Summary Report of the Atlantic Cod Research Track Stock Assessment Peer Review

July 31 - August 3, 2023
Northeast Fisheries Science Center, Woods Hole, Massachusetts

Report prepared by Panel Members:
Jean-Jacques Maguire (Chair), NEFMC SSC
Steven Holmes, NIWA, New Zealand
Noel Cadigan, independent contractor with CIE
Coby Needle, independent contractor with CIE

Introduction

The Northeast Region Coordinating Council (NRCC)[1] has developed an enhanced stock assessment process to improve the quality of assessments. The process involves two tracks of assessment work: 1) a management track that includes routine updates of previously approved assessment methods to support regular management actions (e.g., annual catch limits), and 2) a research track that allows comprehensive research and development of improved assessments on a stock-by-stock or topical basis. The research track assessment process allows for a more thorough review of information available and for the evaluation of different assessment approaches than would be possible in a standard stock assessment process where the results are immediately used for management advice. This Panel reviewed the Research Track Assessment for Atlantic cod.

Previous stock assessments for cod in the Northeast USA were based on two stocks: the Gulf of Maine (an ASAP model most recently updated in 2021) and Georges Bank (age-based stock assessment rejected in 2015, a PlanBsmooth was used in 2017, 2019, and 2021). Based on an extensive review of the information available, the Atlantic Cod Stock Structure Working Group (McBride and Smedbol, eds. 2022) identified five biological units in the area, including two biological units in the Western Gulf of Maine, the winter and spring spawning components. The Atlantic Cod Research Track Working Group opted to combine the winter and spring spawning components into one assessment unit, proposing four assessment units for the area: i) Eastern Gulf of Maine (EGOM), ii) Western Gulf of Maine (WGOM), iii) Southern New England (SNE) and, iv) Georges Bank (GB). There is a separate assessment unit on Eastern Georges Bank managed jointly by the USA and Canada.

The work of the WG has been reviewed by the Atlantic Cod Research Track Peer Review Panel that met via WebeX from July 31, 2023 to August 3, 2023. The Panel comprised three independent scientists selected by the Center for Independent Experts (CIE): Noel Cadigan

[1] Atlantic States Marine Fisheries Commission (ASFMC), Greater Atlantic Regional Fisheries Office (GARFO), Mid-Atlantic Fishery Management Council (MAFMC), New England Fishery Management Council (NEFMC), and Northeast Fisheries Science Center (NEFSC).

(Independent contractor, Newfoundland and Labrador, Canada), Steven Holmes (National Institute of Water and Atmospheric Research, New Zealand) and Coby Needle (Independent contractor, Scotland). The Panel was chaired by Jean-Jacques Maguire (member of the Scientific and Statistical Committee of the New England Fisheries Management Council).

The Working Group report, 20 supporting Working Papers and meeting notes to be reviewed were made available on the data portal (https://apps-nefsc.fisheries.noaa.gov/saw/sasi.php) on July 17, 2023. The Working Group report (441 pages) was the main document to be reviewed with the other material providing additional information. The Panel was also given access to the GitHub repository used by the WG for cod modeling. The review meeting was held by Webex during July 31-August 3 from 12h00 to 17h00 Eastern Standard Time to meet during workable hours for Panel members from New Zealand and Scotland. Individual Panel Members and the Chair took the lead in providing first drafts of various sections of the report, but the entire Panel is responsible for the whole report. Prior to the meeting, members of the Panel met with Michele Traver and Russell Brown to review and discuss the meeting agenda, reporting requirements, meeting logistics and the overall process.

The Panel was assisted by Michele Traver (Chair, NEFSC's Stock Assessment Workshop Process Lead), Russell Brown (Chief, NEFSC Population Dynamics Branch) and Alexander Dunn (Communications Specialist, NEFSC Population Dynamics Branch). Documentation was prepared by the Atlantic Cod Research Track Working Group, and presentations were made by Lisa Kerr (University of Maine), GMRI staff (Amanda Hart and Jamie Behan), Massachusetts State Division of Marine Fisheries staff (Micah Dean), Steven Cadrin (University of Massachusetts Dartmouth), as well as NOAA Fisheries staff (Charles Perretti, Alex Hansell, Richard McBride, Scott Large, Kathy Sossebee, and Tim Miller) (all NEFSC)). Other WG members contributed to the discussions on various topics. Toni Chute, Lisa Hendrickson, Chris Legault, Liz Brooks, Brian Linton, Burton Shank, and Chuck Adams (all from the NEFSC) acted as rapporteurs throughout the meeting (see Appendix 4 for materials provided and Appendix 5 for meeting attendees).

The Atlantic Cod Research Track Working Group (WG) was created with staff from NOAA Fisheries, academia, and Fisheries and Oceans Canada (DFO). This eleven-person WG (chaired by Lisa Kerr, University of Maine) met 22 times from November 2021 through July 2023 to prepare updated assessments for four stocks of cod – Western Gulf of Maine, Georges Bank, Southern New England and Eastern Gulf of Maine. Additional topical meetings (i.e. climate and ecosystem impacts on cod stock dynamics and review of survey indices) were held by subgroups of the WG and additional experts to make progress on technical work. Terms of Reference for the WG are provided in Appendix 1.

The meeting opened at noon on Monday July 31, with welcoming remarks by Russ Brown, Michele Traver, and the Panel Chair. The meeting agenda is provided as Appendix 2. Presentations on the nine Terms of Reference were made on each day. Additional requests by the Panel were reviewed at the beginning of the meeting the following day. Panel members and the Chair drafted material for the Panel's Summary Report in a Google Doc. The Panel Chair compiled and edited this Panel Summary Report with assistance (by correspondence) from the

CIE Panelists, before submission of the report to the NEFSC. Additionally, each of the CIE Panelists will submit their separate reviewer's reports to the Center for Independent Experts.

The analyses conducted by the WG were thorough and of high quality. Their report and presentations were clear and made the Panel's job much easier.

The Panel agreed that all nine TORs had been met, most entirely, but some only partially: for example, there is evidence of lack of fit in EGOM and GB cod assessments, which have some large survey index residual trends. The WG's approaches to estimating biological reference points (BRPs) and making projections for all four stocks were well reasoned; they should form a reasonable basis for providing management advice for the four assessment units when data are updated in the management track assessments.

The Panel's evaluation of the WG's response to the nine TORs is provided below and concludes with a summary of key recommendations.

Evaluation of the Terms of Reference for Atlantic Cod

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 Panel agreed that this TOR has been met for all of WGOM, GB, SNE and EGOM cod stocks.

The panel appreciated the considerable effort that had been put into this ToR. The approach to arrive at the inclusion of ecosystem considerations in assessment model assumptions and parameterization was logical: science reviews to establish indicators, consideration of which indicators affect which life history traits, rationalization of indicators via consistent criteria (theoretical merit, i.e., the ability of the indicator to inform knowledge of a key process and operational merit, i.e., the ability of the indicator to be created and analyzed in a timely manner), and exploratory modeling to test the strength of linkages. Given the strong evidence of recent environmental changes, especially in the Gulf of Maine region, the panel considered that restricting the literature review to papers published after the year 2000 to be reasonable.

The 'ecosystem profile' stage of the work included eliciting fishers' expert knowledge and perspectives at workshops held in February and March 2022. Summaries of the possible influences on cod that were raised during the meeting were made in the main report, WP1 and WP4 and the presentation given to the panel, but the panel noted that these summaries lacked synthesis. It wasn't clear to what extent there had been follow-up to some of the points raised although the

panel also noted that utilizing stakeholder views can be difficult as they are often too specific in time or space. Seal predation was identified both by fishers and from the literature review. The panel noted some papers on seal predation and its specific effect on Northwest Atlantic cod did not appear to have been reviewed. The Panel recommends targeted sampling in time and areas of cod and seal overlap to better inform seal predation effects in any future work on indicators (e.g., Hammill et al., 2014).

Exploratory modeling was able to establish that increasing temperatures were associated with decreases in cod weight at age in WGOM and GB stocks. For the GB region it also identified a distributional shift in cod on GB in spring. The WG recommended "further exploration of density dependent effects and evaluation of the likelihood of spurious relationships identified between ecosystem drivers and aspects of stock dynamics," and the panel supports this. Techniques to test for causal relationships were evaluated briefly but only on WGOM data.

The development of stock size indices by integrating survey data using the Vector Autoregressive Spatio-Temporal (VAST) package was explored but not considered ready for inclusion into the stock assessments. Integrated modeling of survey data can generate a single spatio-temporal survey index based on the contributing surveys. Models can also be used to incorporate environmental data to inform on spatial patterns of stock density. The Panel considers this is important work that should be continued, especially considering the postulated distributional shift in cod on GB in spring and concerns about conflicting signals from the NEFSC and DFO spring surveys for this stock.

Under ToR 1, the WG tested for environmental influences on recruitment, growth (condition and weights-at-age) and distribution, but only models for recruitment were explored further in the WHAM assessments as condition and distribution cannot be implemented directly in the current WHAM models. As the research track progressed WHAM became the preferred model platform across all four stocks (see ToR 4). In only one WHAM model, that for WGOM, was it attempted to use environmental data explicitly, as explanatory covariates (three temperature related metrics). For the most part the WG attempted to accommodate ecosystem/climate influences through inclusion of random effects in the models, although these couldn't always be adopted because of model non-convergence or poor diagnostics. An example was time-varying natural mortality, examined for WGOM, GB, and SNE stocks as annual process errors in M (with and without environmental covariates) but not adopted in final models due to issues with model convergence and based on model diagnostics. The Panel notes that the inclusion of random effects in recruitment and/or M and/or numbers at age (ages 2+) may negate the need to identify specific causal links between environmental drivers and a species' dynamics. However, a reliable causal link, if identified, might increase confidence in projections if the covariate can be predicted. The Panel notes that the WG considered that sensitivity runs for the SNE assessment when using Stock Synthesis showed promise for the inclusion of environmental covariates, but time constraints prevented the same explorations using WHAM.

