Gulf of Maine Haddock Research Track Review: Center for Independent Experts (CIE) Independent Peer Review

Dr Coby Needle, CIE Reviewer

May 2022

Executive Summary

This report summarises my notes and conclusions on the Gulf of Maine (GOM) haddock research track review process, as presented during the review meeting held online during 25-27 January 2022.

I found the presentations and report to be excellent, the process to be extremely well organised, and the staff involved from NMFS and elsewhere to be unfailingly helpful and open. The longer research-track process that has been instigated is not currently possible at ICES because of the overwhelming need for annual (or even more frequent) advice in Europe, but the approach does enable a much more thorough investigation and this is to be highly commended.

Given the models and frameworks used, the work has been very comprehensive and covered most (and maybe all) of the options open to the team. One thing I would have liked to have seen was exploration of a) alternative assessment models, and b) alternative approaches to BRP estimation. There may of course be issues around local expertise and familiarity, but the research track process is long and I would suggest that there could have been opportunities to consider alternatives in these areas. These may have resulted in the same outcome, but without testing you would never know.

With these caveats, I found that the ToRs had generally been met to the extent possible. The assessment was well presented and contained a great deal of relevant information, and I was happy to accept it as a valid representation of GOM haddock stock status. I did not identify any serious weaknesses, but there are a number of issues that I think could improve the assessments still further if addressed which I cover below.

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 now lead the Fisheries Assessment and Advice Programme at Marine Scotland – Science, based in the Marine Laboratory in Aberdeen, Scotland. I am also the Chief Fisheries Advisor for Scotland, and I continue to lead on the ICES assessment of North Sea lemon sole. I am heavily involved in the provision of fisheries advice to the Scottish Government.

Role in the Review Activities

Prior to the online meeting (25-27 January 2022), I thoroughly reviewed the background documents provided for the review panel, along with the extant versions of the stock assessment reports and summary reports for Gulf of Maine haddock (see Appendix 1 for a document list). During the online meeting, I participated in full in the plenary discussions during and after the presentations provided, as well as intersessionally by correspondence with the other review panel members and the review chair. I took copious notes during these discussions which form the basis of my comments below.

Overall Conclusions and Suggestions for Improvements to the SARC Process

My most recent experiences with benchmark-type stock assessment processes was the ICES North Sea lemon sole benchmark meeting, for which I was the principal stock coordinator; and the ongoing North Sea cod benchmark (WKCOD). In comparison with the ICES approach, I found the research-track

process to be highly structured and formalised, and I felt that it certainly benefitted from this. Throughout the meeting it was very clear what was expected, and what the information and subsequent conclusions would be used for, and I considered the meeting and bracketing work to be very instructive and educational. The excellent work by both presenters and rapporteurs helped greatly with this, and the meeting was very well chaired with the schedule being followed closely and all involved being given the opportunity to speak. The WG participants and those who presented at the review meeting had clearly invested a great deal of time and effort in producing work of a very high standard, and everyone involved is to be thoroughly commended.

Although the ICES system has changed recently to much shorter benchmark meetings, the principal advantage of the old ICES process that I felt was lacking in the new US research-track system was the opportunity given to the review panel to make very constructive contributions to the assessments themselves. We were able to comment and review the existing work, and should we have seen the need we could have rejected the assessments, but the time available for the review was too short for further data or model developments, and I found this somewhat frustrating. ICES reviewers under the old ICES system were more involved in the actual assessment work, at least in my experience, and so had more opportunity to make a constructive difference. It is of, course, more difficult to get reviewers to set aside the additional time needed for this, and it is precluded by the timetable of the research-track process, but I think I would have appreciated more opportunity to get more involved, had it been possible.

Regarding the assessment, I found that a tremendous amount of work had been presented and that the ToRs had mostly been fulfilled. The Gulf of Maine haddock assessment did as good a job as could have been expected given the commercial and recreational catch data available, although there was a good deal of data inference and imputation required to generate full age-structured datasets for landings and discards for both of these fisheries. The one element I felt was missing was an exploration and consideration of alternative models for assessment and the BRP estimation, and it would be beneficial in future work to explore the stock dynamics indicated by different approaches. There would likely be issues around local expertise and familiarity with alternatives, and the time and resources for this exploration may well be limited. Given the clear stock signals, particularly from the recent strong year-classes, it is unlikely that these approaches would necessarily lead to different perceptions of stock dynamics, but they would at least increase confidence in the baseline catch-at-age stock assessment.

Overall, I was very impressed by the ability and dedication of the participating scientists, which facilitated a successful and enjoyable meeting.

