Center for Independent Experts independent review report of Georges Bank and Eastern Georges Bank Haddock research track

Anders Nielsen

Executive Summary

The Haddock Research Track Working Group (hereafter referred to as the “research track”) has addressed all terms of references (TORs), has presented their work in two well written and very focused reports (one for Georges Bank haddock (GB haddock) and one for Eastern Georges Bank (EGB haddock)), and has presented the findings at the review meeting. The GB/EGB haddock stock identity issue is complicated, and the research track failed to reach consensus on a few issues related to stock identity and how it is used within the assessment models. The research track has met all TORs, has presented an impressive amount of interesting and relevant research (e.g., on modelling, diagnostics, habitat, and recruitment), and has presented two updated assessment models with corresponding reference points and short-term forecast procedures. In isolation, each of the two models (one for GB haddock and one for EGB haddock) are acceptable with respect to standard model diagnostics (although a fairly high retrospective pattern is seen for GB haddock). The two models do however represent two different hypotheses about the stock dynamics, and hence management based on the difference between the two model’s results should be carefully considered. It is recommended that long term a more consistent and possibly joint modelling approach is considered.

The haddock stock is data rich and dominated by occasional extremely large recruitment events (like many haddock stocks). The sampling effort is impressive and generally above NAFO recommended thresholds. Catch information gives a consistent picture of stock development and is consistent with the surveys. Jointly the catch observations and surveys included in the assessments appear to be sufficient to inform the assessment models.

Both models are implemented in the WHAM framework, which is an exciting and fairly newly developed state-space assessment model. It is developed at the Woods Hole lab (hence plenty of local expertise) and WHAM contains many of the features which are desirable to further develop the haddock assessment. WHAM has the ability to link environmental variables to key model parameters (e.g., natural mortality) and can include time varying components (e.g., selectivity) in response to a slowing growth pattern. Having the two models implemented in the same modelling framework should make it simpler to merge the two models long term.

The research presented on the recruitment process is interesting, because - like for many haddock stocks - the stock development is largely dictated by occasional extremely large recruitment events. Fall bloom correlates with Georges Bank haddock recruitment and egg retention (studied via drifters) is needed for successful recruitment events, but these things cannot be used to predict, because they cannot be predicted themselves. The work presented on stock identity, especially the MacCall basin model, seemed to support a more joint assessment approach. Overall, it seems that density-dependence could be included more directly in the assessment approach and possibly be used (long term) as a way to unite the two assessment approaches.

The previous research recommendations were well responded to by the research track and the new suggestions from the research track are important. Additional research recommendations to consider are suggested here (more details available under TOR 7):

- Resolve the conditional simulation bias issue discussed under TOR 4 for GB haddock (short)

- Merge the two assessment models into one joint assessment model (long term)
- Align modelling approaches - even if keeping the models apart (medium term)

In addition, this reviewer supports the extra research recommendations suggested in the panel's joint summary report.

Background

This report is the independent review report of the 2021-2022 Haddock Research Track Working Group (for Georges Bank and Eastern Georges Bank). The review meeting was held online (28-31 Apr. 2022). The meeting was well organized by the NEFSC's Stock Assessment Process Lead, Michele Traver, and by the chief of NEFSC's Population Dynamics Branch, Russell Brown. A special thanks to lead assessment scientists and presenters: Liz Brooks, Tom Carruthers, Brian Linton, Kevin Friedland, Monica Finley, Yanjun Wang, and Steve Cadrin who presented the work and answered numerous questions. Also, a very big thank you to the rapporteurs who helped the panel enormously. The reading material was provided approximately a week in advance and was very focused on addressing the TORs. The analyses were clearly described and of high quality and constitute an enormous amount of work. The panel was co-chaired by Richard Merrick (NOAA retired, New England Fisheries Management Council Scientific and Statistical Committee) and Allen (Rob) Kronlund (Interface Fisheries Consulting, Ltd.). The panel further consisted of CIE selected reviewers: Joseph Powers (NOAA retired), Kevin Stokes (Stokes.Net.NZ Ltd), and this reviewer, Anders Nielsen (Danish Technical University).

The research track has been ongoing since September 2020 and has had meetings >21 times, where the typical meeting has consisted of two half-day online sessions. The research track has worked on both Georges Bank and Gulf of Maine haddock stocks, but this meeting and review was focused on the Georges Bank and Eastern Georges Bank haddock work. The research track peer review is requested by the Northeast Region Coordinating Council.

Description of the individual reviewer's role

This reviewer has independently read the research track reports, and referenced documents deemed necessary in preparation for this review, participated in an online pre-meeting, participated actively in online research track review meetings (28-31 Mar. 2022), reviewed each individual term of reference, identified key issues, suggested guidance, contributed to the joint panel report, and independently authored this research track review report.

Findings for each TOR:

To ensure that all terms of reference are covered and that comments are interpreted with reference to the correct terms, the terms are listed (with gray highlighting) with corresponding reviewer comments following.

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.

The research track has met this TOR via the work done in the Friedland et al. 2020 paper. Here a lot of environmental variables (30+) are used to calibrate a random forest algorithm via present/absent observations from survey hauls in the period from 1976 to 2019. In the recent period the haddock habitat has increased (doubled its area in spring). The work has identified the variables which best correlate with haddock habitat (March chlorophyll concentration, average spring distribution of *Acartia* spp., December distribution of SST fronts, and fall distribution of *Centropages typicus*).

The random forest algorithm is interesting. Part of it works via cross validation, which means removing and predicting random observations from the dataset to select the variables best describing the observations. When using cross-validation, it is important to remove and predict independent observations (otherwise the correlation between observations interferes with the prediction measure). Spatial observations are often spatially correlated, so a further development of this could consider sampling larger spatial blocks of observations for the cross validation.

A further habitat study using the VAST program was conducted, which showed a similar expansion of the habitat in recent years. The VAST analysis focused on density, temperature, and average age and showed that higher temperatures could be driving the northward shift in the population center.

It was clear from both analyses that the habitat is expanding when the biomass is high. There is a part of GB area which is the preferred and most suitable habitat (in the northeast), but the habitat is expanded in periods after one or more big recruitment events (Figure 1).

