Center for Independent Experts (CIE) Independent Peer Revie of the NEFSC Atlantic Cod Research Track Stock Assessment

Online meeting, 31 July – 3 August 2023

Dr Coby Needle

September 2023

Executive summary

This report summarises my notes and conclusions on the NEFSC Atlantic Cod Research Track Stock Assessments, presented during the online review panel meeting during 31 July – 3 August 2023. Recent research track work on Atlantic cod stock structure had concluded that cod in the north-east of the United States should be assessed as four stocks, rather than two as previously, and so we were required to review data, assessments, reference points and forecast settings for four separate stocks: Western Gulf of Maine (WGoM), Eastern Gulf of Maine (EGoM), Georges Bank (GB) and Southern New England (SNE). The analyses and results presented before and during the review panel meeting were very extensive, with a collated WG report, 20 working papers, and presentations on all nine ToRs (along with follow-up presentations when requested). The WG team are to be commended for producing work of high quality over an extensive time period, with a diligent attention to detail and clear willingness to help the review panel with any questions that arose.

There were difficult issues to address with these assessments. In several cases the availability of data (particularly age data) was less than ideal, to the extent that I questioned whether the use of a full age-structured stock assessment model was appropriate for the SNE stock in particular. Although extensive justifications were provided, the underlying driver seemed to be to produce a full analytical assessment despite some of the data failings, and I'm not convinced this will necessarily lead to the best outcome for stock sustainability. The decision to proceed with four separate stock assessments for stocks which are likely to be mixing to a certain extent out with the spawning season was interesting, and in future work I would like to see consideration of a more multistock approach to account for mixing. The absence of a multispecies model on which to base time- and age-varying natural mortality estimates is also a hindrance that could usefully be addressed. Finally, I was unclear as to how we could usefully review the backup models, when they were imprecisely specified and when no real results were presented.

Overall, however, the work that was presented was extensive, very thorough, convincingly argued for the most part, and certainly a great improvement on the previous assessments with their significant retrospective bias problems. While there were areas of the assessment report that I felt could have been presented with more clarity, all relevant questions were addressed in full by the stock assessors during the review meeting, and I concluded (in agreement with the others in the review panel) that the ToRs had nearly all been met to the extent possible and necessary for subsequent management track assessments (the exceptions being some additional assessment analyses that will be required for two of the stocks). The assessments were well-presented by the lead scientists and WG members, contained a great deal of relevant information, and I was happy to accept them as valid representations of stock status. Finally, I think many other parts of the world could learn a great deal from the research track process in the US: it permits collaborative working for an extended period of time, and leads to work of great depth and breadth.

Background

I am an applied mathematician and modeller by training, and I have worked in quantitative fisheries science since 1996. Having served as the Chair of the ICES Working Groups on the Assessment of Demersal Stocks in the North Sea and Skagerrak (WGNSSK, 2004—2006) and Methods of Stock Assessment (WGMG, 2007—2009), I led the Sea Fisheries Programme and the Fisheries Assessment and Advice Programme of Marine Scotland (around 60 staff) during 2015-2023. Marine Scotland was part of the Scottish Government based at the Marine Laboratory in Aberdeen, Scotland. My current role is as the Chief Fisheries Advisor for Scotland, working in the Marine Directorate of the Scottish Government. One of our key roles is the collection, analysis and interpretation of data from the fishing industry and fishery-independent surveys, and the provision of advice on fisheries and fish stocks to fishery managers in Scotland, the UK and Europe. As well as being the principal provider of advice on demersal stocks in particular to the Scottish Government, I also contribute significantly to assessments

for stocks in the North Sea and West of Scotland through the relevant ICES Working Groups. My main research areas lie in the development of fishery-independent assessment and advice methods, remote electronic monitoring of fishery activity, management strategy evaluations, the modelling of fleet dynamics, and the use of stakeholder information in fisheries management and advice.

Role in the Review Activities

Prior to the online Atlantic Cod Research Track Stock Assessment meeting (31 July – 3 August 2023), I thoroughly reviewed the main WG report and all working papers provided for the review panel (a total of 1141 pages - see Appendix 1 below for a document list). During the meeting I participated in full in the plenary discussions during and after the presentations provided, and I took notes during these discussions which form the basis of my comments below. I conducted exploratory survey-based assessments using data provided by the WG, in order to check stock dynamics and to generate additional survey diagnostics. Finally, I contributed in full to the writing of the Panel Summary Report, and wrote this Individual Independent Review Report.

Overall Conclusions and Suggestions for Improvements to the Research Track Process

I found the Research Track Process to be very well organised. The project contact (Michele Travers) was diligent and extremely helpful, and the contributing WG members were all very responsive and quick to contribute any additional analyses that the Panel members requested. Although there were drawbacks with the online nature of the meeting (for example, we missed the parallel discussions and the ability to easily follow up on some difficult topics with WG members), the technology worked very well, and I felt my contribution was fully facilitated. I have known my fellow Panel members for many years, and while we all brought different skillsets and experiences to the process, I felt that our contributions meshed well. I was highly impressed with the level of scientific ability and professionalism shown by the WG members, and in addition I appreciated the degree of collaboration shown between different institutes, covering government, academic and stakeholder organisations.

I did find the lack of strong stakeholder participation in the review meeting interesting. There were a number of stakeholders present, covering management, industry and eNGO interests, but when the floor was opened to public comment there were relatively few contributions made. This is in stark contrast to my experience in Scotland, where such an audience would always be very vocal in a meeting of this kind. I understand that these groups have made very strong contributions throughout the Research Track process, so perhaps it just reflects a cultural difference – or that previous contributions were felt to be sufficient.

One further issue was the volume of material to cover as a reviewer. Wherever and whenever cod is assessed, it always attracts a great deal of attention, and hence the material provided had to be very comprehensive. The downside of this was that we were required to read and review a total of 1141 pages in a two-week period, which proved to be possible but not very conducive for a thoughtful, considered review. It may be appropriate for future meetings to condense the material somewhat, and if anything crucial is missing the reviewers would always be able to request it.

Overall, I concluded that the ToRs had been met in full (or very nearly), and I was content to recommend the final proposed assessments be taken forward to subsequent management track processes with the same caveats as outlined in the Panel report. The switch from two to four stocks in the north-eastern US area seems to be well-supported, and the resulting assessments should be more appropriate than before. In particular, the argumentation for using the full time-series for recruitment as inputs to reference point estimation was strong.

On the other hand, I found that there was an element of being perhaps over-keen to generate age-based quantitative assessments, when in some cases the data were maybe not strong or

comprehensive enough to support that. There is a good case to be made for designing an assessment model to use the data that are available, rather than constructing new data to fit in with a preferred model, and in some cases (the SNE assessment in particular) this was not necessarily done.

I was not clear on how (or whether) we could review the fallback models (ToR 8), given that the stepwise model development process clearly demonstrated that they are less appropriate than the final proposed model. The Panel did not suggest that any of the proposed final models be rejected outright, so the issue perhaps does not arise, but any proposal to use a "less good" version of the full model as a backup is always likely to lead to this difficulty.

I have mixed views on the recreational LPUE index used in the SNE assessment, but part of that concern may stem from a strong driver in ICES assessments not to use LPUE indices if they can be avoided. However, it is clear that the WG is keen to revisit this index, and it is very unlikely to make any difference currently to the management outcomes. Considering issues more widely, I would like to see future work on potential multistock models, further understanding on WHAM model convergence (or the persistent lack thereof), and the potential use of alternative stock-recruitment models in WHAM (a changepoint model would have been useful for these stocks).

Finally, I wonder if the review meeting might be better organised around detailed discussions about assessment decisions. The reviewers had all read the submitted material in depth and so had a good idea of what the key issues would be, and I'm not convinced about the value of the assessor going through the process and material in detail, given the lack of time overall. On the other hand, the amount of material to cover in this meeting was much greater than in my previous CIE reviews, so perhaps this is not a widespread problem.

In conclusion, I am content to recommend the final proposed models for all four stocks for management purposes, given certain caveats about stock-specific work still to be completed (see detailed comments below). The SNE assessment is the weakest, and would benefit most from improvements, but it is worth retaining to improve both management and future assessment development.

ToR 1: Ecosystem and climate influences

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

I concluded that this ToR had been addressed in full.

Reading the submitted material, it was clear that extensive analyses had been conducted to determine the likely ecosystem and climate influences on the four stocks. As became clear during the meeting, the cod fisheries in the north-eastern US area have diminished over time, and are now probably less influential on stock dynamics than ecosystem and environmental drivers. It was therefore very important to understand these non-fishery drivers, and the WG had made great progress in doing so. I should also note that I very much appreciated the changepoint analysis that was conducted on the environmental drivers, recruitment and the R/SSB ratio – this was very clearly presented with some helpful summary plots that I had not seen before.

The area has seen a significant warming trend in recent years, and there are strong indications that this has affected cod stock dynamics: both negatively, through reduced recruitment, and positively due to increased growth rates (up to a threshold of 12ºC, above which cod will struggle to survive). Changepoint analysis of estimated recruitment suggested the possibility of a regime shift in or around 2010, but subsequent analysis of potential environmental covariates failed to identify an ecosystem driver that was significant and consistent across the area. Furthermore, analyses of recruits per spawner (R/SSB) suggested that the rate of recruitment had varied but was not significantly

diminished in recent years. These two analyses implied that reduced recruitment may just be a function of reduced SSB, which seems feasible.

In terms of predation, it is hypothesised that grey seals may be key predators of cod, but there is a lack of stomach contents sampling to enable this to be verified. More generally, there does not yet exist the kind of multispecies population model that we have in the North and Celtic Seas, and which would facilitate the development of age- and time-varying natural mortality estimates. No plans to develop such a model were mentioned during the meeting, but this could be a fruitful future direction to take.

The ecosystem identifiers which could be used to determine status are: sea surface temperature (affecting recruitment), bottom temperature (growth), Gulf Stream index (recruitment), marine heatwaves (natural mortality) and zooplankton (recruitment). These were used in an exploratory GAM analysis, although there is always a problem of spurious correlation with these kinds of analyses, plus the propensity of relationships to break down at the key moments, as well as the difficulty in predicting many of the potential environmental drivers. The text doesn't go much further than stating that these need to be considered.

Regarding questions raised during the review, I noted that only environmental influences on recruitment were taken forward to further analysis, although growth (condition) and distribution were also tested in WP3 and in the relevant chapter in the main WG report. I queried this during the meeting: these were not taken forward because there is no appropriate way to do so in the proposed WHAM assessments. This is understandable, but it would be helpful to consider this issue further in any future assessment developments – the impact of climate change on distribution would be particularly relevant to a more spatially-defined assessment model, and I understand that WHAM can be configured in this way.