Comments for each ToR of each assessment

ToR 1: Review existing research efforts, data, and habitat information in the Gulf of Maine and Georges Bank, identify any findings relevant to influences of ecosystem conditions on haddock, and consider those findings, as appropriate, in addressing other TORs. For processes that the working group deems important and promising that are not currently feasible to consider quantitatively, describe next steps for development, testing, and review of quantitative relationships and how they could best inform assessments.

I concur with the Panel conclusion for this ToR, and I found this ToR to be *partially completed*, in that the available studies were presented and reviewed, but they are not yet implemented and I didn't see a firm plan for doing so.

The analyses presented for this ToR focussed on a habitat model for GOM haddock, based on random forest machine learning methodologies (Friedland et al., 2020). I was previously unfamiliar with this approach, and unfortunately I didn't find the presentation or report to be sufficient in helping to fully understand the model.

It was noted in the meeting that the Gulf of Maine is one of the most rapidly warming marine areas in the world, and this is very likely to be having an effect on many local ecosystem features (including haddock recruitment). However, the main issue (which the report acknowledges) is that the analysis is tautological. As the text notes: "the habitat scores may be driven by changes in haddock distribution, as opposed to habitat driving haddock distribution." So the argument becomes very circular, and it is difficult to determine the direction of causality. The area of suitable habitat has increased at the same time as stock abundance, so it is not clear if the model is explaining ecosystem effects on haddock, or just the recent increase in haddock abundance. This is partly due to the lack of a mechanistic causal link between the significant model variables and haddock abundance, and this should be addressed.

Finally, it was noted that the "distribution of habitat would be consistent with the hypothesis of a single stock in the GOM/GB region." Follow-up questions on this showed that further work was done during the research track on the stock structure question, but that this will be covered in the Georges Bank review meeting [to be held in March 2022]. This is unfortunate, as the findings will be equally relevant to both areas.

This is a potentially useful approach, given the strong warming signals in the area, but **I would recommend the habitat model needs to be focussed on the prediction or forecasting of future recruitment for Gulf of Maine haddock in particular (or GOM/GB haddock if the assumed stock structure changes).**

The overriding question of stock structure in the GOM/GB region needs to be addressed as a matter of priority, as some evidence (contiguous landings distributions, for example) suggests that there may be less than three stocks in the area, and this would be critical to recognise in the assessment process.

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

I concur with the Panel conclusion for this ToR, and I found this ToR to be *completed in full* – the available catch data was presented in detail, and the collection and collation of the data is as good as could be expected given the difficult nature of some of the fishery components (especially the recreational fishery).

The discard observer coverage is impressive at around 15-20%, which compares very well with the usual 2-3% coverage that we achieve in most European fisheries. The observer programme started in 1989, so surveys are used to reconstruct what pre-1989 historical discards are likely to have been

(based on a survey-filter selectivity assumption). This is a reasonable approach that has been used effectively elsewhere, and can afford to be approximate because those years are only really used in reference point determination.

There were some issues raised during the review meeting around recreational catch, estimation of CPUE indices, and the MREM fleet. These were answered in full on the basis on existing knowledge during the meeting, but some gaps remain and I have suggested that these could be addressed through a set of recommendations (see below).

One comment (which is not really a recommendation) regards the table formatting. When asked about this, the NEFSC staff highlighted that formats were stipulated by regulations on access for different types of reader, and while this is understandable it does result in tables which are large and unwieldy (covering many pages each). This makes following the contents of tables difficult for reviewers.

There remain some shortcomings with the recreational data estimation, which comprises over 50% of the total catch in some years and is therefore of critical importance. While much of this data are gathered by national programmes, and are therefore out of NEFSC control, **the further investigation of the two-fleet model in the assessment could help to address some of the concerns about potential uncertainty and bias in recreational catch data.**

The presentation looked at recent developments in the generation of catch per unit effort (CPUE) data series from the various commercial fisheries. **I would suggest that CPUE data series be considered very carefully before being used, if at all.** The history of fisheries science is full of stocks which have collapsed because of over-reliance on commercial CPUE indices that have remained positive in the face of falling stocks, due to fishery hyper-aggregation, and it is very important to guard against this. Furthermore, the use of commercial CPUE in an assessment that also includes the commercial catch data is in danger of "using the same data twice", and therefore not accounting properly for fishery-independent data.