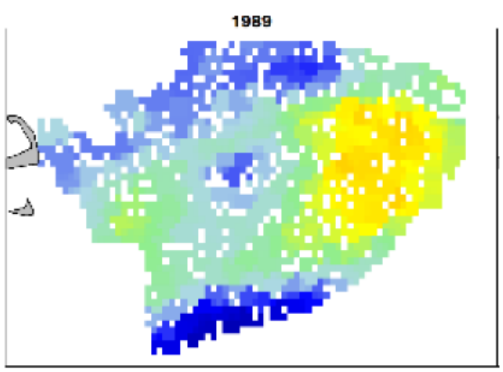

Figure 1: (parts of a presentation): Log scale spring distribution in a low biomass year (1989) compared to distribution in a high biomass year 2019.

These studies were used to support the EGB haddock assessment team's decision to include deeper strata. There is a unique potential possibility to use these spatial analyses further in the assessment modelling. This is possible because both the GB and EGB haddock assessments are implemented in the Woods Hole Assessment Model (WHAM (Stock & Miller 2021)) package and in WHAM it is possible to link certain quantities to environmental variables. For instance, it would be possible to link a measure of habitat to natural mortality (M). It would seem plausible, that in the periods where the haddock stock is pressured (by large stock size) to occupy parts of the area that are less suited for haddock, then the overall mortality would be larger.

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

The research track has met this TOR for both the GB and EGB haddock assessments and the catch data are of high quality with respect to length of time series, sampling intensity, and consistency. Furthermore, the occasional very high recruitment events give high contrast between cohorts, which gives an additional validation of the age reading consistency and the entire data collection and processing.

Catches generally consist of commercial landings, discards, and recreational catches, but recreational catches are so small (due to the offshore location) that they can be ignored. Discards are generally not a large part of the catch and recent Canadian efforts have further reduced them. In certain years the US discards are higher, which coincides with large cohorts entering the fishery. The observer coverage provides a reliable estimate of the discard part of the catch (cv<20% in recent years).

Haddock are landed in gutted condition, so a conversion factor is used to convert weight of gutted to weight of whole fish. The conversion factor dates back to the 1930s, and it is recommended to investigate if this can be improved — especially since over the last 20 years or so a decline in weight and length has been observed in all age classes (Figure B42 and B43). This decline in weight and length is attributed largely to density-dependent growth, but the decline appears very consistent so possibly other factors are also influencing this pattern.

The intensity of age and length sampling is high and generally above NAFO recommended thresholds (e.g., 100 lengths per 200mt landed) (Figure B34 and B35)[1]. The quality of the age-readings is evaluated by comparison studies both within and across labs and the result is that the age readings are very precise[2].

The research track lists a number of sources of uncertainties that could influence the estimated catches (e.g., conversion factor, area allocation, lower observer coverage in certain parts). These are commonly seen in many assessments. Under/over reporting of landings is also commonly suspected, but it appears that this suspicion may be more substantiated for haddock, because a recent court case has documented mis-labeling other species as haddock due to restrictive quotas on the other species.

Overall, the catches have been estimated and described in sufficient detail (spatially and temporally) to confidently use them to inform assessment models to assess the two stocks.

1 In the report available at time of review the two figures B34 and B35 were identical due to a copying mistake. The missing figure will be substituted into the report.

2 At the top of page 18 in the Georges Bank report is a small typo where a cv is listed as 0.42, when it is in fact a cv of 0.42%. This will be corrected in future versions.

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.

The research track has met this term of reference for both GB and EGB. The selected survey time series are long, largely cover the stock area, and are demonstrated to be very consistent.

Three surveys are selected to be used (partly) in both assessments. The two Northeast Fisheries Science Center (NEFSC) bottom trawl surveys (fall and spring) and the Canadian Department of Fisheries and Oceans (DFO) bottom trawl survey. It was also considered to include three others.

a. The NEFSC bottom long line survey, which would potentially be a valuable addition to have a bottom long line survey, because it would cover bottom types that the bottom trawl surveys do not cover, but the NEFSC bottom long line survey only covers a fraction of the stock area, so the decision was made not to use it.
b. The Massachusetts Division of Marine Fisheries (MADMF) bottom trawl survey was similarly excluded, because it only covers inshore areas.
c. The final index considered was a landing per unit effort LPUE index, which was excluded because of unknown effect of different management history and inconsistency with already included indices.

The NEFSC bottom trawl transitioned to a new vessel in 2009. Based on paired tows conducted in 2008 calibration factors were derived. The calibration factors were updated by the research track to be based on a length-based calibration derived in Miller (2013). Splitting the indices at the time of the vessel change was also investigated, but it turned out that the model results became unstable, and the retrospective pattern got worse.

The survey showed a temporal change in weight and length at age over the most recent 20-year period. This largely coincides with large cohorts passing through the population, so can largely be attributed to density dependent growth. It seems plausible that natural mortality could also be influenced (e.g., from more competition for resources or from the part of the population being displaced into areas which are less habitable for haddock). This was a point where the GB haddock and the EGB haddock parts of the research track did not come to the same conclusion. The GB team noticed that catches of 19-year-old haddock are consistent with a natural mortality not larger than 0.2, where the EGB team calculated total mortality from the surveys and concluded that the calculation indicated that the overall natural mortality had increased substantially in the recent period. Possibly both observations could be explained by the natural mortality being spatially different, such that the average natural mortality has in fact increased as a larger part of the stock has been displaced into areas not preferred by haddock in periods where the stock size is low or average. At the same time, it is possible that the part of the population, which remains in the most preferred area, is still only subjected to a natural mortality rate of about 0.2.

The decisions to include or exclude survey indices and to keep the calibrated indices combined (instead of split) are well supported. The selected surveys appear very consistent (both internally when comparing indices).

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.

The research track has met this TOR for both GB haddock and EGB haddock, but with minor reservations concerning compatibility of the two model approaches and concerning a single component of the model validation.