The WG report text notes that warming could influence spawning locations, but it would be interesting to know whether this has been observed in practice. It would also be helpful to know whether the predictions of Drinkwater (2005) been borne out. That is quite an old paper, and it should be clear by now whether what Drinkwater forecasted has happened. The text also notes that the GAM analyses on which the conclusions are based are limited to years up to and including 2019. They may therefore already be out of date, given how fast climate change is operating.

I commented that climate change theory would suggest declines for gadoid stocks in the North Sea, but some of these (haddock and whiting) are currently increasing rapidly towards historical maxima, and others (cod and saithe) are improving. The influence of climate change is therefore difficult to predict for specific fish stocks.

Finally, I felt some of the literature review and stakeholder input sections lacked synthesis – these often feature observations from particular times or places that can be hard to use in wider models. ICES is also working hard to be able to incorporate stakeholder information into advice, which is not straightforward.

I have included below some more editorial notes and questions from the main WG report, WPs and presentations:

- In Figure 1.1, the orange lines are not mentioned in the legend or caption (presumably they are smoothers?)
- The recommendations at the end of this section in the main WG report seem a little out of the blue. For example, the influence of *Calanus* on distribution is raised as a key factor in the summary, but this is not mentioned in the main text. It is covered in WP3, though, and this is an example of the main text not really standing alone.

ToR 2: Fishery data

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

I found that this ToR has almost been addressed in full, with the exception of the spatial distribution of landings (or catch) for which maps were not provided.

The section in the main WG report on fishery data was extensive, at 110 pages, and the data collection and collation of the commercial and recreational landings and discards data for the four stocks was well described. I have some issues with the way in which this was done in some cases, which I cover below, but the approach was covered in full in the report, presentations, and relevant WPs.

I would have found catch distribution maps (by year if possible) very helpful to follow some of the discussion, and to answer the ToR request for "spatial distribution". There is a catch distribution map in WP6 (Figure 2), but it is from VTR only and covers aggregated 5-year blocks, and also is not included in the main report. I also think that a short summary of the findings of the Atlantic Cod Stock Structure Working Group (ACSSWG) would have been helpful to understand the stock definitions, although I realise that we were not being asked to provide a review of that process. A short summary in the WG report of the stakeholder inputs would also have been useful.

While reading the WG report, my initial view had been that the discard mortality rates are generally too low in these assessments. In the 1990s, the Aberdeen Marine Laboratory collaborated with IMR in Bergen on the Survival project, which showed that mortality often occurs sometime after escape, due to injuries received – and this was not even allowing for increased predation risk. However, the acoustic telemetry studies mentioned in WP5 give me more confidence: furthermore, the Survival project used trawls at depth, not hooks in relatively shallow water, and so was perhaps less relevant. During the meeting, it was highlighted that the discard mortality estimates were based on acoustic telemetry at typical depths, with fish being tracked for 30 days to determine post-discard survival. In this work, injury including barotrauma was accounted for, and it was found that hook lacerations were the main impact. This is an interesting and very valid approach to discard mortality estimation, and I am happy to accept the estimates as presented.

The Panel meeting was presented with a summary of the methods used to assign catch to the new stocks from miscellaneous areas. To me, the shapes and sizes of these areas are a strong indication that there is probably mixing between the nominal stocks: for example, it seems unlikely that the area at the bottom of WGoM is entirely WGoM stock fish. I will return to this below when discussing multistock models.

The EGoM stock has quite limited age-composition sampling, with only around half the years covered, and a key fleet for the SNE stock (recreational landings) has no age-sampling at all. These gaps were addressed either by allowing the WHAM state-space model to smooth over them, or by using available catch and survey ALKs to generate numbers at age. An analysis to explore the impact of missing market categories showed that to ignore these categories completely would lead to bias, so in that sense it was preferable to allow the model to generate age compositions (or import them from other catch or survey sources), but it raises the question of whether an age-based assessment is fully warranted when age sampling is sparse. Index-based assessments do not have a great track record in this area, which is behind the obvious strong driver for a quantitative age-based assessment, but this is not without risks when data are limited. I will comment more on this issue when discussing ToR 4 (assessments).

There is a decreasing trend in weights-at-age for older ages, and I questioned whether this could be linked with earlier maturation. The mechanism remains unclear, but it is a widespread feature across many stocks and may be associated with temperature. Maturation age is also declining with time, however, which is similar with what we see in the North Sea.

During the meeting, my fellow panel member Noel Cadigan generated SPAY plots (Standardized Proportions by Age across Years) following fishery age compositions through time. These suggest that the GB and WGoM fishery data follow cohorts quite well (suggesting decent sampling), while this is not the case for the EGoM and SNE fishery data. There is also a possible issue about a chunk of positive residuals around 2005 for WGoM – this would look like a year-effect if this were a survey, and is more difficult to explain for catch. I will return to this issue when discussing ToRs 3 (surveys) and 4 (assessments).

The threshold for inclusion of sampling blocks (> 10 lengths, rather than 100 lengths as before) seems very low to me, and it wasn't clear whether this would have implications for uncertainty in catch estimation. In Scotland we generally use an annual discard rate estimate, as any finer scale is not supported by data.

It is unfortunate that it has not been possible to include cod bycatch in the lobster fishery in the EGoM assessment – although the amount of cod involved is small, it would be a significant proportion of the overall catch, and this should certainly be reconsidered for future work.

I have included below some more editorial notes and questions from the main WG report, WPs and presentations:

- It's not clear to me how weight-at-length models are used to convert catch weight to numbers. I presume this is through length distributions, but this was not stated anywhere in the report or presentations.
- It was explained during the meeting that the noted market categories are not fixed length spans, but vary from market to market depending on how sellers define them. A short note to this effect in the report would have helped my understanding of the process.
- Main WG report, Figure 2.1: a colour key to names of areas would be helpful here.
- Main WG report, Figure 2.4: the syntax here is odd - *Ratio of landings in the 510s that occurred in EGoM vs WGoM.* Does this mean the ratio of landings that were assigned to EGoM vs. WGoM? The landings can only really be said to have "occurred" in 510.
- Main WG report, Table 2.4: for ease of reading, this table could have been reformatted to fit on one page (this is a minor note).
- Main WG report, Figure 2.12 etc – a scale or key would be useful for the bubble plots. I would also suggest running a loess smoother through the weight-at-age data, to make longer-term trends easier to see.
- Discards are hindcasted using fixed discard rates for earlier years. I don't think this text in the main WG report is correct though: "multiplied by the annual total catch (Kall) in years 1964-1988". Kall wouldn't be known for these years, so is it not kall (landings) that is multiplied by the discard ratio to produce Kall?
- Main WG report, Figure 2.17 (for example): comparing these plots would be easier if the plots for all discards and dead discards were on the same scale.
- Main WG report, Table 2.8: again, this table could be reformatted to fit on one page (or two at most).
- Main WG report, Figure 2.25: the sizes of the bubbles are mentioned in the caption, but there is still not a scale or legend.
- The text notes that the EGoM 1989 recreational catch estimate is probably not spurious, but is there an explanation for it – why would it be 50-100 times greater than neighbouring years?

- The WG report text states (p. 101): "in 2021, >60% of harvested fish in WGoM were estimated to be exactly 59 cm." The expanded (stratified) estimates tend to have unrealistic spikes (Figure 2.27, slide 42) which aren't seen with the raw (unstratified) data, and the 59-cm problem appears to be an artefact of this stratification issue. It wasn't made clear why this might be, however.
- Main WG report, Figure 2.32 – this needs some explanation. I am guessing the y-axis labels refer to the blue lines, and the survey and commercial landings are represented by green and red circles? This is quite confusing to try and disentangle.
- Main WG report, Figure 2.33 – it's maybe an optical illusion, but it does look like the fitted curves tend to miss the points at higher ages in some years.
- Main WG report, Figure 2.35 (for example) – this would benefit from a scale or key for the bubble plot.

ToR 3: Survey data

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

I found that this ToR had been addressed in full.

With the decision to move to a four-stock structure made, it was clear to the WG that more data would need to be considered to characterise stock dynamics at a finer spatial scale than previously. Consequently, the WG considered a total of 11 surveys, several of which were split seasonally, and determined whether each could be viewed as sufficiently representative and reliable.

The main WG report, WPs and presentations took a broadly consistent approach to the presentation of this material, which was largely very comprehensive. However, while the bivariate scatterplots from WP8 were very useful, I would have liked to have seen these presented in the main WG report with correlation coefficients (along with catch curves and other survey diagnostics). From the report, it is hard to tell whether the surveys are tracking year-class strength very well – the existing plots would suggest not. Lisa Kerr from the WG was very obliging, however, in generating relevant plots during the meeting for both catch and survey data (as did both Noel Cadigan and myself, using a variety of tools). In summary:

- WGoM: the commercial catch tracks cohort strength well, while the recreational fleet loses cohort signals at lags of more than 1 year. Catch curves for both are well-specified and don't show unwanted hooking at younger ages. The NEFSC trawl spring Albatross survey does well at a lag of one to two years, but shows significantly lower selectivity at younger ages, while the equivalent Bigelow survey does poorly at tracking cohorts. The pattern is similar for the corresponding fall surveys. The spring and fall MENH surveys do not track cohorts well, while the IBS survey does better (although is very limited in the years covered). The BLLS spring and fall surveys are both poor in cohort tracking.
- GB: the commercial fleet data, the DFO survey and the NMFS fall survey all track cohorts well in this area, although the surveys do start to lose signal at greater lags.
- SNE: the commercial fleet and the NEFSC spring survey are both very poor at tracking cohort strength in this area, although with the very low numbers of fish caught this may not be surprising. The recreational catch data are much better in this regard, although we need to remember that the age compositions for these data were not from the recreational fleet itself, but were inferred from commercial and survey data.

- EGoM: the commercial fleet tracks cohort strength reasonably well, but only at short lags, and there are in any case many gaps in the age composition data. Very little can be inferred from the spring and fall MENH or NEFSC surveys, or from the sentinel jig survey.

These were extremely helpful analyses for which I thank Lisa and the WG members. However, they also indicate significant problems with the utility of survey data for determining cod population dynamics in these areas. I will return to this issue when discussing the assessments below.