The Panel observes that climate change would suggest declines for gadoid stocks in the North Sea, but some (haddock, whiting) are currently increasing rapidly towards historical maxima, and others (cod, saithe) are improving. The influence of climate change (and other environmental drivers) can therefore be difficult to predict for specific fish stocks.

Ecosystem/climate changes were accommodated in reference point calculations and projections by recognizing changes in weight-at-age and maturity-at-age (although not for SNE and EGOM where the data are too few) and using the last five years of weight-at-age and maturity-at-age data for the calculations. Conflicting indicators led to some debate within the review about the time period that best represented possible recruitment for reference point and projection calculations. The WG recommended a "broader discussion (i.e., across species) of appropriate methods for defining time windows to characterize prevailing conditions for stocks" and the Panel supports that recommendation.

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.

The Panel agreed that this ToR has been met for all four cod stocks reviewed (WGOM, GB, SNE and EGOM). Uncertainties in landings and discards were discussed during the review but there is a need to quantify uncertainty in a way that can be included in age/length based stock assessment models.

The Panel notes that this ToR did not ask about the reliability/uncertainty in the size/age composition of the catch, which is an important contributor to uncertainty in assessment models. This is particularly important for state-space stock assessment models where the choice of an observation model for age composition information could be critical. Choices by the assessment analysts included the Multinomial, Dirichlet-Multinomial, Dirichlet, and logistic normal distributions. The Atlantic Cod Research Track Working Group chose an observation model based on model selection diagnostics: convergence, residuals, Akaike's Information Criteria (AIC), retrospective consistency, and prediction skill. This choice should also be informed by the sampling programs for age and length compositions, and additional uncertainty related to "data borrowing" when length and age samples are scarce or missing. **The Panel recommends research continue to standardize and quantify the uncertainty and sampling distribution of age compositions (**e.g., Thorson, 2014; Thorson and Haltuch, 2019; Thorson et al., 2023).

Sampling was not extensive for any of the four stocks, but it was particularly scarce for Eastern Gulf of Maine and Southern New England cod. The WG investigated the effect of missing or poorly sampled length data by market categories and found the effect to be significant. When samples were considered insufficient, length/age data were either binned at the higher time stratum or not estimated for the corresponding time and area stratum.

The Panel enquired as to what criteria would lead to moving from an age-based assessment to a data-limited one. The experience in the Region is that age-based assessments, possibly with Mohn's Rho adjustment for retrospective pattern, are preferable to the data-limited approaches used (PlanBsmooth, Ismooth). Also, published work based on simulations using Stock Synthesis (Wetzel & Punt 2011, Cope 2013, Rudd et al. 2022) and WHAM (Legault et al 2023) suggested that age-based assessments performed well compared with data-limited methods.

Age samples are not collected for the recreational fisheries: therefore, lengths are converted to ages using age samples from the commercial fisheries and surveys. Sampling in the commercial fisheries is stratified by market categories, a proxy for stratifying by length. In such cases, **the Panel suggested that age-length-based models that can use the length data directly as sampled should be investigated using the WHAM platform**.

Discard mortality estimates are low compared to studies conducted elsewhere. However, the Panel was satisfied that the field studies were scientifically valid and the discard mortality used were accepted.

The figure comparing the weights at age 1 for SNE cod based on age-length keys and those from a state space growth model showed fewer yearly values for the growth model, likely due to differences in the methods used for filling age-length key holes.

All cod stocks where there were sufficient data to estimate weights-at-age by year showed a decreasing trend in weights-at-age for older ages. This observation has been made for many other stocks of several species including cod in other areas. There is no convincing explanation for this yet.

The Panel agreed with a comment from the public that, when including random effects in modeling, care should be taken that the assumptions are met and that the packages used to calculate them are numerically stable. The TMB software used for this purpose by the WHAM package has been extensively tested, and the WHAM package is also currently undergoing extensive simulation testing in a separate research track, which will conclude soon.

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.

The Panel agreed that this ToR had been met in full for Western Gulf of Maine (WGOM), Eastern Gulf of Maine (EGOM), Georges Bank (GBK) and Southern New England (SNE) cod stocks.

The material submitted prior to the Panel meeting regarding available survey indices was extensive, and was provided via several working papers (WPs 8-11) as well as the main WG report. The WG considered the utility and appropriateness of eleven survey indices, ranging from age-structured indices covering all four stock areas, to juvenile inshore surveys from specific locales. This is many more than were used in the previous GOM and GB stock assessments, and it was noted by the WG that this holistic approach arose from an effort to collate as much relevant data as possible in a situation where stock-specific fishery catch data may be sparse. The extension to four cod stocks from the previous two also made it more practical to include additional survey indices that are more geographically specific to sub-stocks.

In the main report and presentations, the WG was consistent regarding the exploratory and diagnostic summaries presented. These included maps of survey locations, survey strata assignments to stock areas, abundance and biomass time series, age composition summaries, and length distributions. Summary notes were also provided for each survey in the WG report, covering survey timeframe, spatial extent, indices provided, biological sampling, and any calibration experiments. Further information was provided in WPs, and included time series of mean start date, depth and temperature; catch distributions and annual stock centroids; Gini indices; fraction of positive tows; bivariate scatterplots of cohort-based abundance; and weights-at-age, lengths, and relative condition (although these additional analyses were mostly limited to the NEFSC surveys). For each survey and stock, the WG offered a conclusion on the utility and applicability of the resulting indices for stock assessment and advice, along with any caveats or concerns.

The Panel suggested that the material available at the start of the Panel meeting was insufficient to enable a conclusion on whether the available surveys were able to track cohort strength for the four stocks. Suitable diagnostics for this would have included bivariate scatterplots with fitted linear regressions and associated correlation coefficients, survey catch curves, and time-series of log survey indices by age (with cohort as the *x*-axis). Appropriate figures were provided by the WG during the Panel meeting (with comparative plots also produced by Panel members), from which it can be seen that cohort-tracking ability varied between surveys and regions. The specific Panel conclusions on this aspect are covered for each stock assessment in comments for ToR 4 below.

The Panel noted that many ICES assessments in Europe now make use of spatio-temporal GAM (generalized additive model) approaches to develop combined survey indices for given seasons that can incorporate survey data from different vessels, gears and areas, and covering different age ranges, thus facilitating the inclusion of surveys that might be limited in the areas or years covered. The ICES approach also involves a spatiotemporal ALK model to estimate haul-specific survey catch-at-age from catch-at-length, but this may have its own problems (Babyn et al., 2021) which is a complication. The WG has begun an exploration of a similar approach for very coarse length bins using Vector Autoregressive Spatio-Temporal (VAST) models (WP 10), which show promise, **and the Panel recommends that this work be continued for the next Research Track process**.

The Panel discussed the results from the Groundfish LPUE project (WP 6, and summarized in the main WG report). The indices arising from this work led to poor diagnostics in the

assessment models for all but one of the stocks, and the WG proposed retaining the recreational LPUE index for the SNE stock. The driver for this inclusion was the lack of other data for this stock, meaning that assessment model stability was reliant on this LPUE index. However, the Panel noted that the index was based only on positive data (i.e., those recreational fishing events that had led to cod catch), and was therefore likely to be an overestimate of the true recreational cod LPUE and possibly produce biased estimates if there is a trend in the proportion of zero catches over time. Following further discussion during the meeting, the Panel concluded that it was reasonable to persist with the use of the recreational LPUE index for the SNE stock for the time being: there seems to be less potential for hyper-stability of recreational LPUEs; it is one of the few apparently reliable sources of data for the SNE stock; it does not lead to recreational catch data "being used twice" (as the age compositions are only fitted once); and the assessment still suggests significant and severe depletion even if this index may be overestimating recent abundance or suffering from hyperstability. The Panel considered that retaining the index would generate two key benefits: it would encourage both management action and the further development of the index and associated assessment.

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 Panel concluded that ToR 4 had been mostly met for each of the cod stocks. The resulting assessments are accepted for use in subsequent management track processes, although there are some research recommendations that must be addressed before using the models in management track processes (see ToR 7).

The WG had established a clear set of criteria for model selection that was consistent across stocks and this was communicated clearly to the review panel.

Implementing models with time-varying M was not successful for the WGOM and GB stocks. This was also unsuccessful for the EGOM and SNE stocks, but this is less surprising given the data limitations for these stocks. Natural mortality rates are often considered to be among the most important parameters in a fish stock assessment, but they are also among the most difficult parameters to estimate using commonly available data (e.g., Punt at al., 2021). The WGOM and GB stocks included cohort process errors which is an attempt to partially account for time-

varying M. However, these cohort errors can also be due to other process errors such as movement between stock areas. There is ambiguity about the efficacy of using cohort process errors to account for time-varying M (Aldrin et al., 2021).

Many of the exploratory runs seem to have been discarded due to a lack of convergence in the WHAM model run (for example, all the time-varying M runs). The Panel questioned whether this might perhaps be a problem with WHAM itself and similar approaches, rather than the settings attempted, and **recommended** that this issue be explored more fully in future research track processes (for example, by resetting parameter bounds if lack of convergence is due to bounds being reached). The Panel suggests that some reasonable representations of reality may have been missed because the relevant WHAM run wouldn't converge. In addition, the **Panel recommends** that it would be useful if the WHAM model could include leave-one-survey-out diagnostics.

The survey indices and catch information at older ages did not track cohorts well for most stocks. Additional evaluation of the age for a plus group should be conducted.

Western Gulf of Maine (WGOM)

The Panel agreed the ToR 4 had been met in full for the WGOM stock.