The GOM area has a Maximum Retention Electronic Monitoring (MREM) fishery, which includes special permits to enable the capture and landings of undersized fish in return for carrying cameras and no discarding. This encompasses a small proportion of the total fleet thus far. Since 2016, roughly 25 vessels have also been involved in the Audit Model, which tries to replicate human observers by passing fish under a camera for subsequent morphometric analysis. Both programmes run under EFP (a regulation exemption), but as yet are directed at compliance rather than science use. REM technology has great potential to augment (and in some cases replace) the work of onboard observers (see for example Needle et al. 2015), and **I would encourage a close consideration of whether these REM programmes could be used for discard estimation.**

ToR 3: Present the survey data being used in the assessment (e.g., indices of relative or absolute abundance, recruitment, state surveys, age-length data, etc.). Characterize the uncertainty in these sources of data.

I concur with the Panel conclusion for this ToR, and I found this ToR to be *completed in full.* This was another excellent presentation, with (I would say) even more detail and information that the corresponding talk on the commercial catch data. The Gulf of Maine area is well-surveyed and the

resulting data are analysed in considerable depth – in many ways, the survey data look to me to be more reliable that the commercial catch data (given issues around discard estimation and the recreational fishery data).

I appreciated the presentation of bivariate scatterplots to illustrate year-class tracking for some of the surveys, but I would have liked to have seen these for all surveys (and for them to have included correlation coefficients for each subplot). The cohort tracking of some of the inshore surveys did not appear to be very good, based on the summaries presented, but bivariate scatterplots would have helped with this conclusion.

The age-based survey indices are entirely design-based, without the use of a geostatistical modelling approach to smooth between areas (and years). The latter approach is increasingly common is ICES assessments, as it deals well with outliers and allows inference about areas not sampled, and it might be appropriate to consider this for GOM haddock survey indices as well.

The estimation of weights from the ASMFC shrimp bottom trawl survey was quite unusual. Individual fish are not weighed, but the total weight of haddock in a haul was recorded, so individual weights were estimated using the total haddock weight for those hauls for which only one haddock was caught. This is quite limiting, as there may not be many hauls with only one haddock, and could also be biased if there was something unusual about those hauls. It may be better in future to either a) consider whether fish weights could be measured on this survey), or b) not estimate weights at all.

The NMFS bottom longline survey appears to be very important for rough-bottom areas, provides useful information and should certainly be considered again in future. I find this personally interesting, as in Scotland we have several rough-ground areas which we have never been able to survey, but which might be acting as valuable refugia for cod in particular. **I would, therefore, recommend that the NMFS bottom longline survey be considered again for inclusion in the near future.**

In addition to the main NEFSC surveys, there are several state-run inshore trawl surveys which could provide valuable information (i.e. they catch haddock) but are not yet used in the final assessment. In a way this is appropriate, as the inshore areas are not always representative of the full stock extent (and they do not seem to track cohorts very well), but it is also important to use all available data where possible and appropriate. Therefore, **I would recommend that geostatistical methods be considered to account for these smaller surveys in inshore areas.**

ToR 4: 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 model, and evaluate the strength and direction of any retrospective pattern(s) in both the current and the previously accepted model. Enumerate possible sources of the retrospective patterns and characterize plausibility, if possible.

I concur with the Panel conclusion for this ToR, and I found this ToR to be *completed in full*. The final model fitted the data well, was well explained and justified in the main. It was very consistent with previous runs, and presented *reasonable* retro patterns (the main issue is underestimation of the 2013 year-class).

In this analysis, the ASAP model was investigated in considerable detail, with many different formulations tried in progressive steps from the previous model to the new one. The main thing

lacking for me in the presentation and report is an exploration of alternative assessments methods – these could have been different age-based catch-at-age analyses, or more speculatively could have covered time-series approaches, spatial models, length-based models, etc. In ICES benchmarks we would often try something different, at least to see if the stock perception changes (and if so, why) – otherwise there is a danger of interpreting the single model fit as the only definitive truth. **I would, therefore, recommend that alternative models be explored in future research track assessments, rather than just different formulations of the same model.**

There is a penalty function used for annual recruitment deviations, essentially partially constraining recruitments to the long-term geometric mean. The allowed CV is further restricted for the three most recent cohorts. I wonder if this is appropriate, given the sporadic nature of large haddock recruitments? The TSA assessment used until recently for Northern Shelf haddock allowed greater variance for large year-classes. The assessment team reported that they did consider this issue, but that it cannot currently be addressed with the available ASAP implementation. The updated report shows that the 2013 year-class has an estimated CV of around 50%, which is probably reasonable for the purpose (and with an assumption of lognormal recruitment), but for me this is still a good issue to consider further in future work.