Turning to some further survey issues, I noted that the "frequency" y-axes in many of the length distribution plots needed to be clarified (e.g., ToR 3 report, Figure 3.39), as some of them didn't make sense to me. Again, Lisa Kerr from the WG was very responsive in following up on this point, and it transpired that the relevant axis captions were indeed incorrect and were updated accordingly. Another relatively minor issue was around the bubble plots of abundance or biomass at age from the surveys: these were often presented without scale bubbles, which makes it harder for the reader to evaluate the likely impact of any unwanted patterns (an example is Figure 3.26).

Many ICES assessments now make use of spatio-temporal GAM (generalised additive model) approaches to generate survey indices, allowing different surveys covering different locations, gear types and ages to be combined within a season into a single holistic index. With eleven surveys available across the four stocks, this kind of synthesis could be extremely useful, and was considered by the WG through WP 10 which describes Vector Autoregressive Spatio-Temporal (VAST) models that are similar to GAMs, although perhaps more complicated. Unfortunately, this work could not be completed in time for the research track finalisation, and with the Panel I have provided a recommendation to progress this work when time allows.

During the Panel meeting, we asked whether age-length keys (ALKs) are shared between different data sources. This is indeed the case, with ages for several datasets (landings, discards and/or surveys) being inferred by the application of an ALK from another source to length data. This has the potential to introduce both uncertainty and bias, and to a certain extent reflects the strong driver from the WG to develop an age-based assessment for all four of the cod stocks, using both catch and survey data in each case. There is an equally strong argument that assessments should only try and use the data that are available; so, a survey that only measures lengths perhaps shouldn't be used in an age-based assessment. Borrowing data in this way is a widely used approach in fisheries assessments, particularly in areas that may be data-sparse, but it needs to be used with care.

We also noted that the landings-per-unit-effort (LPUE) indices that were developed for commercial and recreational fisheries in the SNE area used only positive observations – that is, data from the hauls which caught cod. This is likely to lead to an estimate which is biased upwards. The WG provided well-reasoned arguments to support this inclusion, but agreed that an index should be developed that accounted for zero observations. This appears to be not all that straightforward and couldn't be completed in time for the end of the Panel meeting, but we have recommended strongly that this be addressed before the next management track process.

I have included below some more editorial notes and questions from the main WG report, WPs and presentations. Time was lacking during the Panel meeting to raise these, but they might help for any future management or research track process, or for minor edits to the current WG report. Figure and table numbers refer to the WG report as submitted to the Panel for review in July 2023.

- NEFSC Bottom Trawl Survey:
    - In assigning strata to stock units, was there ever a problem with a stratum belonging equally to two stock units, and how was this resolved?
    - If the survey was curtailed or cancelled in 2020, why are there 2020 indices in Figures 3.5 – 3.8?

    - Figure 3.9 is missing data for 2021 and 2022 (as is Figure 3.13, but it does state that data are missing after 2019, although no reason appears to be given).
    - In Figure 3.17, some of the SNE data look a bit odd – there over 30 fish in 1994, all of the same length.
    - Figure 3.18 – the caption says that data are only available from 1963-2021, but the plots include 2022.
- DFO Trawl Survey:
    - It is not clear how the "estimated abundance of cod" (Figure 3.24) is generated – I assume through a swept area calculation, but it would be helpful if this were clarified. I would also ask why confidence intervals were not provided?
    - Figure 3.28 – if the shaded areas are 95% confidence intervals as I assume, then this should be stated.
    - Figure 3.29 – I don't think the length-weight equation is given anywhere.
- Maine-New Hampshire Inshore Trawl Survey:
    - It might be helpful to see a map to check the noted "misalignment" between stock areas and survey index areas.
    - Figure 3.31 – do the white areas signify unfishable ground?
    - Figure 3.32 – how are the confidence intervals generated – bootstrapping? The CVs of the larger estimates look to be very high compared with the smaller estimates, but that may be an optical illusion.
    - Figures 3.34, 3.35 – the time series stops in 2020 – are there no data for 2021-22? The previous time-series plots would suggest there are.
- Eastern Gulf of Maine Sentinel Survey:
    - No further comments.
- MDMF Inshore Bottom Trawl Survey:
    - Figure 3.46 – some of the plots for this survey show very sporadic catches, which suggests that it would be difficult to follow stock trends and year-class strength.
- Cooperative GoM Bottom Longline Survey:
    - Figure 3.53 – presumably this is across all years?
- MADMF Industry-based Survey:
    - Figure 3.64 – I found this plot interesting as a preview of what inference could be possible with this survey, but I think it needs further explanation to stand in the report.
- UMASS SMAST survey:
    - I would have been interested to know whether the image analysis for this survey was automated, or done by human viewers – in Scotland we use TV surveys for several species, and it always helpful to know how they are being used elsewhere.
- ASMFC Northern Shrimp Trawl Survey:
    - It wasn't clear whether this was a research-vessel survey, or whether it was carried out on commercial vessels.

- Figure 3.72 – the data contrast is very difficult to see in this Figure, because the plots are long and thin. It might be better to present these side by side.

- Rhode Island (RIDEM) Coastal Trawl Survey:
  - Again, it would be helpful to know whether this survey was conducted on a research vessel or a commercial fishing boat.
- University of Rhode Island GSO Fish Trawl Survey:
  - It wasn't clear how this survey was conducted either: was it with a net towed from two fixed stations on land, or was it done with a small inshore vessel?
- WP 9 compares indices by stock area and season. A short summary of conclusions would have been useful in the main WG report.

Finally, following my request for further survey diagnostics to understand cohort tracking ability, I downloaded the relevant survey data from the NEFSC GitHub repository and conducted SURBAR runs (a survey-based assessment method commonly used in ICES for exploratory data analysis, see Needle 2015). SURBAR is useful for generating bivariate scatterplots and catch curves, but it will also produce estimates of total mortality, along with relative SSB, TSB and recruitment, based on multiple surveys within a specific region. Time did not permit a full exploration of parameter settings for these SURBAR runs, and I only completed a draft analysis for the WGoM stock, but I have included the stock summaries for WGoM below as comparisons with the final proposed assessments can often be instructive (Figures 1-6).

Reference: Needle C. L. (2015). "Using self-testing to validate the SURBAR survey-based assessment model". In: *Fisheries Research* 171. pp. 78–86. DOI: 10.1016/j.fishres.2015.03.001.

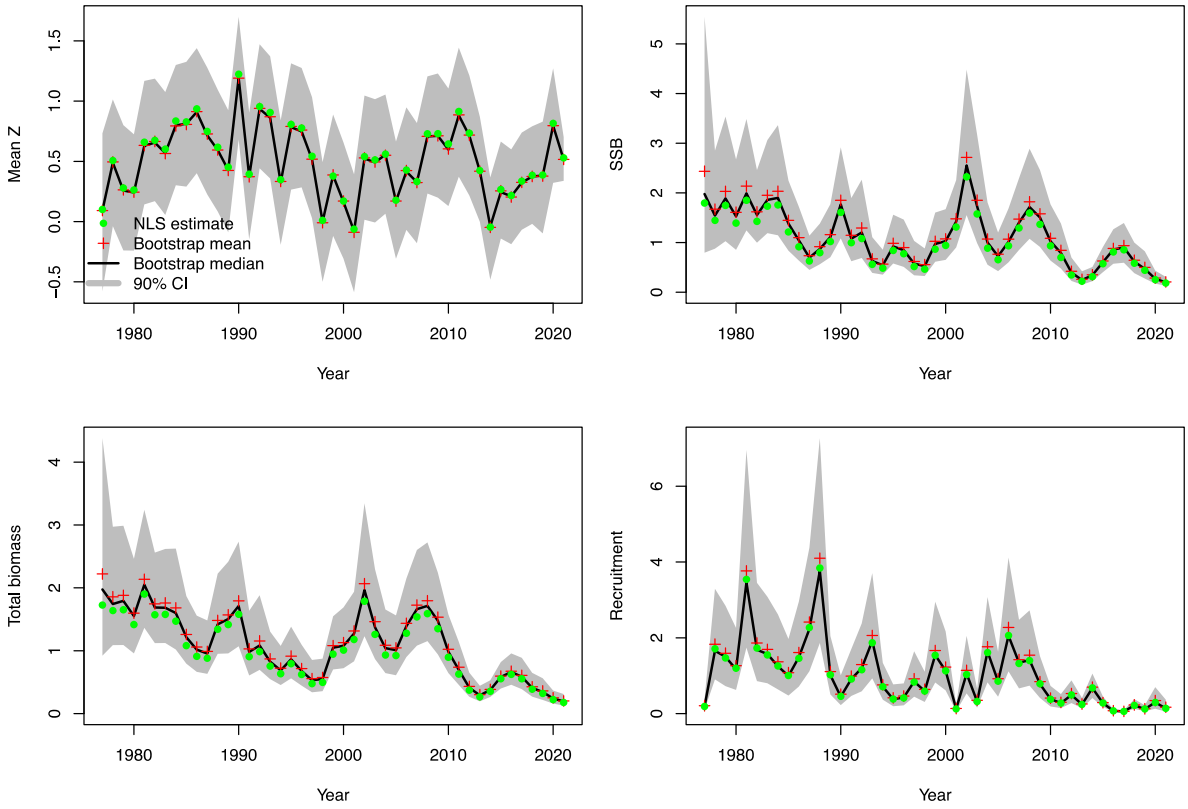


Figure 1: Output from a draft SURBAR run using the NEFSC (Q1 and Q3) and BLSS (Q1 and Q3) surveys for WGoM (ages 1-6 only). Stock summaries.