The assessment of the WGOM stock was presented through the relevant section in the main report under Tor 4, as well as via WP 17 and with a helpful and comprehensive presentation to the Panel meeting. The Panel appreciated the extensive description of research that had gone into the assessment formulation and testing. The new WGOM stock boundaries differ from the former stock area by excluding statistical areas 511 and 512 (which now form part of the EGOM assessment area) but adding statistical areas 521 and 526 (formally part of the GB assessment area) and fishery data from 541, which was described in more detail under Tor 2.

Key differences in the former assessment model and the final model presented to the Panel were summarized in the main document:

	GOM model (NEFSC 2013)	Proposed final WGOM model
Model platform	ASAP	WHAM
Start year	1982	1981
Fleet structure	Single fleet	Two fleet (recreational and commercial)
Fleet selectivity blocks	1982 - 1989, 1990 - 2004, 2005+	1981 - 1989, 1990 - 2011, 2012+
Fleet selectivity function	logistic	Logistic, logistic, age-specific
Age-composition distribution	Multinomial	Dirichlet

	GOM model (NEFSC 2013)	Proposed final WGOM model
Indices of abundance	NEFSC BTS spring, NEFSC BTS fall, MADMF spring	NEFSC BTS spring, NEFSC BTS fall, MADMF spring, MENH spring, MENH fall, IBS spring, BLLS spring, BLLS fall
NEFSC BTS vessel calibration	Calibrated to FSV Albatross units	Uncalibrated (split)
Recruitment	Deviations from mean as fixed effects	Deviations from the mean as random effects
M	M=0.2, M-ramp	M=0.2
NAA random effects	None	All ages

The Panel concluded that the final WGOM WHAM model was appropriate.

The Panel had no issue with the start year being moved to the first year in which length data was recorded from recreational catch. Equally, even if the selection criteria had been equivocal, it is logical to split the commercial and fleet data so that possible differences in selectivity can be accommodated. In the event, adoption of separated fleets led to improvement in fleet age-composition residuals and prediction accuracy and no degradation in other criteria.

The choice of first split in fleet selectivity blocks (1989-1990) in the previous assessment was supported by the fact that legislation changed in 1990 (increases in minimum legal size for both commercial and recreational fleets) but the second split (2004-2005) had been chosen primarily on the basis of model diagnostics. A second split at 2010-2011 is more logical because of elimination of trip limits and introduction of annual catch limits in 2010, as well as spawning closures. Additional catch restrictions were implemented in 2014/2015 which may also have impacted selectivity, possibly resulting in avoidance of older ages (i.e., domed selectivity) for both fleets and four block selectivity was investigated through age-specific selectivity in all four blocks. Ultimately a three-block model was chosen with logistic selectivity in blocks 1 & 2 and age-based selectivity in block 3 (starting in 2012), but a four-block model with logistic selectivity in blocks 1 to 3 and age based in block 4 does not seem to have been considered. The Panel **recommends** a four-block logistic-age based version of the model be tested in future.

Dirichlet age-composition likelihoods produced a clear improvement according to the criteria for model selection with no apparent downsides. However, the Dirichlet does not use any information on age sample sizes, and it will not account for changes in sampling over time. The panel noted that when reliable information on age sampling exists then using the Dirichlet distribution will often not be best practice. There is evidence of mis-specification in terms of residual variability vs age.

The WG had identified nine additional survey indices for possible inclusion to the WGOM assessment in addition to the three used previously. The panel agrees it was justified to not include

the age 0 indices (or MADMF fall survey) as diagnostics suggest these indices would effectively be adding noise to the assessment. This still allowed five additional surveys to be included. The NEFSC BTS spring, NEFSC BTS fall and MADMF BTS spring surveys were used to test for prediction error because they were included in all model variants considered. Prediction error for these surveys improved when the additional surveys were included and the fits to the aggregate indices of the eight surveys was generally good, with age-composition residuals without problematic patterning.

The Panel agreed with the decision to split the NEFSC survey indices at the change point in survey vessels (and gear) from the Albatross to the Bigelow in 2009. The Panel noted the Bigelow to Albatross calibration estimates used previously to convert Bigelow survey stratified mean numbers-at-length to Albatross units were highly uncertain at young and older ages, with the estimates based on only 130 tows.

The WG attempted to estimate a Beverton-Holt stock-recruitment relationship in the model: however, the Beverton-Holt function was estimated to be linear in SSB, which implies an unbounded upper limit on recruitment. On this basis it seems reasonable to model recruitment as deviations from the mean for the time being. The linear relationship between SSB and recruitment implies that using average recruitment in projections may overestimate future recruitment. The **Panel recommends** further investigations of robust approaches to modeling recruitment in the assessment model.

The WG attempted to estimate a constant natural mortality *M* across all ages and a time varying *M* via random effects. Estimation of a constant lifetime *M* was possible but all (seven) variants of the model including a time varying *M* failed to converge. The estimated age- and time-invariant *M* of 0.21 was very close to the value of 0.2 calculated outside of the model and based on life history traits. Given similar model diagnostics, the Panel agrees with the WG decision to retain the fixed *M* model specification on the grounds of parsimony.

The final model applied random effects for numbers at age (NAA). The type of error assumed was 2Dar1 (autocorrelation across ages and between years) and the Panel noted that applying the 2Dar1 process to all ages is a possible model mis-specification. The recruitment process errors would be expected to be considerably different to survival process errors at ages 2-9+. However, for WGOM cod the age correlation in the NAA random effects was low so removing the correlation between age 1 process errors and those for ages 2-9+ may not make much difference to results. It is possible that the option to remove the correlation between age 1 and 2+ will be included in the next update of the WHAM package. If this is the case the **Panel recommends** the effect of using this option be considered in a future assessment.

Environmental effects on recruitment can be directly incorporated into WHAM and the WG tested whether including the ecosystem variables of sea surface temperature (SST), bottom temperature or heatwave index as covariates would improve the assessment model. The covariates tested explained little of the variability in recruitment, in contrast to the run where a Beverton-Holt stock-recruitment relationship had been estimated, and similar diagnostics apart from lower self-test convergence rates led to the WG choosing the proposed final model configuration without environmental covariates. The Panel considers it might be useful to revisit these tests in future,

especially as it was unclear from the presentation and report whether the covariates were only tested individually or also in combination.

NAA random effects seem to be an effective way to improve model diagnostics, but may also act as a convenient way to mask model misspecification and/or biased data. The WG conducted missing catch experiments such that catch data provided to the model were the result of reported catch linearly reduced to 75% or 25% of the original value over the final 20 years. The WG concluded the missing catch led to "generally unaffected" model diagnostics but "increasingly negative process errors with increasing missing catch". The Panel considers the WG correct to caution against assuming that NAA process errors indicate changes in natural mortality, and notes the importance of considering any trends in process errors if random effects are incorporated.

The Panel noted that, in the model development process, if a new or changed feature of the assessment model had been accepted it was retained in all subsequent investigations. It also noted a seemingly high proportion of candidate runs that failed to converge. Given the large number of model aspects that were re-visited and the large number of model configurations that were run, adopting this approach is reasonable. However, the Panel noted it is possible a model change, e.g., age dependent *M*, that failed to converge or provide good diagnostics given the overall model set up in place, may have performed better given differences in model set up elsewhere. As such, **features that led to non-convergence or poor performance during the current research track could still be considered in any future research track, potentially with different permutations of other model features**.

Georges Bank (GB)

The Panel agreed the ToR 4 had been met for the GB stock. If possible, for the subsequent management track assessment, it would be useful to identify the reasons for differences between NEFSC indices and Canadian indices.

The assessment of the GB stock was presented through the relevant section in the main report under Tor 4, as well as via WP 19 and with a helpful and comprehensive presentation to the Panel meeting. The Panel appreciated the extensive description of research that had gone into the assessment formulation and testing. The new GB stock boundaries differ from the former stock area by excluding statistical areas 521 and 526, which was described in more detail under Tor 9.

The GB stock differed from the other three because it has primarily been fished by commercial fleets from the US and Canada. Key differences in the former assessment model and the final model presented to the Panel were summarized in the main document:

	GB model (NEFSC 2015)	Final GB model
Model platform	ASAP	WHAM
Start year	1978	1978
Fleet structure	Single fleet	Single fleet

	GB model (NEFSC 2015)	Final GB model
Fleet selectivity blocks	1978 - 1993, 1994+	1978 - 1993, 1994+
Fleet selectivity function	logistic	logistic
Age-composition distribution	Multinomial	Dirichlet-Multinomial
Indices of abundance	NEFSC BTS spring, NEFSC BTS fall, DFO spring	NEFSC BTS spring, NEFSC BTS fall, DFO spring
NEFSC BTS vessel calibration	Calibrated to FSV Albatross units	Calibrated to FSV Albatross units
Recruitment	Deviations from mean as fixed effects	Deviations from mean as random effects
M	M=0.2	M=0.29
NAA random effects	None	All ages

The Panel concluded that the final GB WHAM model was appropriate. The Panel appreciated the thorough description of model development runs provided in the main report (summarized in Table 4.GB2 of the report) and Table 1 of WP 19.

During the review meeting the Panel requested a sensitivity run to investigate if the assessment output was robust to the somewhat subjective choice of the WHAM likelihood for the age compositions. The assessment team thoroughly responded to this request and demonstrated that the final model was robust to this choice of likelihood, which is illustrated below for SSB, although this was also true for recruitment and Fbar.

