Independent Peer Reviewer Report of the American Plaice Research Track Stock Assessment

July 18-21, 2022

Report prepared by Steven Holmes, NIWA, New Zealand

Executive Summary

This report provides an independent peer review of the 2022 American Plaice Research Track Stock Assessment. Previous assessments of the Gulf of Maine and Georges Bank American Plaice (*Hippoglossoides platessoides*) were based on virtual population analysis (VPA) and suffered from strong retrospective bias. The primary goal of the research track work was to replace the VPA with a more modern stock assessment modeling framework, but the WG also attempted to identify environmental drivers that might influence American Plaice stock dynamics. Three integrated stock assessment modeling frameworks were developed: the Woods Hole Assessment Model (WHAM), Age Structured Assessment Program (ASAP) and Stock Synthesis (SS). One configuration of the WHAM model was recommended to be taken forward to the upcoming management track assessment.

The scientific and statistical analyses and presentations provided by the Working Group (WG) were of a high standard and the WG report very thorough. This reviewer accepts the WG's recommended assessment. The WG made recommendations for future research and the review panel made some additional recommendations all of which I endorse. The WG recommended that any backup assessment for providing scientific advice to managers remain an integrated assessment. I agree with this recommendation given the lack of any major concerns with the current recommended model, i.e., any rejection is expected to be related to concerns over specifics of the model configuration, not a complete rejection of the integrated model.

Introduction

The most recent (NEFSC 2019) assessment of the Gulf of Maine and Georges Bank American Plaice (*Hippoglossoides platessoides*; Fabricius 1780; commonly referred to as 'dab') stock was an operational assessment based on virtual population analysis (VPA). The 2019 assessment had the MADMF (Massachusetts Division of Marine Fisheries) inshore survey index excluded and retrospective pattern adjusted (NEFSC 2019). The 2019 assessment updated fishery catch data, research survey indices of abundance, and the VPA and reference points through 2018. Stock projections were updated through 2022. Based on this updated assessment, the stock status for Gulf of Maine and Georges Bank American Plaice was not overfished and overfishing was not occurring. Although accepted as the basis for management advice it was noted that the assessments for this stock consistently suffered a strong retrospective bias.

Subsequently, the stock was the subject of a research track effort to identify a state-of-art stock assessment model to replace VPA and improve the quality of the assessment with work beginning in June 2021. A Working Group (WG) was created to review the literature to identify environmental drivers that might influence American Plaice stock dynamics; compile fishery, survey, and life history data; configure one or more stock assessment modeling frameworks; evaluate biological reference points for stock status determination; and provide short-term projections. The WG also reviewed all the research recommendations identified in previous reviews, decided if the recommendations had been addressed and whether further

recommendations were needed. Lastly, the WG evaluated possible Plan B options for providing catch advice in the case the recommended stock assessment was rejected. Three integrated stock assessment modeling frameworks were developed: the Woods Hole Assessment Model (WHAM), Age Structured Assessment Program (ASAP) and Stock Synthesis (SS).

The WG attempted to convey all the necessary information for the peer review panel to draw conclusions (on whether to accept the work reviewed) in the WG report, but in addition provided 21 working papers that included additional details and background information. The list of materials provided for review is given in Appendix 1.

The Research Track Stock Assessment Peer Review Panel met via WebEx on July 18-21, 2022. The Panel was composed of three scientists selected by the Center for Independent Experts (CIE): Massimiliano Cardinale (Swedish University of Agricultural Sciences), Steven Holmes (NIWA, New Zealand), and Peter Stephenson (Department of Fisheries, West Australia). The Panel was chaired by Yong Chen (Stony Brook University), as a member of the New England Fishery Management Council's Scientific and Statistical Committee. The Performance Work Statement for CIE reviewers is provided in Appendix 2.