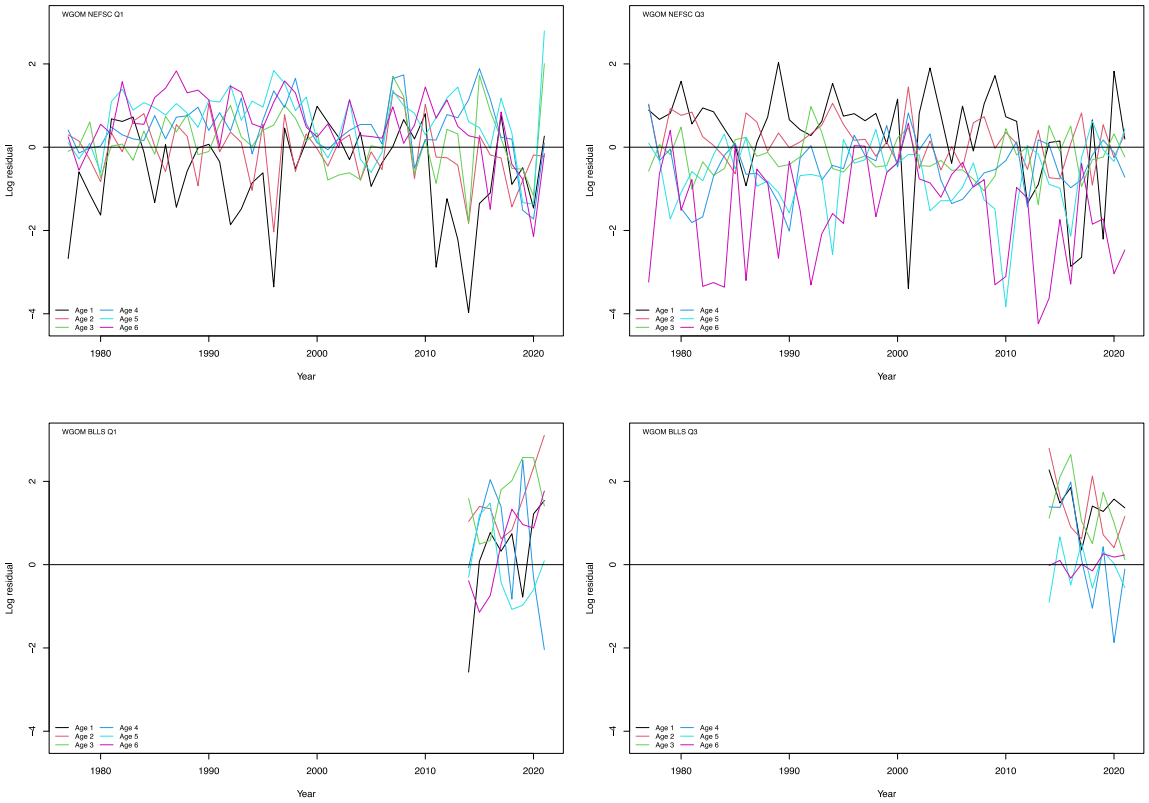


Figure 2: Output from a draft SURBAR run using the NEFSC (Q1 and Q3) and BLSS (Q1 and Q3) surveys for WGoM (ages 1-6 only). Index residuals.

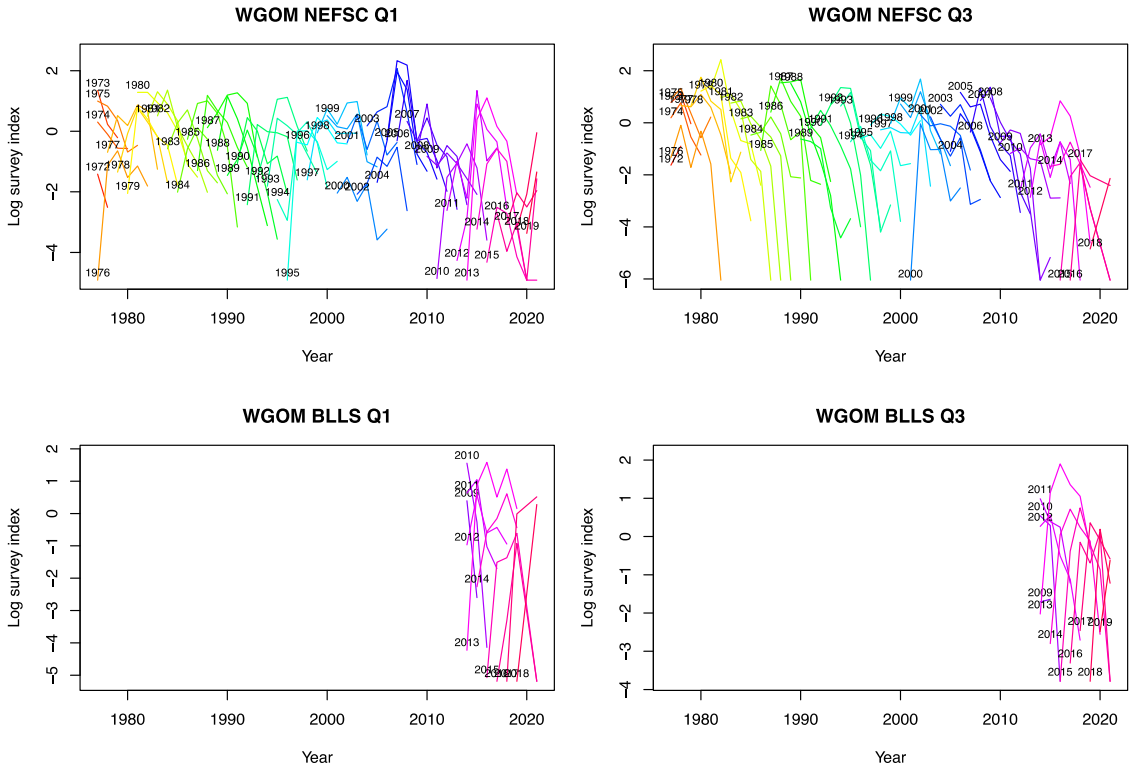


Figure 3: Output from a draft SURBAR run using the NEFSC (Q1 and Q3) and BLSS (Q1 and Q3) surveys for WGoM (ages 1-6 only). Log catch curves.

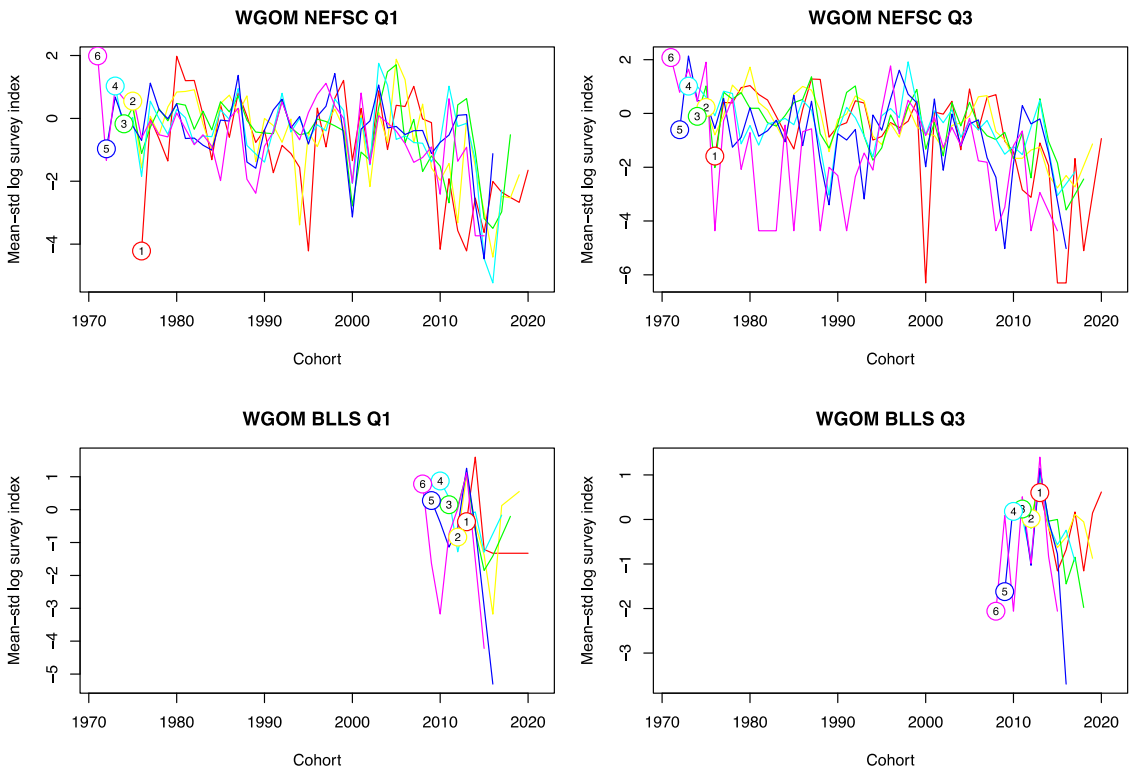
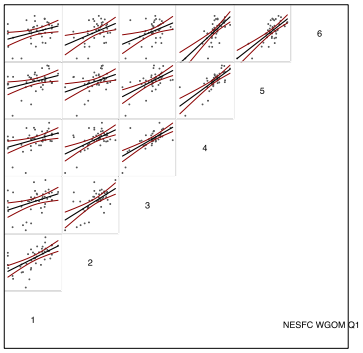
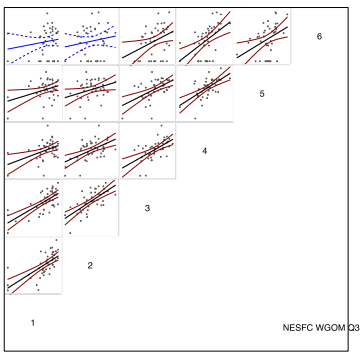


Figure 4: Output from a draft SURBAR run using the NEFSC (Q1 and Q3) and BLSS (Q1 and Q3) surveys for WGoM (ages 1-6 only). Mean-standardised log survey indices plotted by age class and cohort.





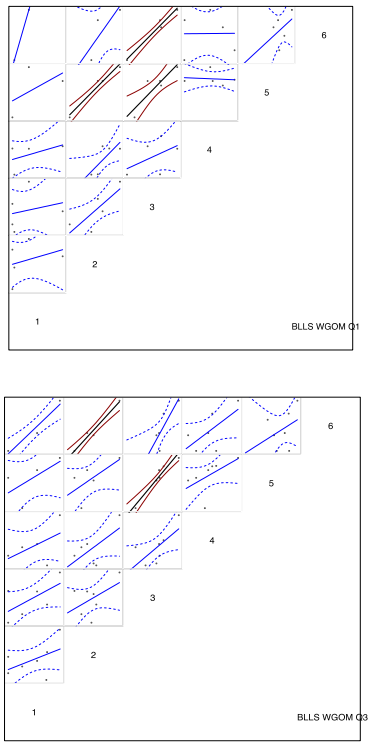


Figure 5:Output from a draft SURBAR run using the NEFSC (Q1 and Q3) and BLSS (Q1 and Q3) surveys for WGoM (ages 1-6 only). Bivariate scatterplots comparing index values for ages x and y, along with fitted regression lines. Significant regressions are highlighted with black and red lines.