Natural mortality is assumed to be age and time invariant at M=0.2. This appears to be the usual guesstimate, without a strong basis in biology or the local ecosystem. Incorrect assumptions about M could lead to bias and retro issues if M is in fact changing, and it will certainly be different for different ages. The assessment team also considered this issue - for haddock, the retro pattern suggests decreasing M, while growth patterns suggest increasing M, but in the absence of multispecies assessments in this region it is difficult to generate better estimates of M. However, predator-prey relationships are likely to be changing as the area warms, and various growth-based approaches to estimating M (e.g. Lorenzen 2000) are possible in the absence of predator-prey information. In the meantime, **I would recommend that the next management track assessment should include a sensitivity analysis exploring whether management advice is robust to the constant M assumption.**

The Review Panel had many questions around this ToR, the majority of which were answered clearly and in full. The structure of the research-track process precludes a very close analysis and consideration by the reviewers of the diagnostics for the intermediate steps between the previous assessment and the new proposed assessment, which makes it harder for us to understand (and agree with) the reasoning behind some of the model choices. **I would, therefore, recommend that some key diagnostics from intermediate runs be presented to reviewers in the next management track process, to facilitate the review process.**

One of the alternative formulations presented set the Effective Sample Size (ESS) for the commercial catch data to 75%, which should downweight the importance of the catch data in the assessment. The fact that this model did not result in a tighter fit to the index time-series suggests that the model is determined to underestimate (to a large extent) the index observations of very large haddock stocks in recent years. I questioned why this might be, but there does not seem to be a ready answer and this would be something to consider again for next time.

Finally, the estimated retrospective bias is larger than would normally be considered acceptable in ICES assessments. However, the most recent peel does suggest that the retro bias may be dissipating in recent years. On questioning this, we were informed that a different criterion is used in the US –

retro bias is classified as either major or minor, based on whether a retro-adjusted estimate falls within or without the assessment CI. It would be interesting to explore whether this approach is more robust and risk-averse than the more drastic measures taken in ICES, where any assessment with a Mohn's rho > 20% is generally downgraded to a data-limited category 3 assessment.

ToR 5: Update or redefine status determination criteria (SDC point estimates or proxies for BMSY, BTHRESHOLD, FMSY and MSY) and provide estimates of their uncertainty. If analytic model-based estimates are unavailable, consider recommending alternative measurable proxies for BRPs.

I concur with the Panel conclusion for this ToR, and I found this ToR to be *completed*. I don't agree fully with all the conclusions, but the ToR has been met as requested.

The report states: "The WG saw no compelling reason to select a different F_{MSY} proxy than the F40% metric that had been adopted previously." In my view, a variable SRR should just lead to large uncertainty in the F_{MSY} estimate, but the median should still be appropriate and I don't see a strong justification for not (at least) considering a more stochastic, production-model-based approach to MSY estimation. F(40%) just seems a little ad hoc to me. On the other hand, I realise that the F(40%) approach is very standard in this region, and in any case the SR plot is so far from a regular relationship that a stochastic production model might not be worthwhile. My issue is more around whether a stochastic approach has been considered, and it might also be more fruitful to investigate length-based proxies for F(msy). **I would therefore recommend that both stochastic production models and length-based proxies be explored to determine whether the F(40%) assumption remains appropriate.**

I also have a concern over the 5-year average of weights-at-age used in the BRP estimation. There is some evidence (although not as strong as I would have expected) of cohort effects on growth and therefore weights, and **I would recommend that the use of cohort-based weights models (e.g. Jaworski 2011) be considered for BRP estimation** (see also comments for ToR 6 and 10).

ToR 6: Define the methodology for performing short-term projections of catch and biomass under alternative harvest scenarios, including the assumptions of fishery selectivity, weights at age, maturity, and recruitment.

I concur with the Panel conclusion for this ToR, and I conclude that this ToR has been *completed* – the methodology has been defined, and is quite clear. My remaining concern is over the way that weights-at-age are generated for short-term forecasts, and my recommendation around this issue for ToR 5 is equally applicable here.

ToR 7: Review, evaluate and report on the status of the Stock Assessment Review Committee (SARC) and Working Group research recommendations listed in most recent SARC reviewed assessment and review panel reports. Identify new research recommendations.

I concur with the Panel conclusion for this ToR, and I conclude that this ToR is *complete with some reservations*.

The main issue for me is the number of SARC and WG research recommendations that are postponed to a wider analysis process, given that they are not specific to GOM haddock. I feel that while this is true, understanding the stock dynamics of GOM haddock is critically dependent on some of these

research recommendations being undertaken, and this research-track process would have been a good opportunity at least to get the work started (this could have been widened to other stocks subsequently). The key examples are those around additional stock-recruitment models, density-dependent selectivity, and density-dependent growth – these are vital to understand for a stock such as GOM haddock that is so dependent on sporadic large year-classes. In this context, it is a shame that many of the recommendations were put back to general groups - there should have been at least a note on how these recommendations apply (or could apply) to GOM haddock. As written, the text suggests that these issues aren't very important.