Overall Comments and Comments by TOR

The WG attempted to convey in the WG report all necessary information for the peer review panel to draw conclusions on whether to accept the work reviewed and to a large degree this was successful. The WG report was well organized which was helped by a logical sequence of TORs. This, in combination with the presentations on each TOR greatly facilitated the panel's work. All members of the WG were responsive to questions and requests for additional information. One such request was for a clearer statement of the criteria used to select between modelling frameworks or between configurations of a given model. Once this was provided it made it easier to see how choices had been made (especially between configurations of a given model). Future research track WGs should be encouraged to produce a similar document to help reviewers, but also to allow WG members to check, while a model or models are in development, that the criteria are appropriate and being applied consistently.

All further comments and recommendations are made against each TOR.

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.

I conclude that this TOR was fully addressed.

The literature review on this subject area appeared very comprehensive. Not surprisingly the literature indicated warming temperatures should be associated with accelerated growth rates, earlier ages at maturity and reductions in body size of post maturity. Warmer temperatures are also expected to lead to higher rates of natural mortality, particularly at younger ages.

Considering the data for the Gulf of Maine and Georges Bank American Plaice stock, highest values of recruitment per spawner (R/SSB) were found to coincide with years of extreme cold temperature. Equally useful to note were other variables (such as salinity) not found to be a significant driver of plaice recruitment. Analyses of the available survey series: Northeast Fisheries Science Center survey (NEFSC), Massachusetts Division of Marine Fisheries trawl survey (MADMF) and the Maine-New Hampshire trawl survey (ME-NH) suggested a general shift to deeper waters and a possible contraction of preferred habitat. Reduced availability to the state surveys was suggested. A Vector Autoregressive Spatio-Temporal (VAST) analysis was performed on the three surveys combined in part to test the idea of distributional shifts. The study showed a declining trend in 'effective area occupied' (estimated biomass divided by average estimated fish density) but this metric does not show where spatially densities are reducing. Color maps of estimated density through time seemed to suggest density reducing in deeper (not shallower) waters over time, but the reproduction of the figures was not clear.

The WG conducted regression analyses comparing key attributes important for stock assessment (recruitment, growth, and maturity) with potential environmental drivers. In this area more research is needed to better understand relationships between environmental changes (trend and variability) and both recruitment dynamics and fish growth/condition. In particular, the seemingly contradictory results for the influence of bottom temperature and the Atlantic Multidecadal Oscillation (AMO) need to be resolved. The regression analyses suggested a negative relationship between fish condition and Atlantic Multidecadal Oscillation (AMO) temperature anomaly, but a positive relationship with bottom temperature anomaly. Regression analysis of R/SSB against the temperature anomaly of the North Atlantic Oscillation (NAO) found a positive relationship between R/SSB and temperature which contradicts the findings from the literature of highest R/SSB from years of extreme cold. The AMO can be measured in terms of temperature anomaly but its influences on plaice may be more indirect, e.g., through altered prey availability, altered larval transport or some other mechanism. In the meantime, I agree with the panel recommendation to consider in future assessments drawing recruit per spawner for projections (and estimation of biological reference points) on the basis that recruits per spawner should implicitly account for environmental influences.

From the literature review the thermal limit for survival and incubation of American plaice eggs was stated as 14˚C and that recent Gulf of Maine sea surface temperatures in the late spawning season have exceeded that threshold in some areas. A potentially useful avenue for future research is detailed studies of sea surface temperature and surface current flows that can be compared to estimated recruitment at age 1 the following year. It is possible that the flow and temperature data could be used to warn of bad recruitment years.

The WG concluded that depth distribution of American plaice is influenced by season and temperature. The distribution of plaice shifts from deeper water in winter to more shallow water in spring/summer. It was also suggested that timing of the return to deeper water in Autumn could be affected by temperature. This might affect the availability to the NEFSC autumn survey. A

spatiotemporal analysis of commercial catches might throw light on the relationship between temperature and migration timing.

The ways environmental effects were considered in later TORs was clearly explained.

TOR 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.

I conclude that this TOR was fully addressed.

I agree the commercial catch (both landings and discards) appears to be well estimated. The assumed 100% discard mortality was well justified. The landing age composition data show good cohort-tracking, which was clearly demonstrated. Use of stock-wide length-weight, age composition and weight-at-age were well justified.