The research track has implemented both the GB haddock and EGB haddock assessments in the Woods Hole Assessment Model (WHAM (Stock & Miller 2021)) package. WHAM is a relatively newly developed assessment package. It has been accepted for assessment of Atlantic butterfish, and other than for GB/EGB haddock, it is being used in the ongoing research track assessments of American plaice, Black sea bass, and Atlantic cod. The WHAM package is part of the NOAA's toolbox (https://nmfs-fish-tools.github.io/) and developed with the source code openly available at github (https://github.com/timjmiller/wham/). WHAM was developed at the Woods Hole lab, so there is plenty of in-house expertise to support the model. All of these factors strengthen confidence in the model. The model is a state-space model and as such allows for both process and observation noise.

GB haddock:

The research track presented an impressively detailed step-by-step model search, which started at the VPA model selected at the last benchmark. At the last benchmark (in 2008) the retrospective diagnostic for the VPA model was unproblematic, but since then the retrospective pattern has increased to a problematic level. The first part of the model search was to update the VPA model to be based on the data decisions taken (including the updated length-based calibration factor). The retrospective pattern was reduced, but it was still at a high level.

The model search continued by configuring a similar (to the extent possible) model in the ASAP assessment package (Legault & Restrepo 1999). ASAP is a statistical catch-at-age model which means that it has different assumptions e.g., with respect to observation noise and selectivity pattern. It was possible to produce assessments in ASAP with similar historic estimates as VPA of important outputs (SSB, average F, and recruitment), but the estimates diverged in the last 15-20 years. Interestingly, the ASAP model did not have a retrospective problem. Different configurations were explored within ASAP in terms of selectivity assumptions (blocking). The general conclusion was that more blocks led to lower AIC, but increased retrospective pattern.

The next step in the model search was to move the ASAP model into the WHAM framework. WHAM is designed to read the same data format and to possibly be configured in almost exactly the same way as ASAP, so this produced negligible differences. Following this many different features of WHAM were explored, and the search ended in a model with: natural mortality fixed at 0.2, separable AR process for survival, constant selectivity before 1963, and separable AR process thereafter.

The entire model search was long, well-structured, and well-argued. The selected base model and many of the alternatively configured models were explored by state-of-art model diagnostics (including focus on prediction ability and composition residuals with correct statistical properties). While most diagnostics were acceptable, two things need to be noticed. The selected model did have a concerning retrospective pattern (Figure 2), which appears to be reducing in the last year (Table 21), but which could increase in the following years.

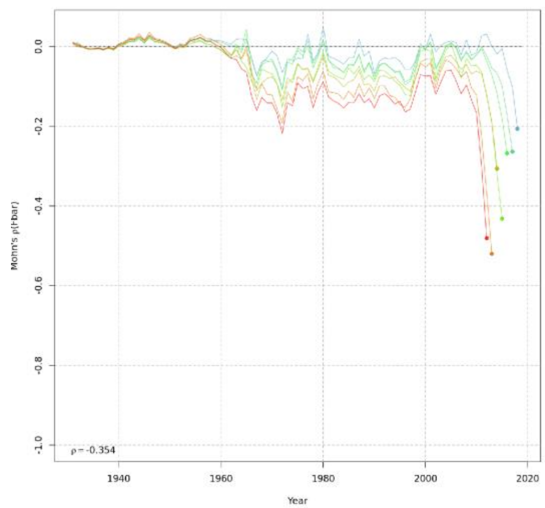

Figure 2: Mohn's rho for average F and spawning stock biomass for the model proposed for GB Haddock.

The other issue is the simulation validation of the model (Figure B134). In this figure, two different simulation schemes are compared. In the first simulation scheme the full model is simulated from only the fixed effect model parameters. This means among other things that for each simulation a time series of the stock sizes (Ns and thereby SSBs) is simulated and then one set of observations are simulated correspondingly. Then for each simulation, the model is used to estimate, and the simulated truth is compared to the estimated. This unconditional self-simulation test is natural for a state-space model and it appears largely unbiased.

The second simulation test is a so-called conditional simulation (Figure B134, right side and figure 3). In this test the estimated time series e.g., the stock sizes (and thereby SSBs) are kept at the estimated values and then a number of datasets are generated by simulating new unbiased observations corresponding to the one estimated time series. This test is not strictly a self-simulation of the state-space model, because the one estimated time series must be expected to have some smoothing bias, and hence it should be interpreted carefully. It does however answer the important question: "If the true system did in fact develop as we have estimated it, then what kind of bias can we expect when estimating via this model?". Such an estimation should not be expected to be perfectly unbiased, because of smoothing bias which is a function of the signal to noise ratio (if we have a signal with a lot of noise then the best we can hope for is to estimate the overall trends and not each detailed fluctuation). Nevertheless, it is an important diagnostic that should provide a reasonably low bias no matter what type of assessment model is being proposed.

The results presented for the conditional simulation are shown for SSB in Figure 3 (left panel). It seems to indicate that a consistent 35-40% bias can be expected in SSB estimates from 1964 until 2019. This would potentially be an important and problematic issue, because it shows that on average and in most of the simulated cases similar to the actual time series of haddock (as agreed by VPA, ASAP, and WHAM figure 3 right panel) a bias of 35-40% should be expected on the estimates of SSB provided by the model. This appears to be the case even in the periods where all the models agree (e.g., the 1980s and the 1990s, Figure 3).

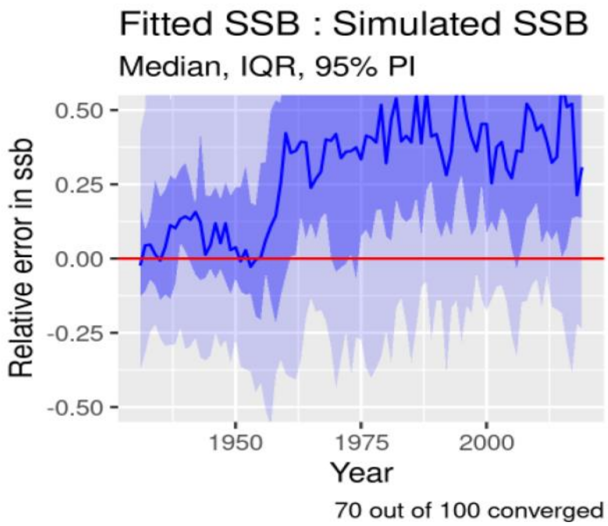

Figure 3: Relative expected estimation error in spawning stock biomass under conditional simulation (left part of figure B134). Spawning stock biomass time series estimated from 4 different models (right part of figure B135).