Another SARC/WG recommendation calls for the development of CPUE or LPUE indices from the commercial fishery. I remain concerned about CPUE or LPUE indices, given the potential for hyper-aggregation of fishing effort as the stock declines. CPUE may be useful for non-target species, but (I would suggest) not otherwise and they are not widely used in ICES assessments now.

Although the GOM and GB assessments were conducted at the same time, the reviews were not which is unfortunate. No attempt was made to model a combined stock – this is also unfortunate, because the specific stock retros are in different directions, and putting them together may iron this out. Overall, **I would recommend that further consideration be given to the question of stock structure in the GOM/GB area, as some of the evidence suggests that the assumption of three independent stocks may not be correct.**

ToR 8: Develop a "Plan B" for use if the accepted assessment model fails in the future.

I concur with the Panel conclusion for this ToR, and I conclude that this ToR is *complete*.

The alternative assessment (Plan B Smooth) is very reminiscent of the ICES 2-over-3 rule applied to a survey-based assessment SSB estimate, and that has been used successfully for data-limited ICES stocks for a number of years. Indeed, assessment scientists in ICES are much quicker than in the US to move to index-based approaches. Consideration does need to be given to the on/off ramp for switching to a Plan B – it is generally easier to move down to a data-limited approach than it is to move back up to a full catch-at-age-based analysis, and the criteria for movement in both directions needs to be well-specified.

On the specific Plan B Smooth, the use of the 0.3 span in the loess smoother seems to have been based on trial-and-error in the past, and is fixed by custom now. **I would recommend an exploration of loess AIC approaches to determine an appropriate loess span for this method.**

I questioned whether the Plan B Smooth output would be included in the management-track report as a matter of course. It seems that this is not done in the US – the alternative assessment is only ever presented if the main model fails, and indeed there is not a tradition of presenting alternative exploratory model runs. The reason is to prevent stakeholders selecting an alternative that is more to their liking, so diluting the main advice, but relying on one model only could lead to unwarranted confidence in that outcome.

In the reviewer's report for SARC59, I was asked (or volunteered) to include a SURBAR run (this is a survey-based assessment model used for exploratory runs in ICES, as well as some advice). **I would recommend that a more model-based approach be considered in the future for a Plan B option.** I

would prefer this myself, as it brings in some stock dynamics structure rather than just the smoothed survey indices, and would enable projections.

ToR 9: Review and present any research related to recruitment processes (e.g., spawning and larval transport, and retention), and potential hypotheses for large recruitment events.

I concur with the Panel conclusion for this ToR. This was a relatively short section in both the report and the presentation, and my conclusion is that the ToR was *partially completed*.

The GOM haddock fishery is currently reliant on one very large year-class, and trying to understand why that happened (and whether it will happen again) is therefore absolutely key to understanding the dynamics of the stock and managing the fishery. The Georges Bank (GB) haddock recruitment may be driven by the fall phytoplankton bloom (haddock are detritus feeders and so could benefit from this directly), but this doesn't seem to apply to the GOM stock. However, only seven years of GOM bloom data have been looked at thus far. **I would therefore recommend that the relationship between the fall bloom and GOM haddock recruitment be updated to consider more recent information.**

The possible linkages between the GOM and GB stocks could also be having an impact on the recruitment in one or both of these areas. For example, it could be that haddock are spawning in GOM, drifting to GB as larvae, and then moving back to GOM to spawn themselves (i.e. natal homing, as observed with haddock spawning in the West of Scotland and moving between there and the North Sea). Linkages of this kind would be useful to explore further in future research tracks – I understand that stock structure will be presented as part of the GB research track meeting, but unfortunately that is a separate review meeting this year.

Finally, the 2013 year-class was so much bigger than any other observed for this stock, that it seems unlikely that there were not stronger density-dependent inhibitors on growth (as seen for large year-classes in most other haddock stocks). **Recommendation: to investigate why growth was not significantly limited for the 2013 year-class, over and above the already reduced growth for the 2010s** (see also comments for ToR 10).

ToR 10: Review and present any research related to density-dependent growth.

I concur with the Panel conclusion for this ToR, that the ToR is *complete*. Research has been presented and reviewed, so in that sense it is done, although I'm not overly convinced about the conclusions.

I remain surprised that cohort effects aren't considered in projection of weights, given that these are generally very significant for haddock (the 2013 year-class in particular must have been growth-impacted, although the analysis suggests it wasn't unusual for the 2010s). I have suggested a recommendation on this issue for ToRs 5 and 9 above.