I agree with the panel that estimating discard age composition using the survey age-length key is applicable because the size range of fish discarded are well selected in the survey. If I understand correctly, the age composition of landings also uses the survey age-length key. The survey is considered to not catch larger fish as well as the commercial fleet, but it is hoped that larger fish are sampled sufficiently for the purposes of the key.

The WG considered both two generalized linear model (GLM) standardizations of landings per unit effort (LPUE) and a spatiotemporal model standardization of catch per unit effort (CPUE) as candidates for a fishery-dependent abundance index. In the section on fisheries data in the WG report it stated one GLM and the spatiotemporal model were taken forward for further consideration, but WHAM model runs only considered the LPUE series. It was not clear why. It would have been good if more time had been given in the review meeting to explain the choice of index to be taken forward for consideration in the WHAM modelling. The reasons for not including the LPUE series in the final proposed WHAM model were valid. During the review meeting a comparison of the standardized LPUE to the recommended model SSB was conducted to show whether there was evidence of hyperstability. It was appreciated that this work was conducted.

I agree with the panel recommendation to continue development of a fishery-dependent abundance index for possible use in the assessment process. The LPUE series developed appears to be tracking the population rather than fishing behavior (the test on the index developed in this research track assessment found the index to be only slightly hyper-stable). The review panel noted the absence of survey data in 2020. A fishery-dependent abundance index is an alternative source of information for this year and any future years where surveys are not possible.

I agree with the panel recommendations on continuing development of the Electronic Monitoring program and processing the otoliths collected by at-sea observers (to complement otoliths collected in the factory otolith sampling).

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.

I conclude that this TOR was fully addressed.

I accept the reasons for excluding the state survey indices as inputs to the proposed assessment. As mentioned under TOR 1 analyses suggesting a general shift of plaice to deeper waters implied reduced availability to both the Massachusetts Division of Marine Fisheries trawl survey (MADMF) and the Maine-New Hampshire trawl survey (ME-NH).

I agree that the NEFSC survey appears to capture the stock dynamics well, i.e., that age composition data show good cohort tracking, and the strong year classes identified in the NEFSC surveys are the same as those suggested in the fishery catch-at-age data (year classes 1987, 1993, 2004, and 2013). I also agree with the decision to treat survey data from the NEFSC fall and spring surveys when conducted using the survey vessel Albatross and survey vessel Bigelow as two separate time series.

A spatiotemporal model, Vector Autoregressive Spatio-Temporal (VAST), was used to develop model-based survey indices, which can integrate all three surveys (NEFSC, MADMF and ME-NH surveys). VAST has the potential to be a very useful tool but effort needs to be put into interpretation of results and their presentation. As mentioned under TOR 1 the study used the output metric of 'effective area occupied' which showed a generally declining trend. This seemed to be used as evidence to support the idea that plaice are moving out from state survey waters to deeper waters, but this metric does not show where spatially densities are reducing. Color maps of estimated density through time seemed to suggest density reducing in deeper (not shallower) waters over time, but the reproduction of the figures was not clear. Effective area occupied should be considered alongside spatial maps showing the modeled fish density by year and a map of bottom depth. It may be informative to run the VAST model simply on the two state surveys and consider the distribution of fish densities compared to histories of temperature within the area of these two surveys.

Using the VAST model, the WG conducted a counterfactual analysis to evaluate impacts of temperature and depth on the center of gravity of predicted spatially explicit abundance. It has taken this reviewer some time to understand what was intended by this analysis. If I understand the results, depth at which fish were caught was found significant in explaining distribution shifts in the density predictions, but in what way depths were changing did not come out of this presentation. Also, this analysis does not explain the underlying causes of the changes in depth distribution.

I do agree with the rest of the review panel that work using VAST should continue because of its ability to include state surveys for the development of integrated abundance indices that cover the entire stock area.