Fortunately, it appears most likely that the problem is not with the estimation model, but with the simulated data sets. This is because for most of the time-series VPA, ASAP, and WHAM show pretty much the same (Figure 3, right panel) for SSB and F, so if WHAM really had a 35-40% bias, then so should the VPA and ASAP, which is not likely, as such a large issue would have been discovered a long time ago for these models. Furthermore, a last-minute modification to the WHAM model was introduced to avoid boundary issues in the selection pattern, which included some rescaling of the catchabilities. I speculate that the issue is caused by a mismatch between the scale of the simulated survey observations and the actual survey observations, which is also consistent with the timing of the observed bias. Problematic simulations are a much less important issue than a large consistent bias in SSB, but it is recommended that the cause of the apparent bias be resolved.

EGB haddock:

The research track also presented a detailed and well-argued model search for the EGB haddock stock component. The assessment challenges for GB and EGB haddock are very similar in terms of data availability and emerging problems. Furthermore, all model explorations for EGB haddock were conducted within the WHAM package, which should make it simpler to compare — and possibly even combine — the approaches long term. Currently the approaches differ substantially.

The initial model configurations and comparative VPA assessments for EGB haddock showed that the retrospective problem was even larger for this stock component than it was for GB haddock (which is not to say that it was small for GB haddock). Mohn's rho of the retrospective pattern for many configurations were around 100% for biomass and 50% for fishing mortality, which is very large (e.g., compared to the rule of thumb of 20% which is frequently used by ICES to disqualify assessments). The model search focused on reducing this retrospective pattern.

Many different model configurations were explored (e.g., different time invariant fixed natural mortalities, different selection blocks, different assumed distributions for the composition observations, and excluding sub-sets of observations), but the retrospective pattern remained. The only model configurations that reduced the retrospective pattern to an acceptable level were model configurations allowing for a time varying natural mortality.

Time varying natural mortality does seem reasonable for EGB haddock. As seen under TOR 1, occasional huge recruitment events expand the haddock habitat into areas not preferred by haddock under low or normal abundance conditions. It seems reasonable that a higher natural mortality could be experienced in the less preferred areas. Furthermore, higher competition for food also seems likely under high abundance conditions, which is further supported by the observed decline in weight and length at age in the most recent period with high abundance (TOR 3).

Four different approaches to include time varying natural mortality were tested (age-invariant AR, separable AR survival, age-invariant externally assigned doubling of natural mortality in the last 10 years, and age-invariant estimated natural mortality in the last 10 years). These four options all reduced the retrospective pattern and the options where yearly mortalities/survivals were estimated somewhat supported the selection of a 10-year window of increased natural mortality. The option to estimate a single age-invariant natural mortality for the last 10 years was selected based on simplicity, best retrospective pattern, and simulation testing.

The proposed model does solve the retrospective problem, and it seems consistent with other observations: Empirical total mortality calculation under TOR 3, habitat expansion TOR 1, and decrease in weight and length at age TOR 3. The proposed model does also have a few issues to consider: The abrupt change in natural mortality from one year to the next may seem unrealistic; the somewhat arbitrary selection of the period (which the model is fairly sensitive to); and selection of what to do when more data years become available (continue to estimate, revert to baseline natural mortality of 0.2, or something else).

GB/EGB Haddock

In isolation both models give acceptable assessments of their respective stock (component), but it seems a bit artificial to consider them as separate assessments. Especially when the management is conducted by considering them both. The very different assumptions concerning natural mortality in the last decade do make the estimates of e.g., stock size and fishing pressure less comparable. The long-term recommendation should be to develop a more joint assessment approach.

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.

The research track has met this TOR for GB haddock and this TOR is not applicable for EGB (where TOR 11 is replacing it).

GB haddock:

Reference points can be calculated within the WHAM package. This is an advantage because it is model consistent, and it propagates the uncertainty in the estimated model parameters to the calculated reference points. The research track proposed to use MSY proxies (SPR40%) as calculated within WHAM, which is different compared to the long-term projections used with the previous VPA model.

To reflect "prevailing conditions" it was proposed to use 5-year averages for stock weights, catch weights, and selectivity. To reduce undue influence of recent extremely high recruitment events it was proposed to use the full time series of estimated recruitments.

The proposed approach seems reasonable with respect to the use of proxies given the sporadic stock-recruitment pattern. As per the year ranges chosen, the relative short 5-year ranges reflect current conditions well, and hence would be well chosen for short term projections. If density-dependence is really the driving force behind the lower weights-at-age, then we may expect the weights to increase back to previous levels and then a longer average may better approximate prevailing (long term) conditions. As mentioned by the research track: recalculation of such reference points and reexamination of the assumptions occur frequently.

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.

The research track has met this TOR for both GB and EGB haddock.

GB haddock:

The research track used evaluations of prediction skills of previous assessment estimates to select the best number of years to average over with respect to weight in the short-term projection. This approach favors 2-year averages for weights. For selectivity at age a 5-year average is chosen consistent with the choice under TOR 5. The recruitment projection is taking advantage of the state-space assessment model, because it can be projected according to the process model assumed for the assessment. This consistency between the assessment model and the forecasts is a nice feature of the assessment, and with further developments in state-space assessment models could be extended to weights and selectivity also.

EGB haddock:

The assessment model was transferred to a forward simulation model (and validated). The uncertainties of the estimated model parameters are included via simulation in the forecast (as estimated via the inverse Hessian). Based on predictive analysis, a 3-year average is used for weights. The estimated average and assumed deviances are used to simulate recruitment forward. Finally, a 3-year average is used for selectivity.

The main difficulty when projecting the EGB haddock model is the assumption about natural mortality (M). Here a low M=0.2 scenario and a high M scenario (where M is set to the estimated value in the last 10 years (M=0.473)) is conducted.

The projections are very sensitive to the M scenario, and it is unclear what is the natural choice for short term projections. Since it is short term, it could be argued that M should stay high as estimated in recent years, but if M is connected to density, then M should be adjusted lower when the density becomes lower. The final choice of an M-scenario should be based on most recent information at the time where the short-term forecast is conducted. A long-term research recommendation could be made to study the relationship between density and natural mortality, so the choice of M-scenario could be less ad-hoc.