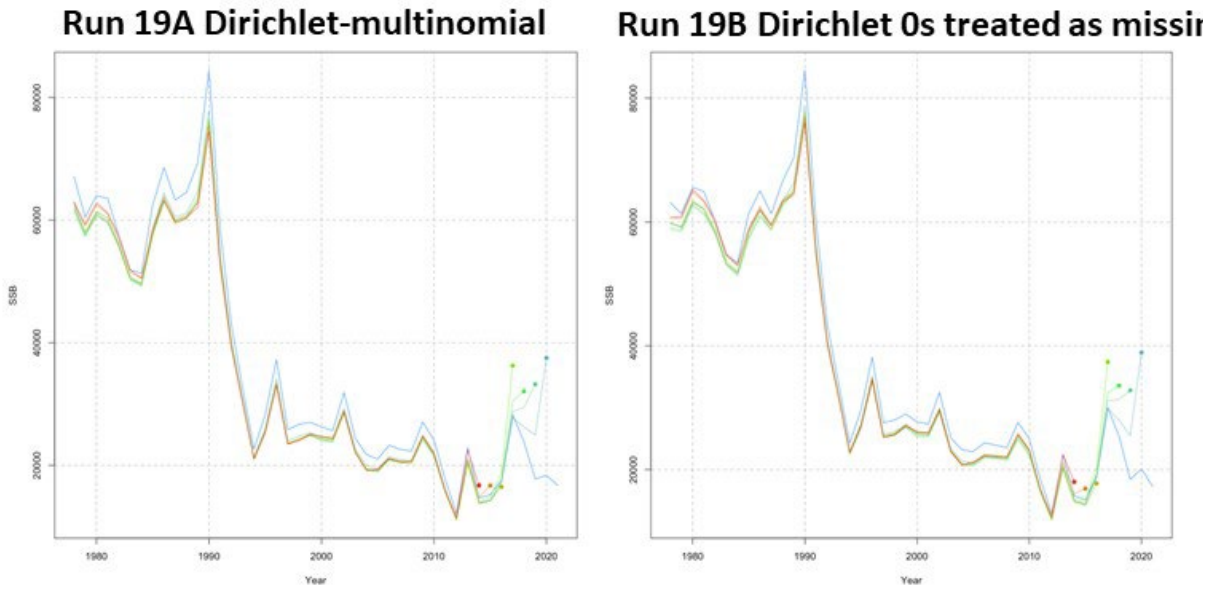

Figure 1. A comparison of GB SSB retrospective patterns for the final model (left-hand panel) and the sensitivity model (right-hand panel) that used the Dirichlet distribution with zeros omitted instead of the Dirichlet-Multinomial distribution.

However, the Panel recommends that further investigations would be useful:

- Further explorations of differences in NEFSC and DFO survey catch rates during 2000-present are required. This should include whether the differences are the same for all sizes. The final model has a residual pattern, with mostly negative residuals for the DFO spring survey indices since 2010, and mostly positive residuals for the NEFSC spring and fall survey indices since 2010. Time-patterns in index residuals will often result in retrospective patterns. Hence, as the number of survey index-years increases, changes to the influence these indices have on WHAM model estimates may be contributing to the retrospective patterns. Resolving the differences in the indices may contribute to reducing the retrospective patterns for this stock.
- Investigate a model using unconverted Bigelow indices, similar to WGOM cod. The uncertainty in the cod conversion factors were large (see Miller, 2013). During those comparative fishing experiments, cod were caught in only 130 tows. Hence, the converted Bigelow indices since 2009 may be providing inaccurate stock trend information. Ideally, if an age and length structured assessment model was used (e.g., the growth branch of WHAM), the catchability of indices would be modeled as length-based (instead of age-based like in WHAM) and the comparative fishing data could be simply included as a likelihood component for the ratio of Albatross-Bigelow catchabilities. This seems like a long-term goal for GB cod, but this approach has been implemented for some other northwest Atlantic stocks to address changes in survey protocols.
- Investigate combining the NEFSC and Canadian spring surveys to provide a single spring index for GB cod.
- If the differences in NEFSC spring and fall survey indices and the Canadian indices cannot be resolved, then two WHAM models should be formulated (one for US indices, one for Canadian indices) to investigate differences in assessment results and management advice.
- The NAA random effects (i.e., cohort process errors) were assumed to be independent across all ages and years. The variance of the age 1 process errors was estimated separately from the variance of process errors at ages 2-9+. This was appropriate. However, the Panel noted that predicted process errors shown in Fig. 22 of WP 19 seemed to have substantial autocorrelation, and also substantial between-age correlation at ages 4-9+. Hence, the panel recommends that the 2DAR1 correlation structure for these random effects be explored in the future, at least for ages 4-9+. However, random effects at age 1 should be independent of the effects at ages 2-9+ because the recruitment random effects (i.e., deviations from the mean) are really different from the survival random effects used for ages 2-9+. Random effects in the first year may also need to be independent of random effects in other years because in the first year these effects may account for survival deviations from the initial equilibrium age-distribution approximation for stock numbers used in the GB WHAM model, and these survival deviations may be substantially different from those in other years. The advantage of

using correlated NAA random effects is that trends in the process errors can be propagated into short-term projections, depending on the strength of the correlations.
- Jan1 and SSB weights-at-age were calculated using the Rivard approach applied to commercial weights. However, these weights are known to be over-estimates of the stock weights at young ages because of the size-selectivity of commercial fisheries. Survey weights-at-age may be better for stock weights because survey gears may be less size-selective. This should be investigated.

The Panel agreed with a comment from the public that it would be worth investigating further the relative perceptions in the NEFSC and DFO surveys. Prior to about 2010, the DFO index is generally above those for the NEFSC surveys. After 2010, the NEFSC indices are consistently above the DFO index. This may be related to the change in vessel and gear in the NEFSC surveys.

Southern New England (SNE)

The Panel agreed the ToR 4 had been almost met for the SNE stock. The recreational LPUE standardization should be updated to include zero trips for the subsequent management track assessment.

The assessment of the SNE stock was presented through the relevant chapter in the main report, as well as via WP 18 and with a helpful and comprehensive presentation to the Panel meeting. The Panel appreciated the extensive work that had gone into the assessment formulation and testing, and noted that data limitations made the finalization of an appropriate assessment complicated and difficult. This is the first year that a specific SNE assessment has been developed, as it was previously considered part of the wider Georges Bank (GB) stock

The SNE cod stock is prosecuted by both commercial and recreational fisheries, with the latter becoming significantly more important in recent years. Age composition sampling in the commercial fleet is limited, with only 7 years covered since 1981, and there has never been age sampling in the recreational fleet (age compositions for the recreational fleet are inferred using ALKs from surveys and available commercial data). The Panel considers that the use of an age-based assessment method (WHAM) for a stock for which ages are not sampled for the principal fishing fleet is ambitious, although the meeting presentation highlighted several other stocks in the area for which similar problems have not led to assessment rejection. The Panel noted that the development work for the assessment had commenced with work with Stock Synthesis (SS), to try and accommodate the lack of age sampling, but this model exhibited poor fit to recreational length data and WHAM had been used as a functional alternative (and for consistency with other NE cod stocks).

Survey indices and biological data are also rather limited for this stock. Four survey indices are available, but three are inshore juvenile surveys which were deemed to be unrepresentative of the wider stock. The remaining survey (NEFSC spring and fall) covers the area comprehensively, but in several years catches very few (or zero) cod and consequently does not track cohort strength very well. There are also two LPUE series, for the commercial and recreational fleets. Concerns were expressed about the commercial LPUE index due to potential hyperstability, and

it was not used. The recreational LPUE index was used, as it tracked cohort strength well (despite ages not being sampled in the recreational fleet) and led to reasonable model convergence and diagnostics. The Panel expressed reservations about this index (see ToR 3 above), but agreed that its use was reasonable for now: however, t**he Panel recommends that this index be re-estimated with zero hauls included** (this couldn't be done during the meeting as it requires agreement on the approach to take, and new data collation). Due to lack of information, weights-at-age, maturity-at-age and natural mortality are assumed time-invariant (although they do all vary by age).

The stepwise development of the model was presented to the Panel. The initial runs using SS and ASAP proved to be unsuccessful, with poor prediction skill and residual patterns in the former, and slow convergence and poor retrospective bias in the latter. Moving to WHAM, the treatment of the catch in two separate "fleets" (commercial and recreational) seemed reasonable and improved retrospective bias. The inshore surveys covering mostly juvenile fish were not retained, as they led to worse model diagnostics. The main NEFSC survey was not split according to vessel, as this split also reduced model fit. The assumption of flat-topped selectivity at age 2 for the NEFSC survey, age 3 for the recreational catch and LPUE, and age 4 for the commercial catch also seemed reasonable. The final model run included a Beverton-Holt recruitment assumption, which was well-supported by the data. Finally, alternative distributional assumptions were tried for random effects on numbers-at-age, but the standard multinomial assumption was retained (in contrast to other stocks in the area).

The Panel concluded that model fits looked appropriate, a**lthough there is some evidence for a selectivity switch in recreational catch data and LPUE around 2000, and a selectivity block at that year could be considered for future work**. Retrospective bias is very low in the final proposed assessment. The Panel observed that the simulation self-test results were quite poor, with bias over 15% in SSB re-estimation. There appears to be no criteria to determine how much bias is too much with these tests, so it is difficult to use them as criteria to accept or reject an assessment, but it is hard to explain this outcome and t**he Panel recommends that consideration be given to this issue in future work**.

The Panel noted that, if this was an ICES assessment, there would be serious consideration of reverting to Category 5 advice (based on catches only), due to significant gaps in data availability and the number of data fill-ins required. However, the Panel agreed that retaining the assessment as the basis for management advice is warranted in this case, as catch-only advice in this area would not encourage management action, and further development of the assessment requires an extant assessment to develop.

Eastern Gulf of Maine (EGOM)

The Panel agreed the ToR 4 had not been fully met for the EGOM stock because of the poor fit of several indices and the absence of estimates of dead discards in the lobster trap fishery. The Panel recommends that these issues be addressed before the next management track process.