References

Jaworski, A. (2011). Evaluation of methods for predicting mean weight-at-age: an application in forecasting yield of four haddock (Melanogrammus aeglefinus) stocks in the Northeast Atlantic. *Fisheries Research*, DOI: 10.1016/j.fishres.2011.01.017.

Lorenzen, K. (2000). Allometry of natural mortality as a basis for assessing optimal release size in fish-stocking programmes. *Can. J. Fish. Aquat. Sci.* **57**: 2374–2381.

Needle, C. L., Dinsdale, R., Buch, T. B., Catarino, R. M. D., Drewery, J. and Butler, N. (2015). Scottish science applications of Remote Electronic Monitoring, *ICES Journal of Marine Science* **72**(4): 1214–1229. DOI: 10.1093/icesjms/fsu225.

Appendix 1: Bibliography of materials provided for review

Kevin D. Friedland, Joseph A. Langan, Scott I. Large, Rebecca L. Selden, Jason S. Link, Reg A. Watson, and Jeremy S. Collie (2020). Changes in higher trophic level productivity, diversity and niche space in a rapidly warming continental shelf ecosystem. Science of the Total Environment, 704:135270. https://doi.org/10.1016/j.scitotenv.2019.135270.

Richard Merrick, Adrian Jordaan and Conor McManus (2021). 2021 Management Track Peer Review Panel Report. [September 2021 Management Track Assessment Report.09222021.docx].

NEFSC (2022). Final Report of the Haddock Research Track Assessment Working Group. [Haddock_WG_Report_DRAFT_GOM_ONLY_text_v2.pdf. Haddock_WG_Report_DRAFT_GOM_ONLY_tables_v2.pdf, Haddock_WG_Report_DRAFT_GOM_ONLY_figures_v2.pdf].

Smith S.L., Golden A.S., Ramenzoni V., Zemeckis D.R., and Jensen O.P. (2020). Adaptation and resilience of commercial fishers in the Northeast United States during the early stages of the COVID-19 pandemic. PLoS ONE **15**(12): e0243886. https://doi.org/10.1371/journal.pone.0243886

Appendix 2: Statement of Work for Dr Coby Needle

<div align="center">

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

Gulf of Maine Haddock

Research Track Peer Review

January 25 -27, 2022

</div>

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]. Further information on the Center for Independent Experts (CIE) program may be obtained from www.ciereviews.org.

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 Gulf of Maine haddock stock. The requirements for the peer review follow. This Performance Work Statement (PWS) also

---

[1] https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/memoranda/2005/m05-03.pdf

includes: **Appendix 2.1**: TORs for the research track, which are the responsibility of the analysts; **Appendix 2.2:** a draft meeting agenda; **Appendix 2.3:** Individual Independent Review Report Requirements; and **Appendix 2.4:** 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. labour and travel) is not covered by this contract.

Each reviewer will write an individual review report in accordance with the PWS, OMB Guidelines, and the ToRs below. Modifications to the PWS and ToRs cannot be made during the peer review, and any PWS or ToRs modifications prior to the peer review shall be approved by the Contracting Officer's Representative (COR) and the CIE contractor. All TORs must be addressed in each reviewer's report. The reviewers shall have working knowledge and recent experience in the use and application of index-based, age-based, and state-space stock assessment models, including familiarity with retrospective patterns and how catch advice is provided from stock assessment models. In addition, knowledge and experience with simulation analyses is required.

Tasks for Reviewers

●   Review the background materials and reports prior to the review meeting.

  ○   Two weeks before the peer review, the Assessment Process Lead will electronically disseminate all necessary background information and reports to the CIE reviewers for the peer review.

●   Attend and participate in the panel review meeting.

  ○   The meeting will consist of presentations by NOAA and other scientists, stock assessment authors and others to facilitate the review, to provide any additional information required by the reviewers, and to answer any questions from reviewers.

●   Reviewers shall conduct an independent peer review in accordance with the requirements specified in this PWS and TORs, in adherence with the required formatting and content guidelines; reviewers are not required to reach a consensus.

●   Each reviewer shall assist the Peer Review Panel (co)Chair with contributions to the Peer Reviewer Summary Report.

●   Deliver individual Independent Reviewer Reports to the Government according to the specified milestone dates.

●   This report should explain whether each research track Term of Reference was or was not completed successfully during the peer review meeting, using the criteria specified below in the "Tasks for Peer Review Panel."

●   If any existing Biological Reference Points (BRP) or their proxies are considered inappropriate, the Independent Report should include recommendations and justification for suitable alternatives. If such alternatives cannot be identified, then the report should indicate that the existing BRPs are the best available at this time.