The review panel noted that the data available to the assessment suggest old and large plaice are not fully selected by the survey but tend to be fully selected in the fishery. A recommendation on future work is made under TOR 7.

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.

I conclude that this TOR was fully addressed.

It was impressive that the WG developed three different assessment models: the Woods Hole Assessment Model (WHAM), Age Structured Assessment Program (ASAP) and Stock Synthesis (SS). All three are appropriate packages for this type of assessment. Equally impressive was the large number of assumptions and configurations explored across models, but particularly in the WHAM model.

To an extent the WG created a rod for its own back. Many results, from different models, presented by different modelers with different presentation styles did make the logic behind model selection and, to a lesser extent, configuration of a given model hard to follow. A summary of model selection procedure (Appendix 4), produced at the request of the panel chair, did help explain the route to the proposed model considerably. I would recommend any future research track WG agree and record their selection criteria in a similar document and that presentations on model development explicitly reference the selection criterion(a) that led to a decision. As stated in the review panel summary report, the criteria for model selection should be clearly listed in the WG report, at the beginning of the model selection process (i.e., before any discussion of individual models).

In part, WHAM and ASAP were preferred over SS because expertise in model development and use is local (east coast USA). Given SS did not seem obviously superior to the other two packages I consider this a valid criterion.

What appeared to be the biggest problem of the previous VPA approach, strong retrospective bias, to a large extent was solved by a re-appraisal of natural mortality. The upward revision (from M=0.2) based on Gulf of Maine and Georges Bank specific growth patterns was well justified. I was concerned that the new value for M had been rounded from 0.27 to 0.3 but Steve Cadrin ran a sensitivity (of ASAP) to show there was almost no difference in model outcomes between these two values. The review panel recommended that future work consider developing an age- or size-dependent M. The current value used is a mean value over all ages/sizes, but it is highly likely that M is greatest on the youngest fish. Age specific M had originally been excluded on the grounds that values at young ages were unrealistic, but the panel suggested the rescaling approach of Lorenzen where $M_a = M*(L_{50\%mature}/L_a)$. An initial sensitivity showed further adjustments to

models would be necessary to accommodate such a change as simply changing the *M* values led to patterns in residuals (model over-estimating numbers of older fish in recent years).

The review panel recommended work to build an ensemble of different plausible configurations and model platforms. I suggest this is only necessary for research track WGs. Hopefully, work to achieve the integration of key environmental covariates into the preferred stock assessment model can be pursued between management track assessments. The review panel summary report states clearly the preferred criteria for model selection and pruning. I would add AIC is still valid if making a final choice between two similarly performing configurations of a model.

TOR 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.

I conclude that this TOR was fully addressed.

I agree with the WG conclusion that no stock-recruitment model can be found for the American plaice stock and that therefore $F_{SPR40\%}$ and $SSB_{FSPR40\%}$ are appropriate proxies for F_{MSY} and B_{MSY}.

The WG used the most recent 5-year estimates of observed weight at age to calculate $SSB_{F40\%}$. I agree with this decision because, as explained in the WG report, it recognized the changes in fish condition and weight-at-age for ages 7+ over time, the last five years of weight-at-age representing the relatively stable weights at ages 1-6 for the entire time series and the current period of relatively lighter weight at ages 7+. The decision to also use the most recent 5-year estimates of selectivity is also supported; estimates of fishery selectivity do change over the full history of the stock assessment but are stable over the most recent five years.

Maturity at age was shown to be relatively stable over time (Figure 4.1 of the WG report) justifying use of maturity at age from the full time series. The WG carefully considered the effect of temperature related environmental factors on recruitment to determine the appropriate timeframe from which to draw estimates. Evidence seemed inconclusive and given the uncertainty in recruitment estimates, the full time-series was chosen for reference point estimates, and I support this. Further work to understand the relationship between recruitment and ocean temperature is a recommendation from the WG. In doing this, careful distinction needs to be made between the effect of bottom temperatures on recruited fish and the effect of surface temperatures on eggs and pelagic larvae. I also recommend the assumption on maturity at age is reviewed regularly as Figure 4.1 of the WG report does show a noticeable reduction in age at 50% maturity after 2011, which was only reversed in the final data year, and it is possible this life history trait is another being affected by environmental drivers.