The assessment of the EGOM stock was presented through the relevant chapter in the main report, as well as via WP 16 and with a helpful and comprehensive presentation to the Panel meeting. The Panel appreciated the extensive work that had gone into the assessment formulation and testing, and noted that data limitations made the finalization of an appropriate assessment complicated and difficult. This is the first year that a specific EGOM assessment has been developed, as it was previously considered part of the wider Gulf of Maine (GOM) stock

As for the other stocks in this Research Track Peer Review, this is a new assessment unit. An age-structured assessment model was developed for the period 1981-2021, using data from commercial and recreational fisheries, as well as from five fishery-independent surveys. Given that a large majority of GOM cod catch (as well as fishery-dependent samples) came from the WGOM, the data available to characterize the EGOM fishery is relatively sparse. However, given five age-structured survey datasets and fishery catch-at-age data from multiple decades, it was deemed worthwhile to pursue an age-structured assessment model.

Catches have historically been small and about 1 mt in 2021 (Figure 1 of WP16)

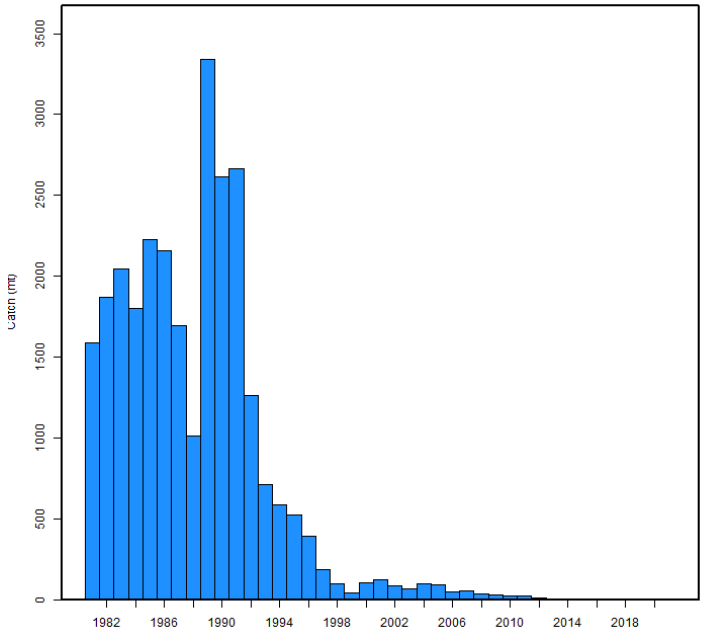

Figure 2 (source: Figure 1 from WP 16 for the WG). Total aggregate catch for EGOM cod, including commercial landings and discards, recreational landings and discards. The estimated value for 2021 is 1 mt. Lobster trap discards, a potential major source of catch, are not accounted for in the figure.

Dead discards in the lobster trap fishery are estimated to have ranged between 2.5 and 10 mt during 2006-2013. While the WG hoped to include updated dead discards in the assessment, those estimates were not provided in time. **The Panel recommends that the assessment be updated with estimates of dead discards in the lobster trap fishery before the management track assessment.**

The Panel suggested re-assessing the criteria to choose which surveys to include in the model. The NEFSC Spring and Autumn surveys seemed to show trends that are different from those in

the other surveys. The model could be run with only the NEFSC Spring and Autumn surveys to compare with a model run using only the other surveys. In this context, the Panel noted that the NEFSC Spring and Autumn surveys, in theory, should be the best, but in fact they had the worst fit in the model. This may be related to the age compositions of the fishery catches used.

5. Update or redefine status determination criteria (SDC; point estimates or proxies for BMSY, $B_{THRESHOLD}$, 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.

The Panel agreed that this ToR was met for all four stocks.

In the previous benchmark of the full Gulf of Maine cod stock, MSY reference points were based on a spawning potential ratio (SPR) of 40% with a lack of a stock-recruit relationship as motivation for the proxy-based approach. The WG had re-examined support for stock-recruit relationships under the new spatial units but concluded the relationships were not well determined and/or not useful to make projections in all cases and using F40% and SSBF40% as proxy reference points was justified for all four stocks.

The results of work conducted under ToR 1 were considered for determining the appropriate time frame for characterizing growth, maturity, natural mortality and recruitment in the calculation of reference points.

Western Gulf of Maine (WGOM)
The panel agrees that the linear stock-recruit relationship is not useful for projections and justifies use of F40% and SSBF40% proxy reference points for the WGOM stock. To calculate the reference points the WG used the most recent five years of weight-at-age and maturity-at-age for the WGOM stock and this was easily justified by the figures for weight-at-age and results for maturity at age presented in the report. Time-varying natural mortality was examined as annual process errors in *M*, but no such runs converged. A time invariant *M* was therefore used to inform reference point calculation. As for the other three stocks, the WG used the full time series of modeled recruitments. Results of a changepoint analysis on NEFSC spring survey data indicated a change point in 2010 for the WGOM stock for both recruitment and recruitment rate (R/SSB) but the R/SSB time series from the proposed assessment model suggested variation without trend. Inclusion in the assessment model of three temperature-based covariates had failed to demonstrate a link between these and recruitment. The panel agrees with the use of the full time series of modeled recruitments and notes it is proposed the projections for WGOM include an autocorrelated random effect for recruitment which should restrict recruitments in early prediction

years to values similar to those late in the time series. The panel, however, **recommends** this issue be re-examined in future, possibly as part of a "broader discussion (i.e., across species) of appropriate methods for defining time windows to characterize prevailing conditions for stocks" as recommended by the WG. For selectivity the WG used the final five-year average of total F-at-age scaled by the maximum F-at-age and this reflects the balance of catch between the commercial and recreational fisheries.

Georges Bank (GB)

The Panel agreed that the status determination criteria and procedures were appropriate. This included using 5-year averages for maturities and weight-at-age, the commercial fishery selectivity in the final year (no recreational fleet), the assessment model M = 0.29, and mean recruitment for the full time-series.

Southern New England (SNE)

F40% and SSBF40% were proposed as proxy reference points for the SNE stock. The SNE assessment does include a relatively well-supported Beverton-Holt curve, but the steepness parameter *h* had to be set with reference to the literature (Myers et al 1999) as it proved to be numerically uncertain, and the WG concluded that the stock-recruit curve could not be used to determine MSY-based reference points. Furthermore, data limitations led to all biological parameters for reference point determination to be time-invariant (stock weight-at-age, maturity, natural mortality, fleet selectivity, and weight-at-age by fleet). The Panel agrees with these conclusions. Application of the standard F40% methodology then led to the required estimate of F40%, and corresponding proxy values for MSY and SSB(MSY).

Eastern Gulf of Maine (EGOM)

The Panel agreed that the status determination criteria and procedures were appropriate. This included using 5-year averages for maturities and weight-at-age, the terminal fishery selectivity, the assessment model M = 0.25, and mean recruitment for the full time-series.

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 Panel agreed this ToR was met for all four stocks.

The WG recommended that WHAM, the proposed assessment model framework for all four Atlantic cod stocks, which includes the capacity to conduct short-term projections internally, should be used for short term projections. The proposed candidate model run for each stock

would be used as a basis for short-term projections. The assumptions of recruitment, growth, maturity, natural mortality, and selectivity that inform stochastic projections of stock size and catches for 2022-2024 would use the same approach as used for the definition of reference points under ToR 5. The use of random effects varied by stock: for WGOM, NAA (2dar1) random effects, for GB NAA with iid random effects, for SNE, iid random effects on R, and for EGOM ar1 were used for recruitment in the model, but the Panel did not find what was used for projections. The panel **recommends** that in the future, projection assumptions are spelled out clearly for all stocks as was done for the GB stock, i.e., including the type of random effect specified.

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.

The Panel agreed that this ToR has been met. The WG reviewed previous recommendations and updated their status. The WG also made new research recommendations. The Panel suggests below a prioritization of those new research recommendations as: i) necessary for the management track, ii) high priority, iii) medium/long term and iv) low priority.

Necessary for management track

1. The WG recommends further work to estimate cod discards from the lobster fishery, development of a workflow for sustainability of this data product into the future, and work to integrate this information into WGOM and EGOM stock assessments. *This recommendation was reiterated in a slightly different form under EGOM:* The WG recommends development of methods to account for cod bycatch in the lobster fishery over the full assessment time series, either by estimating removals-at-age or via process error (e.g., random effects on NAA).
2. The WG recommends continued development of the landings per unit effort (LPUE) time series for the SNE stock. The WG recommends focusing on standardization methods that account for zero observations, new regulations in the fishery for 2022-2023 and spatial changes in the fishery (e.g., spatio-temporal model). Future studies should also explore preferential sampling.
3. The WG recommends investigation of the magnitude of recreational catch in MRIP wave one for the SNE stock.
4. The WG recommends exploration of splitting the NEFSC trawl time series (i.e., Albatross/Bigelow split time series) to evaluate whether this improves model fit and performance for Georges Bank cod. *The Panel notes that this could also be investigated for EGOM cod*.

High Priority

1. The WG recommends new field studies be conducted throughout the year to refine estimates of recreation discard mortality for the SNE stock.
2. The WG recommends expanded efforts to obtain length, weight, and age samples of cod across stocks from the commercial and recreational fishery. *The Panel notes that sampling was originally designed for two management units. The sampling scheme should be reviewed for the four stocks currently used.*
3. The WG recommends future exploration of the utility of indices of abundance that integrate across multiple surveys using tools such as Vector Autoregressive Spatio-Temporal Models.
4. The WG recommends alternative efforts to supplement the collection of fishery-dependent samples to characterize the catch-at-age (e.g., MEDMR or industry collections) *for EGOM.*
5. The WG recommends that a broader discussion (i.e., across species) [of] appropriate methods for defining time windows to characterize prevailing conditions for stocks is needed in the region. This may be most appropriate to explore in a thematic research track focused on biological reference points.

Medium/long term

1. The WG recommends continued work to consider the impacts of changing ocean conditions on aspects of stock productivity (e.g., natural mortality and recruitment) and how to incorporate these effects in short-term projections.