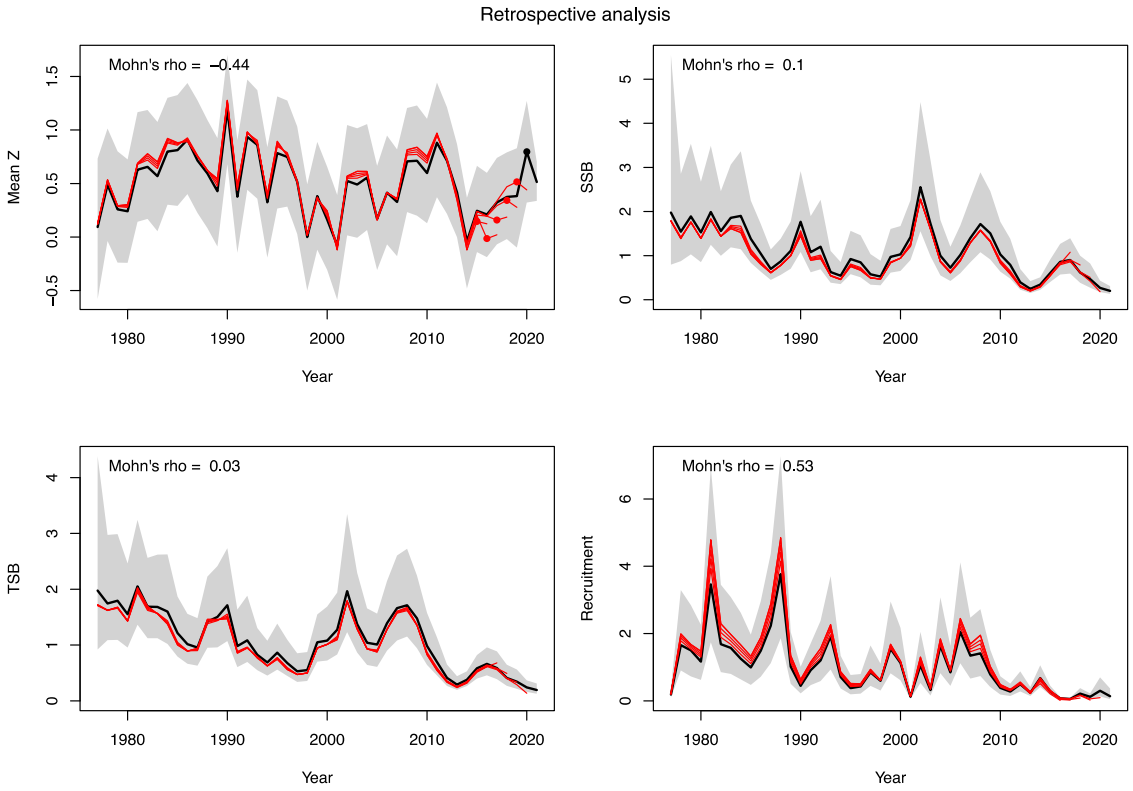


Figure 6: Output from a draft SURBAR run using the NEFSC (Q1 and Q3) and BLSS (Q1 and Q3) surveys for WGoM (ages 1-6 only). Retrospective analysis. The final run is shown is black with grey confidence bounds, while five retrospective peels are shown in red. Values of Mohn's rho are shown for each subplot.

ToR 4: Assessment

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

I concurred with the Panel in finding that this ToR had been met in full for the WGoM and GB stocks, while it had nearly been met in full for the EGoM and SNE stocks. For the latter two stocks, along with the Panel I indicated several further analyses or data sources that need to be considered before the next management track processes.

Considering the assessments as a whole, I found that the compiled report was not always sufficient to be able to evaluate model runs and fits appropriately. The WPs and presentations mostly covered the gaps, and there was a lot of material to cover, but it perhaps should have been possible to be comprehensive within a 450-page report for 4 stocks. Data and code for all assessments were made available to reviewers through a GitHub repository, although time did not permit a full exploration of these before or during the Panel review meeting. On the other hand, I did find that the assessment and statistical modelling knowledge of the analysts presenting the material was at a very high level, higher than perhaps would be the case for some ICES stocks, and the WG team is to be commended on very full investigations and clear presentations of some difficult material.

Many of the exploratory runs seem to have been discarded due to a lack of convergence in the WHAM model runs (for example, all time-varying M runs for the WGoM stock). This indicated to me that there was possibly a problem with WHAM, rather than with the settings attempted, and it did seem that some reasonable representations of reality may have been missed because the relevant WHAM runs wouldn't converge. It transpired during the meeting that many of these runs failed to converge because parameters hit their specified bounds. My experience with state-space models (mostly TSA, but also SAM) is that when this happens, the usual approach is to adjust the parameter bounds until they are no longer hit – if the parameters become unrealistic through this process, then removing the model is appropriate, but otherwise a converged model becomes much more likely. I would encourage this more iterative approach to be taken in future research-track assessments.

Although the WHAM models (particularly for WGoM) were structured so as to allow for possible multistock modelling, this was not taken further during this research track. There is a lack of sufficient genetic data to segregate winter and spring spawners in the WGoM area, so treating these as one stock seems reasonable for now. But it also struck me that the boundaries between the stocks were quite uncertain in places, and in the future, I would like to see some further consideration of a multistock WHAM model to assess the four stocks together, of the type currently proposed for ICES Northern Shelf cod. This can lead to difficulties and requires good genetic or morphometric sampling to enable estimation of both stock fidelity and mixing, but if successful it might be a more realistic assessment of the extant biological situation.

As a minor issue, for the presentation of the test model results, it might have been helpful to present (for example) residual plots for these alongside those for the base model, space permitting. It is hard for the reader to flick through a PDF and check that residuals for an altered model really are "better".

The stepwise experimental design that was taken was well-explained, and was probably all that was possible in the time available. However, I am concerned that it may have missed combinations of changes that could have given better outcomes. For example, if time-varying *M* was tested as the last addition, rather than earlier in the stepwise process, it might have improved the model at that point. Obviously the WG were unable to do all combinations, but approaches involving randomised

experimental design are well-attested in the literature and could have been considered. During the Panel meeting, this was recognised by the analysts as a potential issue. There were a few times when they stepped back and reconsidered inclusion of a specific new element, but the stepwise path was largely pre-chosen. It would have been interesting to see a stepwise comparison of model results as changes were made, rather than just diagnostics – as long as there was no implication that configurations were chosen to suit desired outcomes, of course.

The only recruitment model tried in these models was Beverton-Holt, which for all but one of the stocks became linear when fitted (and therefore was of little use for reference point estimation and forecasts). The only other recruitment model currently available in WHAM is Ricker, which was not considered further as it would suffer from the same problem. To me, a changepoint (or segmented regression) model might be useful here, as it would improve historical recruitment modelling by accounting for the observed linear relationship with SSB, while not allowing unlimited increases in recruitment in forecasts (or for reference point estimation) – the estimate changepoint would probably be at the observed SSB maximum at most. I recommend that a changepoint model (or similar formulation) be considered for inclusion in future versions of WHAM.

The models for all four stocks seem to fit catch data exactly, more or less. This was most likely due to the assumed CVs on catch data being set to a very low level (0.05, lower than the observed CV), along with potentially a lack of cohort tracking (and other noise) in the surveys. We were told that model stability suffered if more deviation was allowed, but this begs the question whether the surveys make any difference to the assessment: and also on what basis the catch CV was set to 0.05 The WG did not conduct leave-one-out assessments to ascertain the influence of the surveys, but my perception is that the assessments would be similar if the surveys were not included. My suggestions here for future assessments would be to always include leave-one-out runs as a standard output, and also to consider using observed (or at least higher) CVs for catch data, as otherwise the inclusion of the survey data becomes relatively uninformative. There is a clear analogy here with the "shrinkage" feature of the XSA model that was previously widely used in ICES, and which has since been largely discarded.

Natural mortality estimates remain time- and age-invariant for all stocks. WP 14 reported that some potential higher values of M were removed as being "unreasonable". To me, 0.2 – 0.3 seems to be on the low side for the lifetime natural mortality rate of a cod: in ICES assessments, cod M starts at 0.8 before falling to 0.2, based on multispecies models, and to me this seems more reasonable.

Maturity-at-age had previously been assumed to be constant through time. In Europe, maturation rates have changed significantly as early growth has increased with rising temperatures. If this pertains in the north-eastern US, a constant assumption is likely to introduce bias.

I questioned whether there was a cut-off threshold for Mohn's rho in the retrospective analysis for these assessments. In ICES advice, 0.2 would be considered the upper limit for F or SSB retrospectives, but there doesn't seem to be a fixed limit here – conclusions on retrospective bias are more qualitative.

For all stocks, a consistent model selection procedure was used, including convergence, residuals, goodness of fit, retrospective bias, forecasting precision, estimation performance (self-testing), and model stability. This was a good approach overall, and I found the explanation of it to be clear and helpful. However, I comment elsewhere on the potential issue with convergence in WHAM, and the quite rigid stepwise approach to model development.

Finally, I appreciated the presentation from Tim Miller on the modelling basis of the WHAM model and the implementations used for these stocks. The differences with the SAM model used in ICES (with which I am more familiar) were highlighted, as were issues such as parameters going towards their bounds during estimation. The additional features of the WHAM model over SAM, such as explicit modelling of environmental covariates, are clearly quite specific to the assessment requirements in the north-eastern US. It would be interesting at some point to investigate

comparative runs of WHAM and SAM applied to the same model, in order to understand more about the characteristics of each. I noted that there is also an ongoing research track process for the WHAM model itself, which will be helpful.

Western Gulf of Maine (WGoM) stock

I concluded that the ToR had been met in full for this stock.

There was discussion about the use of sampled data to develop appropriate weight-length relationships: however, it was not made clear what criteria were being used determine how many such pairs would be considered "enough". Two approaches were presented comparing weight-length models, but it seemed to me that the WG could also just have fitted a growth model to data from each year, to see how the parameters of these have changed through time. This estimate was really measuring fatness or condition, though, and this section isn't actually about "growth" in the sense of fish size changing with time.

I liked the plot in Figure 4.1, which summarised the age at 50% maturity according to the fitted state-space models. However, for WGoM the a50 estimates seem to be settling down a few years before the projection period. It wasn't clear whether this is a function of backwards iteration in the state-space model, or just a coincidence.