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.

The research track has met this TOR by addressing all the previous recommendations well and suggesting a long list of new ones.

Of the suggested new ones I would like to offer additional support to the ones that aim to harmonize the US and the Canadian data collection e.g., weight data and maturity stage classification. Further I offer additional support to any research into identifying factors influencing natural mortality, since the natural mortality assumption is the main inconsistency between the two modelling approaches.

In addition, I would offer the following research recommendations:

- Any efforts into merging the two assessment models into one joint assessment model, or even just two linked assessment models. Running these as two independent assessments seems problematic when the management needs to combine them. If a completely joint model is not feasible, then it could be considered to join them by correlating in certain parts of the model. An example of such correlated state-space assessment models is in Albertsen et al. (2018). This is likely a long-term recommendation.

- Aligning the modelling approaches. Even while keeping the models apart, it would be beneficial if the two modelling teams considered a compromise with respect to key assumptions (e.g., w.r.t. M), such that they could provide comparable estimates.

- Resolving the conditional simulation bias issue discussed under TOR 4 for GB haddock. As mentioned, it is most likely a simulation issue, but resolving it should be a short-term recommendation.

- There seems to be a lot of support for density-dependent effects, and it seems that it could be used more directly in the assessment model (e.g., linked to M).

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

The research track has met this TOR by suggesting the so-called "Plan B smooth" for GB haddock and "I2" approach for EGB haddock.

GB haddock:

The suggested plan B "Plan B smooth" (https://github.com/cmlegault/PlanBsmooth) has been the proposed backup method since 2017. It has never been used for GB haddock and it seems unlikely that it would for such a data rich stock. Plan B smooth has been investigated in a separate research track on Index Based Methods and Control Rules (IBMWG) and it was concluded that the Plan B smooth performed as well as any other index-based method. Briefly, the method works by smoothing average survey biomasses, then fitting a straight line (log scale) through the last three points, finally using the exp(slope) as the catch multiplier. It was suggested to base the method on the two NEFSC surveys and to use a smoothing span of 0.27 (derived via experimentation and considering retrospective performance).

EGB haddock:

Several index-based management procedures were considered for EGB haddock ("Plan B smooth", "I2" which is based on the mean index relative to the previous two years, and additionally proposed "ACI" age composition indices). After extensive simulation evaluation, the "I2" procedure was considered to perform best overall and most similar to 40%SPR.

With respect to the plan B methods proposed, which are calibrated based on the analytical assessment model, then it should be considered if calibration is still valid if the analytical assessment models are "failing", which is the situation when the plan B is to be applied.

GB/EGB haddock:

As mentioned above, it does seem unlikely that any of these (very crude) plan B approaches would be applied for a data rich stock like GB/EGB haddock. Even in the event that it were to happen, then it would not be for many years before an improved analytical approach would be suggested. Hence the long-term performance simulations are likely only theoretically interesting.

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.

This research track has met this TOR by considering studies on fall bloom and on egg retention.

The recruitment process is one of the most interesting parts of haddock assessments; the stock development is characterized by some extremely large recruitment events and the recruitment is to some extent determining the habitat. Hence it will be very useful and interesting if any information can be identified, which can explain (and hopefully predict) these recruitment events.

The research track looked at fall bloom. A previous analysis with only seven points showed a correlation for GB haddock. The analysis was updated and confirmed for GB haddock. Importantly the bloom magnitude was the highest observed for the very big 2013 cohort. The fall bloom appears to be part of explaining large cohorts.

The research track also considered drift and retention of eggs via drifter studies. The egg retention via drifter model showed that years of large cohorts 2003, 2010, and 2013 did have high retention, but so did some years with low recruitment. It was concluded that retention is necessary, but other factors are also needed.

The research reviewed was very interesting and explained important parts of the recruitment process, but at the same time it is difficult to use these things to predict recruitment, because fall bloom and egg retention are themself processes which are difficult to predict for the time horizon typically needed for assessment and management.

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

This research track has met this TOR by an analysis of growth curves against time and an analysis of growth residuals against different covariates (e.g., density, temperature, salinity and geographical location).

A von Bertalanffy growth curve with cohort specific random effects for the three model parameters (t_0, k, and L_∞) was fitted to more than 20,000 age-length samples from the DFO survey. The model with the cohort specific random effects provided a significantly better description of the observations and the growth was lower in the later period where the density is high.