2. The WG recommends further exploration of density dependent effects and evaluation of the likelihood of spurious relationships identified between ecosystem drivers and aspects of stock dynamics.
3. The WG recommends future re-evaluation of ecosystem-stock variable relationships for cod stocks as more data become available and the analysis can be explored over a longer time series of data.
4. The WG came to consensus on the spatial allocation of fishery data to revised stock areas. However, the WG recommends that these data decisions continue to be evaluated and confirmed during future stock assessment processes (e.g., monitor changes in the importance of recreational fishery on GB).
5. The WG recommends empirical validation of fishery length-weight and gutted-to-whole conversion factors.
6. The WG came to consensus on the revision to survey strata assignment to stock area. However, it was noted that there could be a need to rethink strata assignment with changes in the biomass of Atlantic cod over time. The WG recommends that survey strata assignments to stock areas, particularly those that cross stock boundaries, should be revisited during future research track assessments.
7. The WG recommends continued work to consider the impact of changing ocean conditions on aspects of stock productivity (natural mortality and recruitment). *This recommendation was listed under GB but should apply to all stocks.*
8. The WG recommends further development of the candidate WHAM model for the SNE stock including exploration of: 1) environmental covariates on all population dynamics processes, 2) incorporation of time-varying dynamics (random effects on recruitment, numbers at age, natural mortality), 3) time-varying selectivity, and 4) standardizing the recreational landings per unit effort (LPUE) time series in the model (e.g., random effects on catchability).
9. The WG recommends further development of the candidate WHAM model for EGOM. Most modeling decisions for EGOM occurred in the ASAP software framework. The various types of process errors that can be accounted for in WHAM should be more fully explored.
10. The following are two recommendations combined: i) the WG recommends continued efforts to develop and apply techniques to differentiate spring and winter spawners in the catch and survey data, ii) The WG recommends development of a work plan outlining the investments and protocols needed to implement this approach to discriminate mixed stocks in this region.
11. The WG recommends monitoring potential distributional shifts and regularly evaluating their impact on survey catchability *for the GB stock.*

Low priority

1. The WG recommends continued monitoring for changes in Atlantic cod stock dynamics and distribution and evaluation of whether there are improvements to model performance by including environmental covariates within the models or further partitioning variance explained through inclusion of process error.

2. The WG recommends further evaluation of potential ecosystem and climate influences on the SNE stock as limited data availability prevented extensive analysis during the research track.
3. The WG recommends continued development of methods to quantify the amount of catch that has been spatially misallocated across stocks due to fishing behaviour (e.g., Palmer 2017, Hayes and Demerest, work in progress).
4. The WG recommends that when sufficient data is available, attempts are made to apply the multi-stock feature of WHAM, or some other multi-stock assessment model, to separate spring and winter spawning Atlantic cod.
5. The WG recommends monitoring of FSV Bigelow selectivity estimates in future stock assessments and consideration of whether reverting to the empirical calibration estimates of NEFSC trawl time series for WGOM if justified.
6. The WG recommends further evaluation of whether F40% remains an appropriate proxy for FMSY for Atlantic cod.
7. The WG recommends monitoring the accuracy of projections during future assessments and developing methods that continue to improve projections.

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 Panel agreed that this ToR was met for all four stocks.

The Index-Based Research Track Working Group simulation-tested the performance of several empirical Index Based Methods (IBMs) (NEFSC 2020, Legault et al. 2023). The Index-Based Research Track WG concludes: "Overall, none of the IBMs considered in these simulations performed better than the rho-adjusted statistical catch-at-age (SCAA) model. So in situations where an SCAA model is rejected due to a strong retrospective pattern, there should not be an expectation that an IBM will perform better than the rejected model." (NEFSC 2020). The WG recommended that if the proposed WHAM assessment approach is rejected, an alternative simpler WHAM model be developed to integrate information from catch, age composition, and indices and that a retrospective adjustment be applied to the terminal year estimates of F and SSB. Assessment teams did not provide specific backup runs (fits, diagnostics, etc.) for the Panel to review, but the Panel agrees that the proposed general approach seemed reasonable for the four Atlantic cod stocks, while observing that the alternative simpler WHAM model is likely to present worse diagnostics and performance than the proposed WHAM model. The Panel did not review specific backup runs proposed by assessment teams.

9. Apply the findings of the Atlantic Cod Stock Structure Working Group and identify what assessment approaches the available data can support in defining the appropriate scale of Atlantic cod stock assessment. Consider implications for management processes and other practical limitations in the final units and boundaries used for stock assessments.

The Panel agreed that this ToR has been met. Based on the results of the Atlantic Cod Stock Structure Working Group, the Atlantic Cod Research Track Working Group suggested combining two of the five biological units, the winter and spring spawning components, into one assessment unit for the Western Gulf of Maine, mostly because of difficulties allocating existing data (fishery as well as research) to each spawning unit. The other biological units identified by the stock structure WG were adopted as assessment units.

Past studies had indicated significant mixing between the eastern Gulf of Maine and the Bay of Fundy. No recent tagging work has shown mixing, possibly because there are not as many fish to mix.

The biological units identified are believed to be biologically discrete. Spawning site fidelity assures reproductive isolation, but there may be some non-reproductive mixing although at no time are the stocks fully separate from each other.

Larval drift studies indicate that winter-spawned larvae sometimes move around Cape Cod and there is some entrainment of eggs/larvae that were spawned on or near Georges Bank on the Bank because of the gyre there. Generally, in the Gulf of Maine and on Georges Bank, larvae are transported from the NE to the SW. There is connectivity from the GOM to other areas, but not from other areas into GOM.

There were clearly some signs of mis-specification in the cod assessments in the past and the Panel agrees that aligning the stocks with the assessment units is a significant step towards improving the assessments.

Panel Recommendations

1. The Panel recommends targeted sampling in time and areas of cod and seal overlap to better inform seal predation effects in any future work on indicators. (Hammill et al., 2014).
2. The Panel considers that the development of integrated modeling for surveys (e.g. VAST) should be continued, especially considering the postulated distributional shift in cod on GB in spring and concerns about conflicting signals from the NEFSC and DFO spring surveys for this stock.
3. The Panel recommends research continue to standardize and quantify the uncertainty and sampling distribution of age compositions (e.g., Thorson, 2014; Thorson and Haltuch, 2019; Thorson et al., 2023)

4. The Panel recommended that the lack of convergence of WHAM configurations be explored more fully in future research track processes (for example, by resetting parameter bounds if lack of convergence is due to bounds being reached).
5. The Panel recommends that it would be useful if the WHAM model provided leave-one-survey-out diagnostics.
6. For WGOM, the Panel recommends a four block logistic-age based version of the model be tested in future.
7. The Panel recommends further investigations of robust approaches to modeling recruitment in the assessment model. This recommendation was made for WGOM but would apply to all four stocks.
8. The Panel recommends that the effect of removing the correlation between age 1 process errors and those for ages 2-9+ be considered in a future assessment for WGOM.
9. The Panel recommends additional evaluation of the age for a plus group should be conducted, since survey indices and catch information at older model ages did not track cohorts well for most stocks.
10. For GB cod:
    a. Further explorations of differences in NEFSC and DFO survey catch rates during 2000-present are required.
    b. Investigate a model using unconverted Bigelow indices, similar to WGOM cod. If the differences in NEFSC spring and fall survey indices and the Canadian indices cannot be resolved, then two WHAM models should be formulated (one for US indices, one for Canadian indices) to investigate differences in assessment results and management advice.
    c. Jan1 and SSB weights-at-age were calculated using the Rivard approach applied to commercial weights. However, these weights are known to be over-estimates of the stock weights at young ages because of the size-selectivity of commercial fisheries. Survey weights-at-age may be better for stock weights because survey gears may be less size-selective. This should be investigated.
11. The Panel recommends that the recreational LPUE index for SNE cod be re-estimated with zero hauls included.
12. The Panel recommends that criteria be agreed to determine how much bias in self-tests is too large and when this should lead to rejection.
13. The Panel recommends that the poor fit of several indices and the absence of estimates of dead discards in the lobster trap fishery be addressed and that the assessment of EGOM cod must be updated with estimates of dead discards in the lobster trap fishery before the management track assessment.
14. The panel recommends that the issue of what historical period to use in estimating future recruitment in projections be re-examined in the future, possibly as part of a "broader discussion (i.e., across species) of appropriate methods for defining time windows to characterize prevailing conditions for stocks" as recommended by the WG.
15. The panel recommends that in the future, projection assumptions are spelled out clearly for all stocks as was done for the GB stock, i.e., including the type of random effect specified.
16. The Panel suggests that age-length-based models that can use length data directly as sampled should be investigated in the medium-long term.

References

Aldrin, M., Aanes, F.L., Tvete, I.F., Aanes, S. and Subbey, S., 2021. Caveats with estimating natural mortality rates in stock assessment models using age aggregated catch data and abundance indices. *Fisheries Research*, *243*, p.106071.

Babyn, J., Varkey, D., Regular, P., Ings, D., Flemming, J.M. 2021. A gaussian field approach to generating spatial age length keys. Fisheries Research, Volume 240, https://doi.org/10.1016/j.fishres.2021.105956 .

Cope JM. 2013. Implementing a statistical catch-at-age model (Stock Synthesis) as a tool for deriving overfishing limits in data-limited situations. Fisheries Research 142: 3– 14.

Hammill, M.O., Stenson, G.B., Swain, D.P. and Benoît, H.P., 2014. Feeding by grey seals on endangered stocks of Atlantic cod and white hake. *ICES Journal of Marine Science*, *71*(6), pp.1332-1341.

Legault, C.M., Wiedenmann, J., Deroba, J.J., Fay, G., Miller, T.J., Brooks, E.N., Bell, R.J., Langan, J.A., Cournane, J.M., Jones, A.W., Muffley, B. 2023. Data Rich but Model Resistant: An Evaluation of Data-Limited Methods to Manage Fisheries with Failed Age-based Stock Assessments. Can. J. Fish. Aquat. Sci. 80: 27–42.