Based on outputs from the recommended WHAM model the WG estimated F40% to be 0.42 and $SSB_{F40\%}$ 18,000 mt and that there is very high probability the stock is not overfished, and

overfishing is not occurring. The Panel was unanimous in supporting the conclusion. We agreed with the choice of proposed candidate model (see TOR 4) and in addition, stock trends from other configurations of the WHAM model and from the alternative models were similar, suggesting that the stock status determination conclusion is robust.

I support the Panel recommendation to investigate the use of R/SSB, instead of recruitment, in the calculation of $SSB_{FSPR40\%}$, on the grounds this may remove some possible density-dependent influence on recruitment. I also support the recommendation that dynamic BRPs be explored in future research, although I recognize this may be a long-term objective as the relationships between temperature and the parameters for stock productivity are still to be resolved.

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.

I conclude that this TOR was fully addressed.

I accept the WHAM Run 29F4 model configuration as the proposed assessment configuration and therefore also accept its use to produce integrated projections. Assumptions on the time series of data to be used for recruitment and maturity at age (full series) and fishery selectivity and observed weight at age (recent 5 years) were all consistent with the assumptions used to calculate the biological reference point.

The scenarios chosen (F at F40% OFL, 75%F40%=ABC, F2019=status quo, and F=0) were as expected. I agree there is no need for retrospective adjustment (and indeed the lack of retrospective bias seemed to be a common feature across the new models considered).

The assumption to use the full series of recruitment estimates for projections should be re-visited, especially if the nature of the links between temperature (bottom and/or surface) and recruitment become better understood. As noted under TOR 5, maturity at age is possibly beginning to show a trend.

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.

I conclude that this TOR was fully addressed.

I agree with the rest of the panel that the WG produced an excellent summary of all previous research recommendations and whether they had been addressed.

As a review panel we agreed to the new research recommendations of the WG. The summary report of the review panel reproduces the recommendations in the order given by the WG; they are listed below according to my considered order of importance (with numbers in parentheses from the original ordering given in the review presentation)

1. (3) Investments are needed to streamline the estimation of commercial catch and promote reproducibility of estimates.
2. (4) Consider deriving discards from electronic monitoring when an integrated catch monitoring system is developed.
3. (5) As the Gulf of Maine scallop fishery expands, it should be included in discard estimation.
4. (1) Continue to monitor shifts in distributions of plaice, particularly depth and environmental covariates on catchability.
5. (7) The relationship between recruitment and ocean temperature should continue to be monitored.
6. (8) Methods should be developed to compare models with and without environmental covariates.
7. (2) Exploration of spatiotemporal integration of federal and state surveys should continue.
8. (6) Archived otolith samples should be processed (state surveys, at-sea observers, 1975-1979).
9. (9) 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.

This above prioritization reflects the opinion that a sound stock assessment method has been developed in WHAM so that emphasis should now be on ensuring consistent, high quality data inputs. The presentations on TOR 1 and 3 indicate American plaice will be affected by environmental change going forwards suggesting work to link environmental drivers to metrics important to stock status and its monitoring as next priority. Regarding item 7 (Exploration of spatiotemporal integration of federal and state surveys), emphasis needs to be given to interpretation and presentation of results. This was one area where the presentation of results was not as clear and easy to follow as for most of the review (see also my comments under TORs 1 and 3). Proposal 9 is discussed more under TOR 8.

Additional recommendations given in the review panel summary report are listed below. For the same reasons outlined above I suggest order of priorities would be (6), (1), (5), (7), (3), (2), (8). Item (4) is effectively covered by items 2 and 3 of the WG proposals (as listed above). Item (1) is linked to items 4 and 5 of the WG proposals. Item (8) could be considered encompassed by item 7 of the WG recommendations.

(1) Conduct further work to achieve the integration of key environmental covariates in the stock assessment models.