Natural mortality for WGoM was previously fixed at 0.2 for all ages, years and stocks, or with an "M-ramp" model with a linear increase from 0.2 to 0.4 between 1989 and 2002. Life-history based M estimation methods were applied, in the absence of a multispecies VPA for this area, and the median of the different methods was used as a time-invariant estimate. WHAM can apparently estimate natural mortality, so these estimates were used as starting values only, but I would have been interested to hear more about how M can be estimated in a single-stock model without confounding with F.

I was not fully convinced about the value of using a multi-stock WHAM formulation which gives the same results as a single-stock WHAM, when only modelling one stock in WGOM, and when the two-stock approach may or may not be implemented at some future date. This seems like an unnecessary duplication to me, and while I realise it doesn't make any difference, it seems to me that there is still an increased possibility of confusion and error.

The assessment used a total of 8 survey indices, including some that are inshore only or otherwise restricted in spatial extent. As I noted above, methods involving integrated spatio-temporal indices are in development, but in the absence of these there seems to be a danger of the assessment being driven by unrepresentative indices.

WP 17 was very useful to explain and explore some more of the detail behind the assessment, which wasn't possible to include in the main WG report. Some minor points within that WP: in Figure 4, the scale of the bubbles tends to obscure some of the detail; in Figure 9, it is hard to see which indices these selectivity plots are for; and while Figure 11 is useful, I would also like to see the retro plots provided as absolute as well as relative estimates.

The Dirichlet distribution is based on a beta distribution rather than a normal distribution, and so will have strict bounds (unlike the previous multinomial distribution). Was that the driver for the change to Dirichlet in this assessment – to ensure bounded likelihoods? Neither the WG report nor WP 17 are clear on this.

The final proposed version of this assessment included estimated random effects on catch numbers-at-age (NAA). This led to a better model fit, but there is always a concern about what exactly the random effects are trying to model – it could be missing catch, environmental changes, time-varying M, migration to neighbouring areas, etc. Figure 4.WGOM49 does indicate that NAA random effect estimates for age 1 definitely seem to be trying to account for something, but we cannot tell what.

Overall, and notwithstanding my concerns about missing potentially beneficial combinations of parameter settings, the WG did an excellent job of applying the stepwise model fitting process rigorously, and have produced an assessment with much better diagnostics than was previously the case. The model fit to the survey data is not always particularly close, but I feel this is an inevitable consequence of setting a very low CV on the catch data, which means that surveys are ignored to a certain extent. I should also note that the material provided in the WG report was far more extensive for this stock than for the others, which gave the reviewers more to read but also facilitated a more comprehensive review. Unfortunately, the final proposed assessment looks quite bleak, with historical low SSB and no evidence of recruitment recovery.

Georges Bank (GB) stock:

I concluded that the ToR had been met in full for this stock, although I must also say that I found the GB analysis to be less well-structured and more difficult to follow than the analyses for the other stocks. The information seemed to all be there, once I hunted for it, but the run numbering system was a little confusing and the treatment didn't really follow the same logical flow. For example, run 19A is the final proposal for an accepted run, but run 17A is denoted the "revised preferred run", and while reading this section it was sometimes unclear which was which.

The WG noted that splitting the US and Canadian commercial catches (there are only minor recreational catches for this stock) led to worse retrospective bias, and so the split was not retained. However, there weren't any hypotheses put forward as to why this might be, and I think this would be a useful approach to reconsider in future research track work.

The most recent analytical model was rejected in previous management track assessments, leading to advice based on smoothed survey indices. These have lost favour in this area following simulation testing that suggested they were unreliable, so there was a strong driver to generate a new analytical assessment "if possible".

During the review meeting, it was noted that the distributional assumptions made for the age compositions for this stock appeared rather *ad hoc*, and as a Panel we expressed concerns that these might be unduly influential on the outcomes. However, the WG analyst produced additional runs using different assumptions, and demonstrated that these had very little impact.

The remaining model choices seem to be well-presented and justified. The main issues for me were those raised at the start of this section relating to all the assessments (model convergence, recruitment models, etc), along with retrospective bias which remains stubbornly high for this stock – and which probably would lead to a Category 3 (data limited) assessment and advice in the ICES context. I was interested to learn that even though high, the bias does not appear to be high enough to warrant retrospective rescaling in the provision of advice, but I would be concerned that this level of bias might render the advice unreliable.

Below are some specific comments regarding figures in the collated WG report:

- Figure 4.GB1 – does the caption mean that the purple line has one parameter estimated to cover selectivity at ages 7-9+?
- Figure 4.GB2 – as with WGoM, the catch data seem to be fit exactly – is this just an optical illusion, or is that catch really that strong a driver in the assessment? The residuals in Figure 4.GB3 still look quite large.
- Figure 4.GB4 – does this Figure indicate that naïve prediction is actually better for all years except the first?
- Figure 4.GB6 – again, I would like to see retrospective runs plotted as absolute values, alongside these relative plots.

- Figure 4.GB7 – what age range is used for "fully-selected F"? I think this summary should also include recruitment time-series estimates.
- Figure 4.GB8 – this would be more informative if shown not on a log scale.

Southern New England (SNE) stock:

I concluded that the ToR has not quite been met in full for this stock: there were some outstanding issues that would need to be resolved before the next management track assessment.

The SNE stock was previously considered to be the western part of the GB stock, and this year was the first in which a separate SNE assessment had been attempted. Model development had begun for this stock using the previous ASAP model, along with some exploratory runs in Stock Synthesis (SS), and the decision to change to the WHAM model came quite late in the WG process. This was due in part to diagnostic problems with both earlier models: SS had poor prediction skill, and residual patterns in fits to recreational length data; ASAP had poor convergence rates, and worse retrospective bias. Consistency with other stocks seems to have been the main reason to change to WHAM in particular. The late switch meant that the number of different candidate runs explored for the SNE stock was much less than for other stocks. This was unfortunate, as SNE is probably the most questionable of the four stock assessments covered by this review.

Unusually, the recreational catch is higher for this stock than the commercial catch (and has been since 1981, if not before). Recreational estimates are available from 1981 onwards only, so that fixed the earliest start point of the assessment.

Age sampling is quite limited for this stock, with (for example) ALKs from surveys and the commercial catch being used for the significant recreational catch, and weights-at-age, maturity-at-age and natural mortality all assumed to be time-invariant. I did question if there had ever been a consideration that there might not be enough data to do a full catch-at-age assessment. Even though age sampling has been carried out, for commercial landings, it is worth noting that age composition data are missing for commercial landings for more than half of the years covered (with none since 2014). The WHAM model is able to proceed with several years of missing data for both overall catches and age compositions, and the WG presented analyses supporting the retention of missing years rather than using interpolations, but the question still remains: is it fully appropriate to continue with an age-based assessment? To my mind, were this an ICES assessment, there would be a strong consideration of treating this as a data-limited stock, for which a Category 3 assessment would be more appropriate.

Runs using alternative indices of abundance tried to use the three available age-0 indices in the assessments (SS or WHAM), but fits were poor or conflicting. This isn't surprising to me, as those indices seem to be only sampling a small part of the stock area each, and these runs were not retained. The split survey run divided the NEFSC survey based on the vessel used, but produced much greater retrospective bias (oddly). The commercial LPUE index was not retained, due to concerns about hyperstability (targeting behaviour), but the recreational LPUE was used. Stock-recruit model runs were attempted, and one with Beverton-Holt with a fixed value of steepness *h* from Myers et al (1999) was retained. Figure 4.SNE5 shows a good fit to Beverton-Holt for the stock-recruit data, but it was explained that attempting to estimate *h* directly led to model instability. Nine alternative age-composition likelihood runs were explored (WP 18), but none were better than the standard multinomial – which is in contrast to the WGoM and GB assessments. Random effects were also included on recruitment deviations – random effects on other parameters led to unconverging WHAM models.

The simulation self-tests do not perform well for this stock, with bias in SSB estimates using data generated by the model itself measured at over 15% in some years. This would suggest data or model

misspecification somewhere, but there are no guidelines on how much self-test bias is too much, and this must remain as an issue to be checked in future.

The additional diagnostics provided by the WG during the meeting highlight that the survey and commercial LPUE indices do not track cohort strength well. The recreational LPUE index does do this well, but there are several significant issues with it:

- Ages are not sampled from the recreational fishery, so they are inferred using ALKs drawn from a composite of survey and commercial catch data, which is not generally considered best practice.
- The recreational LPUE is based only on positive fishing events (no zero hauls included), and only from for-hire boats (not private anglers). These are both likely to have an impact on the estimation, particularly the former as it will overestimate the degree to which cod are encountered by such boats.

I also questioned whether the use of a recreational LPUE meant that recreational catch data were somehow "being used twice" – this has often been cited as a reason not to use fishery CPUE or LPUE in ICES assessments. However, it turns out that the relevant age compositions are only fitted once, and subsequently used in both catch and LPUE, so that seems more reasonable.

The Panel recommended strongly that if it is to be used, the recreational LPUE needs to be updated (at the very least) to include zero fishing events, and I fully concur with this. The WG agreed, but were unable to provide an update during the review meeting, and this will need to be done before the next management track process.

Overall, I felt that this was the hardest assessment to agree with. In many respects, the data are insufficient to warrant a full age-based assessment, and I got the impression that the development of such an assessment was almost a pre-determined outcome – no matter what. The catch and survey data problems would in many cases have led to rejection, but I was persuaded to accept the assessment for three main reasons: catch-only advice (a standard fall-back in ICES) would not encourage any management action in this area; for the assessment to develop there needs to be an existing assessment to work on; and the stock status is unlikely to be any different with a different assessment approach.

The assessment suggests that the stock remains in a very poor state, with high *F* and very low SSB. The assessment shows a very high recreational *F* estimate of over 1.0 in particular, meaning that more than 60% of the stock is removed by the recreational fishery every year.

Below are some specific comments regarding figures in the collated WG report:

- Figure 4.SNE3 – the age-0 indices look very noisy, and extremely inconsistent, but they're not being used here.
- Figure 4.SNE6 – the commercial catch is fit almost exactly here.
- Figure 4.SNE7 – the residuals for the index age-comps are poor. The text does note low sample sizes, but are these data likely to be still representative?
- Figure 4.SNE9 – I do prefer this style of retrospective plot, although it would have been even better with the confidence interval around the most recent fit. Retrospective bias is small.

Eastern Gulf of Maine (EGoM) stock:

I concluded that the ToR has not quite been met in full for this stock: there were some outstanding issues that would need to be resolved before the next management track assessment. These relate

principally to bycatch data from the lobster fishery, which is currently not included, but some further exploration of the impact of different survey combinations would also be beneficial.