McBride RM, and RK Smedbol (Eds.). 2022. An Interdisciplinary Review of Atlantic Cod (Gadus morhua) Stock Structure in the Western North Atlantic Ocean. NOAA Technical Memorandum NMFS-NE-273.

Miller TJ. 2013. A comparison of hierarchical models for relative catch efficiency based on paired-gear data for US Northwest Atlantic fish stocks. Can J Fish Aquat Sci. 70(9):1306–1316. https://doi.org/10.1139/cjfas-2013-0136

Myers RA, Bowen KG and Barrowman NJ. 1999. Maximum reproductive rate of fish at low population sizes. Can. J. Fish. Aquat. Sci. 56: 2404–2419.

Punt, A.E., Castillo-Jordán, C., Hamel, O.S., Cope, J.M., Maunder, M.N. and Ianelli, J.N., 2021. Consequences of error in natural mortality and its estimation in stock assessment models. *Fisheries Research*, *233*, p.105759.

Rudd MB, Cope JM, Wetzel CR and Hastie J (2021) Catch and Length Models in the Stock Synthesis Framework: Expanded Application to Data-Moderate Stocks. Front. Mar. Sci. 8:663554. doi: 10.3389/fmars.2021.663554.

Thorson, J.T., 2014. Standardizing compositional data for stock assessment. ICES Journal of Marine Science, 71(5), pp.1117-1128.

Thorson, J.T. and Haltuch, M.A., 2019. Spatiotemporal analysis of compositional data: increased precision and improved workflow using model-based inputs to stock assessment. Canadian Journal of Fisheries and Aquatic Sciences, 76(3), pp.401-414.

Thorson, J.T., Monnahan, C.C. and Hulson, P.J.F., 2023. Data weighting: An iterative process linking surveys, data synthesis, and population models to evaluate mis-specification. Fisheries Research, 266, p.106762.

Wetzel CR and Punt AE (2011) Performance of a fisheries catch-at-age model (Stock Synthesis) in data-limited situations. Marine and Freshwater Research 62: 927–936.

Appendix 1 - Terms of Reference for Atlantic Cod Research Track Stock Assessment

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.
9. Apply the findings of the Atlantic Cod Stock Structure Working Group and identify what assessment approaches the available data can support in defining the appropriate scale of Atlantic cod stock assessment. Consider implications for management processes and other practical limitations in the final units and boundaries used for stock assessments.

Appendix 2 – Initial agenda for Atlantic cod Research Track Assessment Peer Review meeting, July 31-August 3, 2023.

Monday, July 31, 2023

Time	Topic	Presenter(s)	Notes
12:00 p.m. - 12:15 p.m.	Welcome/Logistics Introductions/Agenda/Conduct of Meeting	Michele Traver, Assessment Process Lead Russ Brown, PopDy Branch Chief JJ Maguire, Panel Chair	
12:15 p.m. - 12:45 p.m.	Introduction and Overview	Lisa Kerr (WG Chair)	
12:45 p.m. - 1:45 p.m.	Term of Reference (TOR) #9	Lisa Kerr/Rich McBride	

Time	Topic	Presenter(s)	Notes
1:45 p.m. - 2:45 p.m.	TOR #1	Scott Large/Jamie Behan	Ecosystems
2:45 p.m. - 3:00 p.m.	Break		
3:00 p.m. - 4:30 p.m.	TOR #2	Charles Perretti/Kathy Sosebee	Catch
4:30 p.m. - 5:00 p.m.	Discussion/Summary	Review Panel	
5:00 p.m. - 5:15 p.m.	Public Comment	Public	
5:15 p.m.	Adjourn		

Tuesday, August 1, 2023

Time	Topic	Presenter(s)	Notes
12:00 p.m. - 12:05 p.m.	Welcome/Logistics	Michele Traver, Assessment Process Lead JJ Maguire, Panel Chair	
12:05 p.m. - 12:30 p.m.	Wrap up TOR #2		
12:30 p.m. - 1:00 p.m.	WHAM overview	Tim Miller	
1:00 p.m. - 2:30 p.m.	TOR #3	Lisa Kerr	Survey Data
2:30 p.m. - 2:45 p.m.	Overview of approach for TORs #4-6 and #8	Lisa Kerr	All stocks
2:45 p.m. - 3:00 p.m.	Break		
3:00 p.m. - 4:30 p.m.	TORs #4-6 and #8	Charles Perretti	WGOM - Models, BRPs, Projections, and Alternative Assessment Plan

Time	Topic	Presenter(s)	Notes
4:30 p.m. - 5:00 p.m.	Discussion/Summary	Review Panel	
5:00 p.m. - 5:15 p.m.	Public Comment	Public	
5:15 p.m.	Adjourn		

Wednesday, August 2, 2023

Time	Topic	Presenter(s)	Notes
12:00 p.m. - 12:05 p.m.	Welcome/Logistics	Michele Traver, Assessment Process Lead JJ Maguire, Panel Chair	
12:05 p.m. - 2:00 p.m.	TORs #4-6 and #8	Amanda Hart	GB - Models, BRPs, Projections, and Alternative Assessment Plan
2:00 p.m. - 2:15 p.m.	Break		
2:15 p.m - 4:15 p.m.	TORs #4-6 and #8	Alex Hansell and Steve Cadrin	SNE - Models, BRPs, Projections, and Alternative Assessment Plan
4:15 p.m. - 4:45 p.m.	Discussion/Summary	Review Panel	
4:45 p.m. - 5:00 p.m.	Public Comment	Public	
5:00 p.m.	Adjourn		

Thursday, August 3, 2023

Time	Topic	Presenter(s)	Notes
12:00 p.m. - 12:05 p.m.	Welcome/Logistics	Michele Traver, Assessment Process Lead JJ Maguire, Panel Chair	
12:05 p.m. - 2:00 p.m.	TORs #4-6 and #8	Micah Dean	EGOM - Models, BRPs, Projections, and Alternative Assessment Plan
2:00 p.m. - 2:15 p.m.	Break		
2:15 p.m. - 3:15 p.m.	TOR #7	Lisa Kerr	Research Recommendations
3:15 p.m. - 4:15 p.m.	Panel Wrap-up and Discussion/Summary	Review Panel	
4:15 p.m. - 4:30 p.m.	Public Comment	Public	
4:30 p.m.	Adjourn		

Appendix 3 - Performance Work Statement (PWS) - Center for Independent Experts (CIE) Program – Atlantic Cod Research Track 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
Atlantic cod are an iconic species in New England and in recent years, cod stocks have experienced dramatic declines. Both of the currently managed stocks in the U.S., Gulf of Maine and Georges Bank, are overfished and managed under rebuilding plans established to promote population growth. Reductions on fishing rates as well as area closures and gear restrictions are all used to help manage cod stocks. Both the Gulf of Maine and Georges Bank stocks of cod were last assessed in 2021. The Gulf of Maine stock utilized an analytical assessment model, while the Georges Bank assessment utilized empirical methods. In 2018, an Atlantic Cod Stock Structure Working Group was formed to determine the most appropriate representation of cod stock structure for use in regional stock assessments. The findings of this group identified five biological stocks in U.S. waters. The Atlantic Cod Research Track was convened to consider the implications of these newly proposed biological cod stocks and attempt to develop analytical stock assessments to support practicable management actions.

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 Atlantic cod stocks. The requirements for the peer review follow. This PWS also includes: Annex 1: Terms of Reference (TORs) for the research track, which are the responsibility of the analysts; Annex 2: a draft meeting agenda; Annex 3: Individual Independent Review Report Requirements; and Annex 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, model diagnostics from various population models, and how catch advice is provided from stock assessment models. In addition, knowledge and experience with simulation analyses is helpful.

Tasks for Reviewers

- Review the background materials and reports prior to the review meeting
    - Two weeks before the peer review, the Assessment Process Lead will electronically disseminate all necessary background information and reports to the CIE reviewers for the peer review.
- Attend and participate in the panel review meeting
    - The meeting will consist of presentations by NMFS 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
- 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. Individual reviewer perspectives should be provided in their individual reports, and any lack of consensus should be clearly described in the panel's summary report.
- Each reviewer shall assist the Peer Review Panel Chair with contributions to the Peer Review Panel's Summary Report.
- Deliver individual Independent Reviewer Reports to NMFS 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 Atlantic Cod 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. Again, the CIE reviewers are not required to reach a consensus. 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 remote, via WebEx video conferencing.

Period of Performance
The period of performance shall be from the time of award through October 2023. 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.

Timeline	Action
Within 2 weeks of award	Contractor selects and confirms reviewers
Approximately 2 weeks later	Contractor provides the pre-review documents to the reviewers
July 31 – August 3, 2023	Panel review meeting
Approximately 2 weeks later	Reviewers submit draft peer-review reports to the contractor for quality assurance and review
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.

1. 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

Annex 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.

9. Apply the findings of the Atlantic Cod Stock Structure Working Group and identify what assessment approaches the available data can support in defining the appropriate scale of Atlantic cod stock assessment. Consider implications for management processes and other practical limitations in the final units and boundaries used for stock assessments.

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.

