke and Alex Hansell	Survey Data

Time	Topic	Presenter(s)	Notes
8 a.m. - 9:30 a.m.	TOR #4	Amanda Hart, Tim Miller, Steve Cadrin, Dan Hennen, and Alex Hansell	Assessment Models
9:30 a.m. - 9:45 a.m.	Break		
9:45 a.m. - 11:45 a.m.	TOR #4 cont.	Amanda Hart, Tim Miller, Steve Cadrin, Dan Hennen and Alex Hansell	Assessment Models
11:45 a.m. - 12 p.m.	Break		
12 p.m. - 12:30 p.m.	TOR #4 cont.	Amanda Hart, Tim Miller, Steve Cadrin, Dan Hennen and Alex Hansell	Assessment Models
12:30 p.m. - 12:45 p.m.	Summary/Discussion	Review Panel	
12:45 p.m. - 1 p.m.	Public Comment	Public	
1 p.m.	Adjourn		

Wednesday, July 20, 2022

Time	Topic	Presenter(s)	Notes
7 a.m. - 7:05 a.m.	Welcome/Logistics	Michele Traver Yong Chen	
7:05 a.m. - 8 a.m.	TORs # 5 and #6	Steve Cadrin Paul Nitschke, Jamie Cournane	Reference Points Projections
8 a.m. - 9 a.m.	TOR # 7	Steve Cadrin	Research Recommendations
9 a.m. - 9:15 a.m.	Break		
9:15 a.m. - 10:45 a.m.	TOR #8 and Near Term Plans	Steve Cadrin	Alternative Assessment Approach
10:45 a.m. - 11 a.m.	Break		
11 a.m. - 11:15 a.m.	Summary/Discussion	Review Panel	
11:15 a.m. - 11:30 a.m.	Public Comment	Public	
11:30 a.m. - 12 p.m.	Key Points/Follow ups/Panel Wrap ups	Review Panel	
12 p.m.	Adjourn		

Thursday, July 21, 2022

Time	**Topic**	**Presenter(s)**	**Notes**
7 a.m. - 12 p.m.	Report Writing	Review Panel	

Appendix 4. Attendees of American plaice Research Track Peer Review Meeting. July 18-21, 2022

Yong Chen – Review Meeting Chair
Steven Holmes, Peter Stephenson, Massimiliano Cardinale - CIE Panel

Russ Brown - NEFSC, *Population Dynamics Branch Chief*
Michele Traver - NEFSC, *Assessment Process Lead*
Alicia Miller, Alex Dunn, Alex Hansell, Charles Adams, Charles Perretti, Jason Boucher, Kathy Sosebee, Dan Hennen, Mark Terceiro, Paul Nitschke, Tim Miller, Tony Wood - NEFSC
Steve Cadrin, Cole Carrano, Amanda Hart, Max Grezlik - SMAST
Libby Etrie - NEFMC
Jamie Cournane, Robin Frede, Angela Forristall, Chris Kellogg - NEFMC Staff
Jamie Behan, Lisa Kerr – GMRI
David McCarron - MADMF (retired)
Jackie ODell - Executive Director of Northeast Seafood Coalition
Mark Alexander - Asst. Director (retired), Connecticut Dept. of Energy & Environmental Protection

NEFSC - Northeast Fisheries Science Center
GARFO - Greater Atlantic Regional Fisheries Office
NEFMC - New England Fisheries Management Council
SMAST - University of Massachusetts School of Marine Science and Technology
GMRI - Gulf of Maine Research Institute
MADMF - Massachusetts Division of Marine Fisheries

Appendix 5. Model Selection Procedure for American Plaice Research Track 2022

The plaice WG considered WHAM model variants (alternative 'runs') based on structural aspects of greatest relevance to plaice, as identified by Terms of Reference or recommendations from previous assessments. In particular, environmental effects (ToR1), index selection (e.g., 2019 decision to exclude MADMF index and recommendation to consider separate Albatross and Bigelow indices; NEFMC 2020 recommendation to consider fishery CPUE), earlier start year (NEFSC 2002 recommendation), time varying selectivity, various random effect structures, and different age composition likelihoods.

1. Initial model acceptance initially focused on the requirement that the model converge on a solution. This convergence criterion eliminated some of the variants considered (e.g., notably run 37E with estimation of selectivity at age for multi-survey VAST stock indices did converge).
2. Model validation then focused on residual analyses. Non-random residual patterns for the inshore state surveys and calibrated Albatross-Bigelow series were used to justify excluding state surveys and splitting Albatross and Bigelow surveys as separate indices. In addition to conventional residuals, one-step ahead residuals were also used to judge model fit.
3. AIC was used to compare candidate models that were fit to the same data, fit the data well, and assumed the same statistical distributions and therefore had comparable likelihoods. AIC was similar among candidate runs but lowest the run used for status determination and projections (29F-4).
4. We examined retrospective patterns for all candidate model runs and measured retrospective inconsistency as Mohn's rho for spawning stock biomass and fully selected fishing mortality. All runs using the revised natural mortality assumption had similarly high retrospective consistency (rho<0.1).
5. We evaluated prediction skill of all candidate model runs using error of forecast values. Mean absolute scaled error (MASE) was similar among candidate runs but was lowest for the run used for status determination and projections (29F-4).
6. Self-tests were conducted on the three candidate runs (29F2, 29F4, 29F5). Candidate runs performed similarly in self-tests.

With criteria 1-6 generally being similar among the three candidate runs, run 29F4 was selected to present results, status determination and short-term projections, because it had the best retrospective consistency, AIC, prediction skill, and estimation performance for spawning stock biomass with 100% convergence in self-tests. In summary, the WG conducted model selection initially using traditional convergence and residual diagnostics for age-based assessments as well as some more recently developed diagnostics to determine the three candidate runs and the run selected for status determination and projections.