This is the first year of a separate EGoM assessment, having previously been considered part of the wider Gulf of Maine stock, and (similarly to SNE) it transpired that catch and survey data were actually quite limited for this area. Commercial landings and discards are mostly from trawls, although historically gillnets were also important: however, while discards from the lobster fishery may be significant, there are no consistent data series for these. Recreational landings and discards are also included, but at a lower level than for SNE. Five survey indices are available, and fishery LPUE is also available but was not considered further. Age composition data is missing for commercial landings for more than half of the years covered (with none since 2014): however, the most recent landings estimate (for 2021) was 1 mt, at which point sampling does become difficult. Weight-at-age and M-at-age (see WP 14) are assumed to be time-invariant. Maturity data from surveys was modelled through time using a state-space logistic regression model (see WP 12). The available fishery and survey data were summarised through a useful plot (Figure 4.EGOM1) that I appreciated. Similarly, I found Figure 4.EGOM13 to be a very useful summary of the available catch-at-age data.

As with other stocks, a stepwise approach was taken to implementing changes, although the paucity of data meant that there were fewer alternatives available than (for example) with the WGoM stock. I think the key issue with this assessment is the missing data on discarded bycatch in the important lobster fishery. Consistently with other stocks, the WG did try to use WHAM to model missing lobster fishery bycatch through random effects on numbers-at-age. However, runs with random effects on NAA did not converge, and the lobster bycatch remains a potentially significant problem (recent indications are that such bycatch may be up to 6 times the commercial landings). The Panel also suggested rerunning the assessment using only the NEFSC surveys, as these seem to track cohort strength best, and I concur with this recommendation.

The Beverton-Holt stock-recruit model had been tried without success in the initial ASAP models, and was not attempted in the final WHAM models, but I feel the effort should at least have been made. It is also worth noting that there doesn't seem to be a stock-recruit plot given anywhere in the collated WG report or the relevant WPs, although the analyst (Micah Dean) did helpfully show one during his presentation to the review panel.

In the final proposed model, fleet selectivity was split into two blocks around 1997, which was well justified as this coincides with the start of spatial management, commercial trip limits and mesh size limits. Retrospective performance of the final model is good, as are the results of self-testing, but prediction skill is much less so. The stock remains at a very low level, although example forecasts indicate a rapid increase in SSB, and it is not clear where this would be coming from. The assessment is reminiscent for me of Rockall cod – historically a significant fishery but the target is now different (haddock in both cases), and the stock is at a very low level, but there are glimmers of recovery.

Overall, I would conclude that the assessment requires two additional aspects to be dealt with before being used in a future management track process – these are the missing lobster catch data, and runs with only the NEFSC surveys.

Below are some additional comments regarding figures in the collated WG report:

- Figure 4.EGOM5: as with the other stocks, I find the fits to the catch data to be suspiciously close. There are some positive residuals for the NEFSC survey during long periods as well.
- Figure 4.EGOM6: the model doesn't seem to pay much attention to the NEFSC survey at younger ages, and appears to be driven more by inshore indices.
- Figure 4.EGOM8: I would like to see this on absolute scales as well.

ToR 5: Status Determination Criteria (reference points)

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

I concluded that this ToR had been met in full. Once the assessment is accepted, the generation of reference points is quite formulaic in this area, and the formula has been applied appropriately for these stocks. I did, however, make some observations.

In the WG report, the stock-recruit relationships for the four stocks are referred to as "not well-determined", which I found odd. Three of the stocks have very linear stock-recruit relationships, while one follows a rather clear Beverton-Holt model, so to say that they are not "well-determined" doesn't seem quite right to me. The key point is that a linear stock-recruit model is not very useful for forecasts, medium-term simulations or reference point estimation, so I would have preferred to see this aspect highlighted in the text. Alternatively, a changepoint (hockey stick) model would possibly help here, as it would track the linearity over most (or all) of the observed SSB range, while not allowing recruitment to increase without limit beyond that range.

The analysis on recruitment time-periods for use in reference point estimation was comprehensive, and the case was made persuasively for using the full recruitment time-series for this purpose for all four stocks, but this still implies that the large recruitments seen in the past remain possible now – which seems ambitious, and may lead to higher reference points than would be justified on the basis of recent observations. For example, just considering recruitment in isolation, the WGoM down-step around 2010 looks very clear. However, the WG also showed that there are no consistent environmental drivers that apply across all stocks which could explain a change in recruitment success, and further plots were provided on R/SSB for all stocks which show that this is relatively consistent. This evidence does support the general use of the full-time series mean for recruitment in reference point estimation.

In terms of other biological parameters for reference point estimation, five-year means for weight-at-age, maturity-at-age and fishery selectivity were used, and these decisions were well supported. Despite some evidence that natural mortality M may be changing through time, only time-invariant M could be used in assessments (and hence reference point estimation) as none of the relevant WHAM assessment models would converge (I have noted some suggestions regarding this aspect above for ToR 4).

In the WG report and the relevant presentations, I would have appreciated some more information on how the reference points are estimated, once the parameters are defined. They were presumably based on the YPR and SSBPR curves, and will doubtless be very familiar to most of the WG, but a short statement of how it was done for these stocks would have been helpful for me. For example, it would have been useful to add lines to the YPR and SSBPR plots to show the fishing mortality, SSB and yield implied by fishing at F40%.

Finally, I had some more minor comments on the WG report plots: these are given below for the WGOM stock, but they would apply equally to the other stocks as the output template is consistent:

1. Figure 5.WGOM.1 – it is not clear what the colours represent.
2. Figure 5.WGOM.6 – this would be better with estimation lines added, as I noted above.
3. Figure 5.WGOM.8 – the SSB/SSB40% plot (upper) could be rescaled.

ToR 6: Projection methods

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

I concluded that this ToR has been met in full. As with ToR 5, once the runtime settings are determined, the generation of the forecasts seems semi-automatic for these stocks. The same runtime settings were used in projections as for reference point estimation, so most of my comments to ToR 5 above apply equally well here.

One additional issue was whether TAC constraints were used for the intermediate year in this area. That is, for an assessment being conducted in 2023, an ICES forecast for 2024 would sometimes assume the quota (TAC) was going to be taken in full in 2023, and sometimes it wouldn't – it would depend on whether this was thought to be reasonable and appropriate, given knowledge of the stock and fishery. I couldn't see this mentioned anywhere in the WG report, which suggests it isn't done in this area, but it would be helpful to clarify this.

I also noted that the forecasts are presented as simple $F = F_{40\%} \sim F_{msy}$ projections, without a sliding scale in F of the kind familiar to me from the ICES MSY approach. The forecasts therefore could be considered to be lacking an element of precautionarity, but whether that is critical would need to be tested through simulation. It looks like all of the stocks considered here will require rebuilding, however, so it may be that a different management approach altogether will be used.

Finally, although presented only for illustrative purposes in this research track, I did see that forecast F is well above the current F for the GB and EGOM stocks, while it is well below the current F for the SNE stock. For future management track processes, the WG would need to be careful to ensure that the forecast Fs seemed reasonable, given the recent history of the stock and knowledge about any ongoing developments in the fishery.

ToR 7: Research recommendations

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.

I concluded that this ToR had been met in full.

The WG began this process with quite an extensive list of recommendations, and (as ever with cod assessments) seemed to end up with a new list that was very nearly as extensive. As a Panel we went through these recommendations and provided a prioritisation of them, to compare and contrast with the WG's own prioritisation, and we also provided some recommendations of our own (some of which I have alluded to throughout this report, and all of which are listed in full in the Panel report).

Regarding the initial recommendations with which the research track process started, my reflections on these by ToR are as follows.

- ToR 1: the WG conducted extensive work on environmental drivers, although these tended not to be used in the assessments – quite often because of WHAM convergence issues, and I think a lot of potentially fruitful models fell down at this hurdle.

- ToR 2: the WG attempted to evaluate unaccounted-for catch. The main area where this could be an issue is EGoM where there is likely to be significant lobster fishery bycatch of cod, but data were not available by the time of the review meeting to include this. The Panel have recommended that this be addressed if possible in time for the next management track process. In the meantime, the WG used random effects to try and account for missing catch, which has helped with model misspecification but might also be masking other potential problems.
- ToR 3: the WG considered the inclusion of different surveys as requested, and retained several of them. The switch to the WHAM assessment model helped with this process, as it is able to handle 2020 survey data that were missing due to the covid pandemic.
- ToR 4: The TRAC cod assessment of EGoM cod is not yet resolved (TRAC: Transboundary Resource Assessment Committee, jointly between USA and Canada). The WG revised natural mortality estimates, and allowed random effects to account for some elements of time-varying mortality. They are now looking to follow up on previous runs which failed to converge – there is a lack of understanding yet on why this happened, and that will need to be addressed. The catch was split into two fleets (commercial and recreational) when possible. Stock-recruitment models were generally quite linear and so not very helpful for reference point estimation, although I have suggested in this report that a changepoint (hockey stick) model might be useful. It is worth noting that retrospective bias (long a major problem with assessments in this area) has greatly improved in the new assessments.
- ToR 5: the request to consider BRPs that accounted for changes in natural mortality was not really progressed – M was considered to be time-invariant throughout, largely due to convergence issues with WHAM when M was allowed to vary.
- ToR 6: the recruitment time period to use for projections was explored. The WG provided good argumentation around environmental drivers and the R/SSB time series, and the general conclusion to use the full time-series seemed reasonable.
- ToR 8: in previous years, there had been concerns about the ISmooth backup approach (which was advice based solely on survey indices). Fallback models were proposed for each stock this year, although these could not be reviewed as they had been discarded during the model development process.

ToR 8: Backup Assessment Approach

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.

This ToR was met in the sense that a backup assessment model was suggested for each stock, although these are imprecisely specified and the Panel was therefore unable to review them.

The US system requires there to be a backup model for all agreed assessment models, to be used in the situation where the final proposed model fails for whatever reason, or is no longer accepted. The approach used for the backup model here for all four stocks has been to recommend a "simplified" version of the WHAM model. I have two main concerns to raise here:

- The stepwise run evaluation presented in ToR 4 showed these backup models to be less appropriate than the final proposed model, so the backups actually look worse (in terms of fit and diagnostics) than the final proposed models. If the final models are rejected, the backup models will potentially come in for even greater criticism.
- If the Panel objections relate to the WHAM model itself (unlikely, but the high number of failed runs could be questioned), there are no alternatives given.