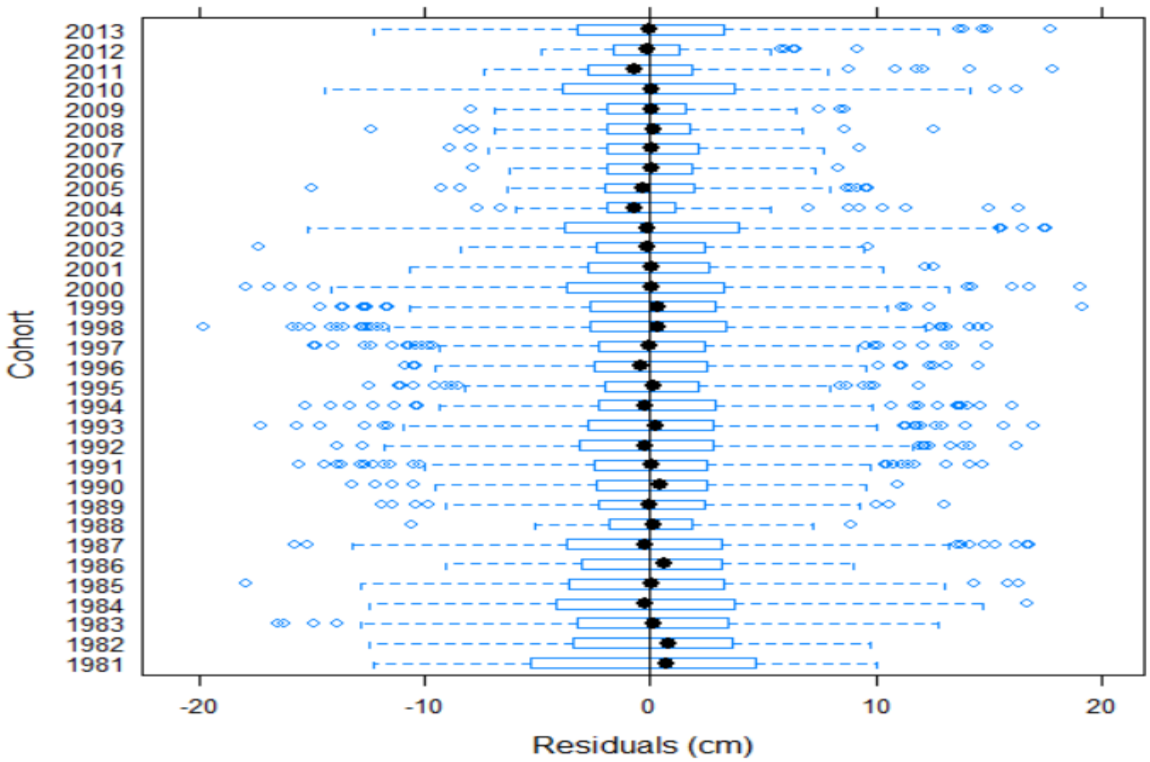

Figure 4: Residuals by cohort for the von Bertalanffy growth curve with cohort specific model parameters.

An interesting observation from the model diagnostics is that if we consider the last period (say after 2000), then the three very large cohorts (recruited in 2003, 2010, and 2013) are also the cohorts where the distribution of the residuals is the widest (Figure 4). This further supports the idea that when large recruitment events occur, a part of the cohort is displaced to areas which don't support haddock growth as well, while a part of the cohort remains in areas which do. This will lead to the mean effect of density dependent depressed growth, but also to a larger variation in growth, because part of the cohort continues to grow at normal levels.

If a similar thing is happening with respect to natural mortality, then it can explain the observation that the EGB haddock assessment sees a general increase in natural mortality while at the same time the GB haddock team sees 19+ year old fish; this indicates that the mortality is no greater than M=0.2. It does not seem unreasonable that depressed growth could correlate with higher mortality.

In the second analysis a year-invariant spline growth curve is fitted to the age-length observations. It would seem to have been more consistent to have used the residuals from the time invariant von Bertalanffy model, but the difference is likely minimal. The residuals are described via a number of covariates (density, sst, bst, salinity, phytoplankton bloom, lon-x-lat) in a general additive model. Density-dependence was identified as the most important factor, but temperature also had an effect.

11) For Eastern Georges Bank, provide advice to TMGC on appropriate reference points.

The research track has met this TOR by demonstrating a procedure for evaluating reference points.

The current F-reference point F_{ref}=0.26 was derived in 2002 and based on 40%SPR. The growth and thereby the selectivity and possibly the natural mortality has changed since, so it should be investigated if the reference point is still appropriate.

The proposed approach first turns the assessment into a forward simulation model (and validates it), then the resulting yields and ssbs were compared when simulating forward from 1987 with the reference point updated every 2nd, 3rd, 4th, 5th, 6th, or 7th year and based on averages of the most recent 5, 10, or 15 years.

If the high natural mortality scenarios are used, then the resulting reference levels become unreasonably high, so only the low M scenarios were explored further.

The different scenarios were fairly robust in terms of the selection of update intervals and average years. A 10-year average that is updated every 4 years is proposed, as that combination gave the highest ssbs and among the highest yields. Using that definition resulted in a current reference point of $F_{ref(40\%SPR)}$=0.488.

The procedure demonstrated should provide valuable advice to TMGC for selecting appropriate reference points and could easily be expanded to include other alternatives.

12) Review data related to stock structure of haddock on Georges Bank (including Eastern Georges Bank management area) and implications for assessments conducted on the whole bank and on subareas of the bank.

The research track has met this TOR via a well-structured document and presentation on knowledge about GB haddock stock structure. The GB haddock stock structure is complicated and the non-consensus regarding stock structure is less related to the perception of the stock structure and more concerned with the treatment within assessment models.

The MacCall basin model seems to be a useful way to summarize the stock structure and dynamics. At low abundance levels, the stock stays in the optimal habitat, which means that parts of the stock do not mix. At intermediate abundance levels, part of the stock is displaced into sub-optimal habitat and the mixing is greater, because the sub-stocks which were isolated before are now connected via the sub-optimal habitat. At high abundance levels (e.g., after a few extremely large recruitment events), part of the stock is pushed to sub-optimal and even to marginal habitat and there is high mixing. When the stock is pushed into marginal habitat it seems plausible that average growth, maturity, and mortality can be affected while, at the same time, parts of the stock which manage to remain in the optimal habitat remain largely unaffected.

It was concluded that there are no persistent stock boundaries for GB haddock and that boundaries which were defined during low abundance levels may need to be reconsidered. Current conditions suggest that haddock on GB is one single stock.

The recruitment time series is highly correlated and the length distribution suggests a well-mixed single stock. It would seem that assessing it only as a single stock would be beneficial (possibly augmented with a spatial analysis of the distribution to assist management).

Comments on the review process:

The review meeting was efficient and well organized. However, having an assessment review online is not a good substitute for an actual review meeting. The discussion is slower, and hence fewer issues are raised. It is also not possible to stand up and make an illustrative drawing where needed. Furthermore, the sharing of knowledge, which for other review meetings has been substantial (e.g., sharing tips and tricks of modelling, or introduction to new tools or software), does not happen when all breaks are in isolation. Having informal discussions in person is much better for networking between assessment panels and reviewers, and overall makes the physical meetings more productive.

References:

Albertsen, CA, Nielsen A, and Thygesen UH, 2018. Connecting single-stock assessment models through correlated survival. ICES Journal of Marine Science. 75: 234-244.

Friedland, KD, et al. 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.

Legault, CM and Restrepo, V. 1999. A Flexible Forward Age-Structured Assessment Program. ICCAT Coll. Vol. Sci. Pap. 49.

Miller, TJ. 2013. A comparison of hierarchical models for relative catch efficiency based on paired-gear data for US Northwest Atlantic fish stocks. Canadian Journal of Fisheries and Aquatic Sciences 70 (9), 1306-1316.