●   During the meeting, additional questions that were not in the Terms of Reference but that are directly related to the assessments and research topics may be raised. Comments on these

questions should be included in a separate section at the end of the Independent Report produced by each reviewer.

● The Independent Report can also be used to provide greater detail than the Peer Reviewer Summary Report on specific stock assessment Terms of Reference or on additional questions raised during the meeting.

Tasks for Review panel

● During the peer review meeting, the panel is to determine whether each research track Term of Reference (ToR) was or was not completed successfully. To make this determination, panellists 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 Haddock Research Track Working Group.

1) The Peer Review Panel (co)Chair, with the assistance from the reviewers, will write the Peer Reviewer Summary Report. Each reviewer and the (co)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. Reviewers are not required to reach a consensus.

The (co)chair's objective during this Peer Reviewer Summary Report development process will be to identify or facilitate the finding of an agreement rather than forcing the panel to reach an agreement. The (co)chair will take the lead in editing and completing this report. The (co)chair may express their opinion on each research track Term of Reference, either as part of the group opinion, or as a separate minority opinion. The Peer Reviewer Summary Report will not be submitted, reviewed, or approved by the Contractor.

Place of Performance

The place of performance shall be held remotely, via WebEx video conferencing.

Period of Performance

The period of performance shall be from the time of award through April 2022. Each reviewer's duties shall not exceed **14** days to complete all required tasks.

Schedule of Milestones and Deliverables: The contractor shall complete the tasks and deliverables in accordance with the following schedule.

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
January 25-27, 2022	Panel review meeting
Approximately 2 weeks later	Contractor receives draft reports
Within 2 weeks of receiving draft reports	Contractor submits final reports to the Government

* The Peer Reviewer Summary Report will not be submitted to, reviewed, or approved by the Contractor.

Applicable Performance Standards

The acceptance of the contract deliverables shall be based on three performance standards:

(1) The reports shall be completed in accordance with the required formatting and content;

(2) The reports shall address each ToR as specified;

(3) The reports shall be delivered as specified in the schedule of milestones and deliverables.

Travel

No travel is necessary, as this meeting is being held remotely.

Restricted or Limited Use of Data

The contractors may be required to sign and adhere to a non-disclosure agreement.

NMFS Project Contact

Michele Traver, NEFSC Assessment Process Lead

Northeast Fisheries Science Center

166 Water Street, Woods Hole, MA 02543

Michele.Traver@noaa.gov

Appendix 2.1. Haddock Research Track Terms of Reference

1.   Review existing research efforts, data, and habitat information in the Gulf of Maine and Georges Bank, identify any findings relevant to influences of ecosystem conditions on haddock, and consider those findings, as appropriate, in addressing other TORs. For processes that the working group deems important and promising that are not currently feasible to consider quantitatively, describe next steps for development, testing, and review of quantitative relationships and how they could best inform assessments.

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 being used in the assessment (e.g., indices of relative or absolute abundance, recruitment, state surveys, age-length data, etc.). Characterize the uncertainty in these sources of data.

4.   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 model, and evaluate the strength and direction of any retrospective pattern(s) in both the current and the previously accepted model. Enumerate possible sources of the retrospective patterns and characterize plausibility, if possible.

5.   Update or redefine status determination criteria (SDC point estimates or proxies for B_{MSY}, $B_{THRESHOLD}$, F_{MSY} and MSY) and provide estimates of their uncertainty. If analytic model-based estimates are unavailable, consider recommending alternative measurable proxies for BRPs.

6.   Define the methodology for performing short-term projections of catch and biomass under alternative harvest scenarios, including the assumptions of fishery selectivity, weights at age, maturity, and recruitment.

7.   Review, evaluate and report on the status of the Stock Assessment Review Committee (SARC) and Working Group research recommendations listed in most recent SARC reviewed assessment and review panel reports. Identify new research recommendations.

8.   Develop a "Plan B" for use if the accepted assessment model fails in the future.

9.   Review and present any research related to recruitment processes (e.g., spawning and larval transport, and retention), and potential hypotheses for large recruitment events.

10.   Review and present any research related to density-dependent growth.

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.

Appendix 2.2. Draft Review Meeting Agenda

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

Gulf of Maine Haddock

Research Track Assessment Peer Review Meeting

January 25 - 27, 2022

WebEx link: https://www.google.com/url?q=https://noaanmfs-meets.webex.com/noaanmfs-meets/j.php?MTID%3Dmac73d9098b946224d02f64d3d429d0b3&sa=D&source=calendar&ust=1633797763460762&usg=AOvVaw12N6T8o0JhfZohwPjxr3UL

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

DRAFT AGENDA* (v. 1/5/2022)

*All times are approximate, and may be changed at the discretion of the Peer Review Panel chair. The meeting is open to the public; however, during the Report Writing sessions we ask that the public refrain from engaging in discussion with the Peer Review Panel.