The WG presented arguments to support the use of simplified WHAM models in this context, however, citing simulation studies that showed that data-limited approaches are not necessarily better. This was reasonable, but I would have preferred the proposed alternatives to be more clearly specified – currently their settings are expressed rather vaguely, and there is a lot of text about models that "could" be used. This has made it impossible for the Panel (including myself) to review the backup models, other than to say that they are likely to be less appropriate than the final proposed models.

The Panel report concurs with this approach, although I have reservations as outlined above.

ToR 9: Atlantic Cod Stock Structure

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.

I concluded that this ToR had been met in full.

The Atlantic Cod Stock Structure Working Group (ACSSWG) met prior to the current research-track process, and the Panel were not required to provide a review of their work or conclusions. The reasoning and justifications for the current 4-stock structure were presented clearly in the relevant WPs and the presentations to the review Panel, although I would have found a more complete description of the process within the WG report itself very useful. The WG report section is very short (just 5 pages) and covers only the process used and the final output.

One aspect that I would like to learn more about is the potential impact on management of the change in stock structure assumptions. Over the last 2½ years, I have been involved in an ICES process to revise the assumed stock structure for cod in the North Sea, West of Scotland and Skagerrak, and one remaining (huge) problem is the implications for quota allocation and other management measures. During the review meeting, I questioned whether there were any such problems in the north-eastern US, and responses were provided during the presentation to show that new management structures were being considered. No decision has been made yet though, and this is something I would like to follow up on for my own interest.

I also questioned whether the distinctions between the four stocks are very clear, with no mixing, as the extensive fish movements described would seem to argue against this. Genetic analyses are convincing for reproductive isolation, but there does seem to be mixing at other times. The WG concluded that the closed boundaries are probably not entirely correct, but could be viewed as a decent approximation.

The decision to conduct four entirely separate assessments for the four stocks is interesting, given the likely mixing out with spawning times. The approach taken at ICES has been a multistock cod model in which stocks are assumed separate (and in their stock areas) during Q1, but with an unknown degree of mixing during Q2-4. This is possibly a more transparent approach, and concurs better (I would argue) with the available evidence, but it could lead to considerable management problems and would not necessarily produce more sustainable stocks in the north-eastern US.

Genetic sampling has been proposed on both sides of the Atlantic as a way to distinguish more clearly between substocks of a multistock complex out with the spawning period. This was also mentioned during the review meeting as a possible way to disaggregate cod from the winter and spring spawning populations in the WGoM stock. Such sampling has not yet been implemented, however, and it would take considerable time and resources to do so.

Appendix 1: Bibliography of materials provided for review

Working Group report

Atlantic Cod Research Track Working Group (2023). Research Track Assessment of Atlantic Cod. July 14, 2023.

Working papers

Lisa Kerr, Steve Cadrin and Scott Large. WP1: Stakeholder Meeting Summary.

Abigail Tyrell and Scott Large. WP2: Development of ecosystem indicators for Northwest Atlantic Cod.

Jamie Behan and Lisa Kerr. WP3: Environmental Influences on Atlantic Cod (*Gadus morhua*) Stock Dynamics.

Lisa Kerr and Steve Cadrin. WP4: Cod Stakeholder Meeting Summary.

Micah Dean. WP5: Recreational Discard Mortality for US Cod Stocks.

Lucy McGinnis, Keith Hankowsky, Max Grezlik, Gavin Fay, Steve Cadrin and Alex Hansell. WP6: Exploration of Fishery Dependent Data for Atlantic Cod (*Gadus morhua*).

Lucy McGinnis, Gavin Fay, Alex Hansell and Steve Cadrin. WP7: Standardizing landings per unit effort from fishery data with a comparison of spatial partitioning.

NEFSC. WP8: NEFSC Trawl Survey Expanded Figures.

NEFSC. WP9: Groundfish Survey Time Series Correlations.

Katie Lankowicz. WP10: Atlantic Cod VAST Model Technical Document.

Micah Dean, WP11: Eastern Gulf of Maine Sentinel Jig Survey.

Micah Dean and Charles Perretti. WP12: Time-Varying Maturity in US Cod Stocks.

Micah Dean. WP13: Weight-at-length in US cod stocks.

J. Roger Brothers, Lisa Kerr, Steve Cadrin and Abby Tyrell. WP14: Approximation of Natural Mortality Rate of Cod in Four Proposed Stock Areas Using Life History Traits

NEFSC. WP15: Atlantic Cod Model Selection Procedure (modified from American Plaice Research Track 2022).

Micah Dean. WP16: Assessment model development for the Eastern Gulf of Maine cod stock.

Charles Perretti. WP17: WGOM Cod Stock Assessment Model Development Working Paper.

Alex Hansell, Cole Carrano and Steve Cadrin. WP18: Assessment model development for Southern New England Cod.

Amanda Hart and Lisa Kerr. WP19: Assessment Model Development for the Georges Bank Atlantic Cod Stock.

Lisa Kerr and Jamie Behan, WP20: Defining Projections and Biological Reference points for Atlantic Cod in a Changing Ecosystem.

Presentations

Lisa Kerr and Rich McBride: ToR 9 - Stock Structure. 31 July 2023.

Scott Large and Jamie Behan: ToR 1 - Ecosystem and Climate Influences. 31 July 2023.

Charle Perretti and Kathy Sosebee: ToR 2 - Fishery Data. 31 July 2023.

Lisa Kerr: ToR 3 - Survey Data. 1 August 2023.

Lisa Kerr: Presentation on Data Processing Methods [additional presentation to address Panel questions]. 1 August 2023.

Charles Perretti: Western Gulf of Maine Cod (ToRs 4, 5, 6, 8). 1 August 2023.

Lisa Kerr: Presentation of Cohort Tracking Diagnostics and R/SSB Time Series [additional presentation to address Panel questions]. 2 August 2023.

Amanda Hart: Georges Bank Survey Explorations [additional presentation to address Panel questions]. 2 August 2023.

Alex Hansell, Cole Carrano, Steve Cadrin: Southern New England Cod (ToRs 4, 5, 6, 8). 2 August 2023.

Lisa Kerr: ToR 4,5,6,8 - Assessment, reference points, projections, back-up assessment. 2 August 2023.

Amanda Hart, Lisa Kerr: Georges Bank Cod (ToRs 4, 5, 6, 8). 2 August 2023.

Micah Dean: EGOM Cod Assessment Model (ToRs 4, 5, 6, 8). 3 August 2023.

Alex Hansell, Cole Carrano, Steve Cadrin: Southern New England Cod (ToRs 4, 5, 6, 8). 3 August 2023.

Lisa Kerr: ToR 7 - Research Recommendations. 3 August 2023.

Lucy McGinnis, Gavin Fay, Alex Hansell, Steve Cadrin: Standardizing Landings per Unit Effort from Cod Fishery Data. Presentation to the Cod Research Track WG, provided to the Panel for information.

Appendix 2: Performance Work Statement

Performance Work Statement (PWS)

National Oceanic and Atmospheric Administration (NOAA)

National Marine Fisheries Service (NMFS)

Center for Independent Experts (CIE) Program

External Independent Peer Review

Under Contract #1305M219DNFFK0025

Atlantic Cod Research Track Peer Review

July 31 – August 3, 2023

Background

The National Marine Fisheries Service (NMFS) is mandated by the Magnuson-Stevens Fishery Conservation and Management Act, Endangered Species Act, and Marine Mammal Protection Act to conserve, protect, and manage our nation's marine living resources based upon the best scientific information available (BSIA). NMFS science products, including scientific advice, are often controversial and may require timely scientific peer reviews that are strictly independent of all outside influences. A formal external process for independent expert reviews of the agency's scientific products and programs ensures their credibility. Therefore, external scientific peer reviews have been and continue to be essential to strengthening scientific quality assurance for fishery conservation and management actions.

Scientific peer review is defined as the organized review process where one or more qualified experts review scientific information to ensure quality and credibility. These expert(s) must conduct their peer review impartially, objectively, and without conflicts of interest. Each reviewer must also be independent from the development of the science, without influence from any position that the agency or constituent groups may have. Furthermore, the Office of Management and Budget (OMB), authorized by the Information Quality Act, requires all federal agencies to conduct peer reviews of highly influential and controversial science before dissemination, and that peer reviewers must be deemed qualified based on the OMB Peer Review Bulletin standards[1].

Scope

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

The purpose of this meeting will be to provide an external peer review of the Atlantic cod stocks. The requirements for the peer review follow. This Performance Work Statement (PWS) also includes: **Annex 1**: TORs for the research track, which are the responsibility of the analysts; **Annex 2:** a draft

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

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.

Schedule	Milestones and Deliverables
Within 2 weeks of award	Contractor selects and confirms reviewers
Approximately 2 weeks later	Contractor provides the pre-review documents to the reviewers
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.

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

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

Atlantic Cod Track Assessment Peer Review Meeting

July 31 –August 3, 2023

WebEx link: TBD

DRAFT AGENDA* (v. 4/27/2023)

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

Tuesday, August 1, 2023

Time	Topic	Presenter(s)	Notes

9 a.m. - 9:15 a.m.	Welcome/Logistics	Michele Traver, Assessment Process Lead Panel Chair	
9:15 a.m. - 10:30 a.m.	TOR #3		
10:30 a.m. - 10:45 a.m.	Break		
10:45 a.m. - 11:45 a.m.	TOR #3 cont.		
11:45 a.m. - 12:15 p.m.	Discussion/Summary	Review Panel	
12:15 p.m. - 12:30 p.m.	Public Comment	Public	
12:30 p.m. - 1:30 p.m.	Lunch		
1:30 p.m. - 3 p.m.	TOR #4		
3 p.m. - 3:15 p.m.	Break		
3:15 p.m. - 4:15 p.m.	TOR #4 cont.		
4:15 p.m. - 4:45 p.m.	Discussion/Summary	Review Panel	
4:45 p.m. - 5 p.m.	Public Comment	Public	
5 p.m.	Adjourn		

Wednesday, August 2, 2023

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

3:15 p.m. - 4:15 p.m.	TOR #8-9		
4:15 p.m. - 4:45 p.m.	Discussion/Summary	Review Panel	
4:45 p.m. - 5 p.m.	Public Comment	Public	
5 p.m.	Adjourn		

Thursday, August 3, 2023

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

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.

Appendix 3: Panel membership

The review panel consisted of the following 4 members:

Dr Jean-Jaques Maguire (chair, SSC)

Dr Coby Needle (independent CIE contractor)

Dr Noel Cadigan (independent CIE contractor)

Dr Steven Holmes (NIWA, New Zealand)

Other participants:

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