Stock, BC., and Miller, TJ. 2021. The Woods Hole Assessment Model (WHAM): A general state-space assessment framework that incorporates time- and age-varying processes via random effects and links to environmental covariates. Fisheries Research, 240: 105967. doi: 10.1016/j.fishres.2021.105967

Appendix 1: Bibliography of materials provided for review

Brooks, E.N, M.L. Traver, S.J. Sutherland, L. Van Eeckhaute, and L. Col. 2008. In. Northeast Fisheries Science Center. 2008. Assessment of 19 Northeast Groundfish Stocks through 2007: Report of the 3rd Groundfish Assessment Review Meeting (GARM III), Northeast Fisheries Science Center, Woods Hole, Massachusetts, August 4-8, 2008. US Dep Commer, NOAA Fisheries, Northeast Fish Sci Cent Ref Doc. 08-15; 884 p + xvii. http://www.nefsc.noaa.gov/publications/crd/crd0815/

Friedland, K.D., J.A. Langan, S.I. Large, R.L. Selden, J.S. Link, R.A Watson, and J.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.

Hordyk, A.R., Q. Huynh, and T.R. Carruthers. 2022. *OpenMSE R Package* (1.0.0) [Computer software]. Blue Matter Science. https://cran.r-project.org/package=openMSE.

Legault, C.M. and V.R. Restrepo. 1998. A flexible forward age-structured assessment program. ICCAT. Col. Vol. Sci. Pap. 49: 246–253.

NEFSC (Northeast Fisheries Science Center). 2017. Operational Assessment of 19 Northeast Groundfish Stocks, Updated Through 2016. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 17-17; 259p. doi: 10.7289/V5/RD-NEFSC-17-17.

NEFSC (Northeast Fisheries Science Center). 2019. Georges Bank haddock 2019 assessment update report. Unpubl. Rpt. 10 pp. nefsc.fisheries.noaa.gov/saw/sasi/sasi_report_options.php

NEFSC (Northeast Fisheries Science Center). In Review. Final Report of the Haddock Research Track Assessment Working Group. Unpubl. Rpt. 65 pp. https://apps-nefsc.fisheries.noaa.gov/saw/sasi/sasi_report_options.php

Stock, B.C., and T. J. Miller. 2021. The Woods Hole Assessment Model (WHAM): A general state-space assessment framework that incorporates time- and age-varying processes via random effects and links to environmental covariates. Fisheries Research 240, 105967. https://doi.org/10.1016/j.fishres.2021.105967

TRAC (Transboundary Resources Assessment Committee). 2019. Eastern Georges Bank haddock [5Zjm; 551,552,561,562]. TRAC Status Report 2019/01.

TRAC (Transboundary Resources Assessment Committee). 2021. Eastern Georges Bank haddock [5Zjm; 551,552,561,562]. TRAC Status Report 2021/01.

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

Eastern Georges Bank and Georges Bank Haddock
Research Track Peer Review

Background

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

Scientific peer review is defined as the organized review process where one or more qualified experts review scientific information to ensure quality and credibility. These expert(s) must conduct their peer review impartially, objectively, and without conflicts of interest. Each reviewer must also be independent from the development of the science, without influence from any position that the agency or constituent groups may have. Furthermore, the Office of Management and Budget (OMB), authorized by the Information Quality Act, requires all federal agencies to conduct peer reviews of highly influential and controversial science before dissemination, and that peer reviewers must be deemed qualified based on the OMB Peer Review Bulletin standards[3]. 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

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

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 Eastern Georges Bank and Georges Bank and haddock stocks. The requirements for the peer review follow. This Performance Work Statement (PWS) also includes: **Appendix 1**: TORs for the research track, which are the responsibility of the analysts; **Appendix 2:** a draft meeting agenda; **Appendix 3:** Individual Independent Review Report Requirements; and **Appendix 4:** Peer Reviewer Summary Report Requirements.

Requirements

NMFS requires three reviewers under this contract (i.e. subject to CIE standards for reviewers) to participate in the panel review. The chair, who is in addition to the three reviewers, will be provided by either the New England or Mid-Atlantic Fishery Management Council's Science and Statistical Committee; although the chair will be participating in this review, the chair's participation (i.e. labor and travel) is not covered by this contract.

Each reviewer will write an individual review report in accordance with the PWS, OMB Guidelines, and the TORs below. 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, 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, <u>and</u> 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 June, 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
April 4 – April 7, 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 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.

11. For Eastern Georges Bank, provide advice to TMGC on appropriate reference points.

12. Review data related to stock structure of haddock on Georges Bank (including Eastern Georges Bank management area) and implications for assessments conducted on the whole bank and on subareas of the bank.

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. Draft Review Meeting Agenda
{Final Meeting agenda to be provided at time of award}

Eastern Georges Bank and Georges Bank Haddock
Research Track Assessment Peer Review Meeting

April 4 – April 7, 2022

WebEx link: TBD Phone: TBD

DRAFT AGENDA* (v. 1/6/2022)
**All times are approximate, and may be changed at the discretion of the Peer Review Panel chair. The meeting is open to the public; however, during the Report Writing sessions we ask that the public refrain from engaging in discussion with the Peer Review Panel.*

Monday, April 4, 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 TBD Panel Chair	
9:30 a.m. - 10:30 a.m.	TOR #2	Liz Brooks, Monica Finley	Catch data
10:30 a.m. - 10:45 a.m.	Break		
10:45 a.m. - 11:45 a.m.	TOR #2 cont.	Liz Brooks, Monica Finley	Catch data
11:45 a.m. - 12:15 p.m.	Discussion/Summary	Review Panel	
12:15 p.m. - 12:30 p.m.	Public Comment	Public	
12:30 p.m. - 1:30 p.m.	Lunch		
1:30 p.m. - 3 p.m.	TOR #3	Liz Brooks, Monica Finley	Survey data

3 p.m. - 3:15 p.m.	Break		
3:15 p.m. - 4:15 p.m.	TOR #3 cont.	Liz Brooks, Monica Finley	Survey data
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, April 5, 2022