(2) Explore an ensemble modeling approach to incorporate different plausible configurations and model platforms selected and weighed by a comprehensive diagnostic against performance criteria agreed beforehand to provide stocks status and management advice for American plaice.

(3) Continue developing LPUE which can be used in the assessment process, possibly including bottom temperature in LPUE standardization.

(4) Continue developing Electronic Monitoring program and Observer monitoring to quantify the discards and biological information.

(5) Explore dynamic BRPs with consideration of environmental covariates given the large change of thermal habitat in the stock area and its potential impacts on American plaice life history processes.

(6) Continue developing built-in diagnostic tools for WHAM (e.g., plots of MASE).

(7) Consider developing an age- or size-dependent M. The current value used is a mean value over all ages/sizes, but it is highly likely that M is greatest on the youngest fish. Miss-specification of M by size may lead to biased estimates of selectivity and hence BRPs. One common approach is to scale the Lorenzen weight based Ms to the overall mean derived from meta-analyses.

(8) Further work on the development of VAST based survey indices in the stock assessment models.

A recommendation made by the review panel (under TOR 3) but not listed above is work to explain why old and large plaice are not fully selected by the survey, but tend to be fully selected in the fishery. This could possibly be achieved by analysis of the fleet and survey data at high spatial resolution to evaluate possible differences in stock availability between the survey and fishery.

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.

I conclude that this TOR was fully addressed.

The WG recommended that if the proposed assessment approach (WHAM run 29F4) was rejected in a future management track assessment that the backup be a re-configured WHAM model or possibly a version of the ASAP model.

The logic behind this recommendation is sound. Age compositions from the fishery and NEFSC surveys suggest older plaice are not fully selected in the surveys meaning fishery catch is not directly comparable to survey biomass estimates for a relative exploitation rate estimate. The survey not fully selecting older plaice also undermined their use for catch curve analysis (the assumption for constant mortality is violated) while catch curves from the fishery led to negative estimates of fishing mortality (estimated total mortality less than assumed value for natural mortality). Given the limitations in the empirical approach alternatives, I accept the argument that, in general, an integrated assessment should be the preferred approach. It was also put by the WG that, if the proposed assessment approach were replaced by an empirical backup, it could take a decade to reinstate an integrated assessment and I accept this strengthens the case not to put forward an empirical approach as the backup option.

There is a risk to this approach. The proposed assessment approach is obviously considered the optimal configuration of the preferred assessment model. For the upcoming management track assessment there is the possibility to revert to the most preferred ASAP run (Run 43), but in future years, if effort is concentrated on the one model and if that approach were rejected by a management track assessment it seems there would be an inevitable delay before modifications to the model (or an alternative integrated model) could be put forward for approval. I consider the risk of the approach being rejected as low, however, and so I accept the WG's arguments for only considering integrated assessments going forwards.

Appendix 1. Materials provided for review

American Plaice WG Report

Model Selection Procedure for American Plaice Research Track 2022

Ecological Influences (ToR1)

WP_14. Ecosystem and Climate Influences, by Jamie Behan, Lisa Kerr, Amanda Hart, Alex Hansell, Tyler Paklovitch and Steve Cadrin (November 16, 2021)

WP_16. Plaice Ecosystem Drivers by Jamie Behan and Lisa Kerr (June 21, 2022)

Fishery Data (ToR2)

WP_5. Fishing Industry Knowledge of American plaice, by Tyler Pavlowich, David Richardson, John Manderson and Greg DeCelles (November 9, 2021)

WP_6. Exploration of Fishery Data to Evaluate Catch Rates of American Plaice, by Max Grezlik, Lucy McGinnis, Keith Hankowsky, Gavin Fay, Steve Cadrin and Alex Hansell (November 10, 2021)

WP_7. Catch Rates of American Plaice Trawl Fishery, by Keith Hankowsky, Max Grezlik, Lucy McGinnis, Gavin Fay, Steve Cadrin and Alex Hansell (November 12, 2021)