Tuesday, January 25, 2022

Time	Topic	Presenter(s)	Notes
9 a.m. - 9:30 a.m.	Welcome/Logistics Introductions/Agenda/Conduct of Meeting	Michele Traver, Assessment Process Lead Russ Brown, PopDy Branch Chief Richard Merrick, Panel Chair	
9:30 a.m. - 10:30 a.m.	ToR #2	Charles Perretti	Catch data
10:30 a.m. - 10:45 a.m.	Break		
10:45 a.m. - 11:15 a.m.	ToR #2 cont.	Charles Perretti	Catch data
11:15 a.m. - 11:45 a.m.	Discussion/Summary	Review Panel	
11:45 a.m. - 12 p.m.	Public Comment	Public	
12 p.m. - 1 p.m.	Lunch		
1 p.m. - 2:30 p.m.	ToR #3	Charles Perretti	Survey data

Time	Topic	Presenter(s)	Notes
2:30 p.m. - 2:45 p.m.	Break		
2:45 p.m. - 3:45 p.m.	TORs #1, #9, and #10	Charles Perretti	Ecosystem, Recruitment Processes, and Density Dependent Growth
3:45 p.m. - 4:15 p.m.	Discussion/Summary	Review Panel	
4:15 p.m. - 4:30 p.m.	Public Comment	Public	
4:30 p.m.	Adjourn		

Wednesday, January 26, 2022

Time	Topic	Presenter(s)	Notes
9 a.m. - 9:15 a.m.	Welcome/Logistics	Michele Traver, Assessment Process Lead Richard Merrick, Panel Chair	
9:15 a.m. - 10:15 a.m.	ToR #4	Charles Perretti	Mortality, Recruitment and Biomass Estimates
10:15 a.m. - 10:30 a.m.	Break		
10:30 a.m. - 11:45 a.m.	TORs #5, #6, and #8	Charles Perretti	BRPs, Projections, and Alternative Assessment Approach
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. - 2 p.m.	ToR #7	Brian Linton	Research Recommendations
2 p.m. - 2:30 p.m.	Discussion/Summary	Review Panel	
2:30 p.m. - 2:45 p.m.	Public Comment	Public	
2:45 p.m. - 3 p.m.	Break		
3 p.m. - 4 p.m.	Follow-ups/Key Points	Review Panel	
4 p.m.	Adjourn		

Thursday, January 27, 2022

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

Appendix 2.3. 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.   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.

   a. Reviewers should discuss their independent views on each ToR even if these were consistent with those of other panellists, but especially where there were divergent views.

   b. Reviewers should elaborate on any points raised in the Peer Reviewer Summary Report that they believe might require further clarification.

   c. 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 2.4. Peer Reviewer Summary Report Requirements

1. The main body of the report shall consist of an introduction prepared by the Research Track Peer Review Panel chair that will include the background and a review of activities and comments on the appropriateness of the process in reaching the goals of the peer review meeting. Following the introduction, for each assessment /research topic reviewed, the report should address whether or not each Term of Reference of the Research Track Working Group was completed successfully. For each Term of Reference, the Peer Reviewer Summary Report should state why that Term of Reference was or was not completed successfully.

To make this determination, the peer review panel chair and reviewers should consider whether or not the work provides a scientifically credible basis for developing fishery management advice. If the reviewers and peer review panel chair do not reach an agreement on a Term of Reference, the report should explain why. It is permissible to express majority as well as minority opinions.

The report may include recommendations on how to improve future assessments.

2. If any existing Biological Reference Points (BRPs) or BRP proxies are considered inappropriate, include recommendations and justification for alternatives. If such alternatives cannot be identified, then indicate that the existing BRPs or BRP proxies are the best available at this time.

3. The report shall also include the bibliography of all materials provided during the peer review meeting, and relevant papers cited in the Peer Reviewer Summary Report, along with a copy of the CIE Performance Work Statement.

The report shall also include as a separate appendix the assessment Terms of Reference used for the peer review meeting, including any changes to the Terms of Reference or specific topics/issues directly related to the assessments and requiring Panel advice.

Appendix 3: Panel membership or other pertinent information from the panel review meeting.

CIE panel membership: Dr Kevin Stokes, Dr Anders Nielsen, Dr Coby Needle.