Time	**Topic**	**Presenter(s)**	**Notes**
9 a.m. - 9:15 a.m.	Welcome/Logistics	Michele Traver, Assessment Process Lead TBD, Panel Chair	
9:15 a.m. - 10:30 a.m.	TORs #1 and #9	Kevin Friedland, Liz Brooks, Scott Large	Ecosystem and Recruitment Processes
10:30 a.m. - 10:45 a.m.	Break		
10:45 a.m. - 11:45 a.m.	TORs #1 and #9 cont.	Kevin Friedland, Liz Brooks, Scott Large	Ecosystem and Recruitment Processes
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.	TORs #10 and #12	Liz Brooks, Steve Cadrin, Yanjun Wang	Density Dependent Growth and Stock Structure
3 p.m. - 3:15 p.m.	Break		

3:15 p.m. - 4:15 p.m.	TORs #10 and #12 cont.	Liz Brooks, Steve Cadrin, Yanjun Wang	Density Dependent Growth and Stock Structure
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, April 6, 2022

Time	**Topic**	**Presenter(s)**	**Notes**
9 a.m. - 9:15 a.m.	Welcome/Logistics	Michele Traver, Assessment Process Lead TBD, Panel Chair	
9:15 a.m. - 10:30 a.m.	TOR #4	Liz Brooks, Tom Carruthers	Mortality, Recruitment and Biomass Estimates
10:30 a.m. - 10:45 a.m.	Break		
10:45 a.m. - 11:45 a.m.	TOR #4 cont.	Liz Brooks, Tom Carruthers	Mortality, Recruitment and Biomass Estimates
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.	TORs #5, #6, and #11	Liz Brooks, Tom Carruthers	BRPs, Projections and EGB Reference Points
3 p.m. - 3:15 p.m.	Break		
3:15 p.m. - 4:15 p.m.	TORs #5, #6, and #11 cont.	Liz Brooks, Tom Carruthers	BRPs, Projections and EGB Reference Points

4:15 p.m. - 4:45 p.m.	Discussion/Summary	Review Panel	
4:45 p.m. - 5 p.m.	Public Comment	Public	
5 p.m.	Adjourn		

Thursday, April 6, 2022

Time	**Topic**	**Presenter(s)**	**Notes**
9 a.m. - 9:15 a.m.	Welcome/Logistics	Michele Traver, Assessment Process Lead TBD, Panel Chair	
9:15 a.m. - 10:30 a.m.	TOR #8	Liz Brooks, Tom Carruthers	Alternative Assessment Approach
10:30 a.m. - 10:45 a.m.	Break		
10:45 a.m. - 11:45 a.m.	TOR #7	Brian Linton	Research Recommendations
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. - 2:30 p.m.	Follow-ups/Key Points	Review Panel	
2:30 p.m. - 5 p.m.	Report Writing	Review Panel	
5 p.m.	Adjourn		

Friday, April 7, 2022

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

Appendix 3. Individual Independent Peer Reviewer Report Requirements

1. The independent Peer Reviewer report shall be prefaced with an Executive Summary providing a concise summary of whether they accept or reject the work that they reviewed, with an explanation of their decision (strengths, weaknesses of the analyses, etc.).

2. The report must contain a background section, description of the individual reviewers' roles in the review activities, summary of findings for each TOR in which the weaknesses and strengths are described, and conclusions and recommendations in accordance with the TORs. The independent report shall be an independent peer review, and shall not simply repeat the contents of the Peer Reviewer Summary Report.

a. Reviewers should describe in their own words the review activities completed during the panel review meeting, including a concise summary of whether they accept or reject the work that they reviewed, and explain their decisions (strengths, weaknesses of the analyses, etc.), conclusions, and recommendations.

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

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

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

3. The report shall include the following appendices:

Appendix 1: Bibliography of materials provided for review
Appendix 2: A copy of this Performance Work Statement
Appendix 3: Panel membership or other pertinent information from the panel review meeting.

Appendix 4. Peer Reviewer Summary Report Requirements

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

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

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

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

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

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

Appendix 3: Panel membership and meeting attendance

NEFSC - Northeast Fisheries Science Center
GARFO - Greater Atlantic Regional Fisheries Office
NEFMC - New England Fisheries Management Council
DFO - Department of Fisheries and Oceans (Canada)
SMAST - University of Massachusetts School of Marine Science and Technology
MAMFI - Massachusetts Marine Fisheries Institute

~~~~~~

Richard Merrick - US Co-Chair
Allen (Rob) Kronlund - Canadian Co-Chair
Joe Powers - CIE Panel
Anders Nielsen - CIE Panel
Kevin Stokes - CIE Panel

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

Abigail Tyrell - NEFSC
Alain d'Entremont - Scotia Harvest Inc., TMGC Canadian co-chair
Alex Hansell - NEFSC
Alicia Miller - NEFSC
Angela Forristall - NEFMC Staff
Anthony Wood - NEFSC
Brian Linton - NEFSC
Cate O'Keefe - Fisheries Applications Consulting Team
Catriona Regnier-McKellar - DFO
Charles Adams - NEFSC
Charles Perretti - NEFSC
Irene Andrushchenko - DFO
Jamie Cournane - NEFMC Staff
Jason Boucher - NEFSC
Kathryn Cooper-MacDonald - DFO
Kathy Sosebee - NEFSC
Kevin Friedland - NEFSC
Kris Vascotto - Atlantic Groundfish Council, Executive Director
Larry Alade - NEFSC
Libby Etrie - NEFMC Member
Liz Brooks - NEFSC
Liz Sullivan - GARFO
Lottie Bennett - DFO
Mark Terceiro - NEFSC
Melanie Griffin - MAMFI
Michelle Greenlaw - DFO
~~~~~~

Mike Simpkins - NEFSC
Monica Finley - DFO
Paul Nitschke - NEFSC
Robin Frede - NEFMC Staff
Sarah Salois - NEFSC
Scott Large - NEFSC
Steve Cadrin - SMAST
Tara McIntyre - DFO
Tara Trinko Lake - NEFSC
Tim Miller - NEFSC
Tom Carruthers - Blue Matter Science, consultant for DFO
Tom Nies - NEFMC, Executive Director
Toni Chute - NEFSC
Yanjun Wang - DFO