WP_8. American plaice catch rate analysis using a spatial model, by Andy Jones, Tyler Pavlowich, David Richardson and Anna Mercer (November 13, 2021)

WP_9. Fishery Dependent Data Indices of Abundance (LPUE or CPUE) for American Plaice, by Mark Terceiro (November 16 2021)

WP_10. Electronic Monitoring Data: American Plaice, by Cate O'Keefe, Mel Sanderson and Liz Moore (December 4 2021)

WP_19. Fishery Data, by Larry Alade

Survey Data (ToR3)

WP_11. Seasonal Variation in Size-at-Age of American Plaice from Survey Data, by Steve Cadrin (November 22 2021)

WP_12. Spatio-temporal dynamics of American plaice (*Hippoglossoides platessoides*) in US waters of the northwest Atlantic, by Alexander Hansell, Larry Alade, Andrew Allyn, Lauran Brewster, Steve Cadrin and Lisa Kerr (December 1 2021; updated July 2022)

WP_13. Relative efficiency of a chain sweep and the rockhopper sweep used for the NEFSC bottom trawl survey and biomass estimates for American plaice, by Timothy J. Miller, David E. Richardson, Andrew Jones and Phil Politis (December 9 2021)

WP_20. Survey Data, by Larry Alade

Biology (ToR4)

WP_1. Size distribution analysis of American plaice, by Tyler Pavlowich (August 2021)

WP_2. Overview of American Plaice ageing in the Northwest Atlantic, by Josh Dayton and Eric Robillard (September 10 2021)

WP_3. Updating Parameters for Length and Weight Relationships and Length at Age of American Plaice, by Ashley Silver, Tyler Pavlowich and Larry Alade (September 10, 2021)

WP_4. Maturity Analyses of American Plaice in the Georges Bank and Gulf of Maine region, by Shakira Goffe, Daniel Hennen and Larry Alade (September 10, 2021)

WP_15. Approximation of Natural Mortality Rate for American Plaice in US Waters Based on Life History Traits, by Steve Cadrin (January 6, 2022)

Assessment Models (ToR4)

WP_17. American Plaice Assessment Model Developed in Stock Synthesis, by Dan Hennen and Alex Hansell (April 25 2022)

WP_18. A state-space assessment of American plaice using the Woods Hole Assessment Model (WHAM), by Amanda Hart, Lisa Kerr and Tim Miller (June 27 2022)

Projections (ToR4)

WP_21. Projections, by Larry Alade

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

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.

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

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 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.

   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.

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. Attendees for July 18-21 American plaice research track peer review meeting.

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

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

~~~~~~

Yong Chen - Chair
Steven Holmes - CIE Panel
Peter Stephenson - CIE Panel
Massimiliano Cardinale - CIE Panel

Russ Brown - NEFSC, Population Dynamics Branch Chief
Michele Traver - NEFSC, Assessment Process Lead

Alex Dunn - NEFSC
Alex Hansell - NEFSC
Alicia Miller - NEFSC
Amanda Hart - SMAST
Angela Forristall - NEFMC Staff
Charles Adams - NEFSC
Charles Perretti - NEFSC
Chris Kellogg - NEMFC Staff
Cole Carrano - SMAST
Dan Hennen - NEFSC
David McCarron - MADMF (retired)
Jackie ODell - Executive Director of Northeast Seafood Coalition
Jamie Behan - GMRI
Jamie Cournane - NEFMC Staff
Jason Boucher - NEFSC
Kathy Sosebee - NEFSC
Libby Etrie - NEFMC Member
Lisa Kerr - GMRI
Mark Alexander - Asst. Director (retired) of the Fisheries Division, Connecticut Dept. of Energy & Environmental Protection
Mark Terceiro - NEFSC
Max Grezlik - SMAST
Paul Nitschke - NEFSC
Robin Frede - NEFMC Staff
Steve Cadrin - SMAST
~~~~~~

Tim Miller - NEFSC
Tony Wood - NEFSC

Appendix 4. 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.