Annex 2. Draft Review Meeting Agenda

Atlantic Cod Research Track Assessment Peer Review Meeting

July 31 - August 3, 2023

WebEx link: https://noaanmfs-meets.webex.com/noaanmfs-meets/j.php?MTID=m35f7a63c1ebaa546af3e814f1a269e1d

Meeting number (access code): 2762 857 0886
Meeting password: PAvVKXGV333

Phone: +1-415-527-5035 US Toll

AGENDA* (v. 7/28/2023)

**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 31, 2023

Time	Topic	Presenter(s)	Notes
12:00 p.m. - 12:15 p.m.	Welcome/Logistics Introductions/Agenda /Conduct of Meeting	Michele Traver, Assessment Process Lead Russ Brown, PopDy Branch Chief JJ Maguire, Panel Chair	
12:15 p.m. - 12:45 p.m.	Introduction and Overview	Lisa Kerr (WG Chair)	
12:45 p.m. - 1:45 p.m.	Term of Reference (TOR) #9	Lisa Kerr/Rich McBride	
1:45 p.m. - 2:45 p.m.	TOR #1	Scott Large/Jamie Behan	Ecosystems
2:45 p.m. - 3:00 p.m.	Break		
3:00 p.m. - 4:30 p.m.	TOR #2	Charles Perretti/Kathy Sosebee	Catch
4:30 p.m. - 5:00 p.m.	Discussion/Summary	Review Panel	

Time	Topic	Presenter(s)	Notes
5:00 p.m. - 5:15 p.m.	Public Comment	Public	
5:15 p.m.	Adjourn		

Tuesday, August 1, 2023

Time	Topic	Presenter(s)	Notes
12:00 p.m. - 12:05 p.m.	Welcome/Logistics	Michele Traver, Assessment Process Lead JJ Maguire, Panel Chair	
12:05 p.m. - 1:30 p.m.	TOR #3	Lisa Kerr	Survey Data
1:30 p.m. - 1:45 p.m.	Overview of approach for TORs #4-6 and #8	Lisa Kerr	All stocks
1:45 p.m. - 2:15 p.m.	WHAM overview	Tim Miller	
2:15 p.m. - 3:15 p.m.	TORs #4-6 and #8	Charles Perretti	WGOM - Models, BRPs, Projections, and Alternative Assessment Plan
3:15 p.m. - 3:30 p.m.	Break		
3:30 p.m. - 4:30 p.m.	TORs #4-6 and #8 cont.	Charles Perretti	WGOM - Models, BRPs, Projections, and Alternative Assessment Plan
4:30 p.m. - 5:00 p.m.	Discussion/Summary	Review Panel	
5:00 p.m. - 5:15 p.m.	Public Comment	Public	
5:15 p.m.	Adjourn		

Wednesday, August 2, 2023

Time	Topic	Presenter(s)	Notes
12:00 p.m. - 12:05 p.m.	Welcome/Logistics	Michele Traver, Assessment Process Lead	

Time	Topic	Presenter(s)	Notes
		JJ Maguire, Panel Chair	
12:05 p.m. - 2:00 p.m.	TORs #4-6 and #8	Amanda Hart	GB - Models, BRPs, Projections, and Alternative Assessment Plan
2:00 p.m. - 2:15 p.m.	Break		
2:15 p.m - 4:15 p.m.	TORs #4-6 and #8	Alex Hansell and Steve Cadrin	SNE - Models, BRPs, Projections, and Alternative Assessment Plan
4:15 p.m. - 4:45 p.m.	Discussion/Summary	Review Panel	
4:45 p.m. - 5:00 p.m.	Public Comment	Public	
5:00 p.m.	Adjourn		

Thursday, August 3, 2023

Time	Topic	Presenter(s)	Notes
12:00 p.m. - 12:05 p.m.	Welcome/Logistics	Michele Traver, Assessment Process Lead JJ Maguire, Panel Chair	
12:05 p.m. - 2:00 p.m.	TORs #4-6 and #8	Micah Dean	EGOM - Models, BRPs, Projections, and Alternative Assessment Plan
2:00 p.m. - 2:15 p.m.	Break		
2:15 p.m. - 3:15 p.m.	TOR #7	Lisa Kerr	Research Recommendations

Time	Topic	Presenter(s)	Notes
3:15 p.m. - 4:15 p.m.	Panel Wrap-up and Discussion/Summary	Review Panel	
4:15 p.m. - 4:30 p.m.	Public Comment	Public	
4:30 p.m.	Adjourn		

Annex 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.

Annex 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. It should also include whether they **accept or reject** the work that they reviewed, with an explanation of their decision (strengths, weaknesses of the analyses, etc.)

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 4 - Materials provided or referenced during the Atlantic Cod Research Track Stock Assessment Peer Review meeting

Working papers and presentations were available on a NEFSC website (https://apps-nefsc.fisheries.noaa.gov/saw/sasi.php) by selecting the species and year of assessment.

Working Papers:

WP1 Stakeholder Meeting Summary
WP2 Development of Ecosystem Indicators
WP3 Environmental Influences on Cod
WP4 Stakeholder Meeting 2 Summary
WP5 Rec Discard Mortality
WP6 FDD Exploration

WP7 Cod LPUE
WP8 NEFSC Trawl Survey Expanded Figs
WP9 Survey Time Series Correlations
WP10 Integrated Survey Indices (VAST)
WP11 EGOM Sentinel Index Modification
WP12 Time Varying Cod Maturity
WP13 Time Varying Cod LW
WP14 Estimating Cod M by Stock
WP15 Atlantic Cod Model Selection Procedure
WP16 EGOM Assessment Model ToR 4
WP17 WGOM Assessment Model ToR 4
WP18 SNE Assessment Model ToR 4
WP19 GB Assessment Model ToR 4
WP20 Reference Points

Presentations

Atlantic Cod Research Track Stock Assessment. Lisa Kerr
ToR 9 - Stock Structure. Lisa Kerr and Rich McBride
ToR 1 - Ecosystem and Climate Influences. Scott Large and Jamie Behan
ToR 2 - Fishery Data. Charle Perretti and Kathy Sosebee
Data Processing Methods [additional presentation to address Panel questions]. Lisa Kerr
ToR 3 - Survey Data. Lisa Kerr and Steve Cadrin
ToR 4,5,6, & 8 – Assessment, reference points, projections, back-up assessment. Lisa Kerr
Woods Hole Assessment Model. Tim Miller
Western Gulf of Maine Cod (ToRs 4, 5, 6, 8). Charles Perretti
Cohort Tracking Diagnostics and R/SSB by Stock [additional presentation to address Panel questions]. Lisa Kerr
Georges Bank Cod (ToRs 4, 5, 6, 8). Amanda Hart
Southern New England Cod (ToRs 4, 5, 6, 8). Alex Hansell, Cole Carrano, Steve Cadrin
EGOM Cod Assessment Model (ToRs 4, 5, 6, 8). Micah Dean
Age composition likelihood comparisons for GB candidate model (19A) [additional presentation to address Panel questions]. Amanda Hart
ToR 7 - Research Recommendations. Lisa Kerr.
Standardizing Landings per Unit Effort from Cod Fishery Data. Lucy McGinnis, Gavin Fay, Alex Hansell, Steve Cadrin. (Provided to the Review Panel but not presented)

Appendix 5 - Meeting attendees at the Atlantic Cod Research Track Stock Assessment Peer Review meeting

Atlantic Cod Research Track Peer Review Attendance
July 31-August 3, 2023

DFO - Department of Fisheries and Oceans (Canada)
GARFO - Greater Atlantic Regional Fisheries Office

GMRI - Gulf of Maine Research Institute
MADMF - Massachusetts Division of Marine Fisheries
MEDMR - Maine Department of Marine Resources
NEFMC - New England Fisheries Management Council
NEFSC - Northeast Fisheries Science Center
SMAST - University of Massachusetts School of Marine Science and Technology

~~~~~~

JJ Maguire - Chair
Steven Holmes - CIE Panel
Noel Cadigan - CIE Panel
Coby Needle - CIE Panel

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

Alex Dunn - NEFSC
Alex Hansell - NEFSC
Alicia Miller - NEFSC
Alison Frey - NEFSC
Amanda Hart - GMRI
Angela Forristall - NEFMC Staff
Andy Jones - NEFSC
Anna Mercer - NEFSC
Brian Linton - NEFSC
Burton Shank - NEFSC
Caira Clark - Nature Conservancy of Canada
Carla Guenther - Maine Center for Coastal Fisheries
Cate O'Keefe - NEFMC Executive Director
Charles Adams - NEFSC
Charles Perretti - NEFSC
Chris Kellogg - NEFMC Deputy Director
Chris Legault - NEFSC
Cole Carrano - SMAST
Dave McElroy - NEFSC
Doug Butterworth - University of Cape Town (South Africa)
Frank Blount - Frances Fleet
Gareth Lawson - Conservation Law Foundation
Gary Nelson - MADMF
Irene Andruschchenko - DFO
Jackie Odell - Northeast Seafood Coalition
Jamie Behan - GMRI
Jamie Cournane - NEFMC Staff
Jessica Blaylock - NEFSC
John Pappalardo - Cape Cod Hook Fishermen's Association
Jon Deroba - NEFSC
Julie Nieland - NEFSC
~~~~~~

Kathy Sosebee - NEFSC
Katie Lankowicz - GMRI
Kelly Whitmore - MADMF
Kiersten Curti - NEFSC
Kristan Blackhart - NEFSC
Libby Etrie - NEFMC Member
Lisa Hendrickson - NEFSC
Lisa Kerr - University of Maine
Liz Brooks - NEFSC
Liz Sullivan - GARFO
Mark Terceiro - NEFSC
Max Grezlik - SMAST
Melanie Barrett - DFO
Melanie Griffin - MADMF
Micah Dean - MADMF
Nicholas Calabrese - SMAST
Paul Nitschke - NEFSC
Rebecca Peters - MEDMR
Rebecca Rademeyer - Independent Consultant (South Africa)
Rich McBride - NEFSC
Rick Bellavance - NEFMC Member
Robin Frede - NEFMC Staff
Robyn Linner - Stony Brook University
Roger Brothers - GMRI
Scott Large - NEFSC
Spencer Talmage - GARFO
Steve Cadrin - SMAST
Susan Wigley - NEFSC
Tara Dolan - MADMF
Tim Barrett - DFO
Tim Miller - NEFSC
Tim O'Donnell - Gloucester Marine Genomics Institute
Toni Chute - NEFSC
Yanjun Wang - DFO