Center for Independent Experts (CIE) Independent Peer Review of SARC 65 (Sea Scallops and Atlantic herring)

NEFSC, Woods Hole MA, 26-29 June 2018

Individual Independent Review Report

Dr Coby Needle, CIE Reviewer

August 2018

Executive Summary

This report summarises my notes and conclusions on the SARC 65 assessments of Sea Scallops and Atlantic herring, presented during the SARC 65 meeting at the Northeast Fisheries Science Center (NEFSC), Woods Hole MA, during 26-29 June 2018. While there were areas of both assessment reports that I felt could have been presented with more clarity, all relevant questions were addressed in full by the stock assessors during the review meeting, and I concluded (in agreement with the others in the SARC panel) that the ToRs had all been met to the extent possible and necessary for the assessment. The assessments were well-presented by the lead scientists and SAW working group chairs, contained a great deal of relevant information, and I was happy to accept them as a valid representation of 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. Overall, I would like to commend the NEFSC teams for their open, accommodating and non-defensive discussion of the material, and for all the help they gave me and my fellow reviewers during the process. The meeting was conducted with a great deal of tact and good humour under what might otherwise have become difficult circumstances, and there is a great deal that other parts of the world can learn about stakeholder engagement from the SARC process.

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 have led the Sea Fisheries Programme of Marine Scotland (around 60 staff) for the past three years: part of the Scottish Government based at the Marine Laboratory in Aberdeen, Scotland. One of our key roles is the collection, analysis and interpretation of data from the fishing industry and fishery-independent surveys, and the provision of advice on fisheries and fish stocks to managers in Scotland, the UK and Europe. As well as being the principal provider of advice on demersal stocks in particular to the Scottish Government, I also still lead on the ICES assessments of North Sea lemon sole, and contribute significantly to other North Sea stocks through the relevant ICES Working Group. My main research areas lie in the development of fishery-independent assessment and advice methods, remote electronic monitoring of fishery activity, management-strategy evaluations, and the modelling of fleet dynamics.

Role in the Review Activities

Prior to the SARC 65 meeting in Woods Hole (26-29 June 2018), 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 sea scallops and Atlantic herring (see References below for a document list). During the SARC 65 meeting, I participated in full in the plenary discussions during and after the presentations provided, as well as intersessionally with the other review panel members and the SARC chair. I took copious notes during these discussions which form the basis of my comments below. Finally, I contributed in full to the writing of the SARC Summary Report, and wrote this Individual Independent Review Report.

Overall Conclusions and Suggestions for Improvements to the SARC Process

In general, I found the process to be very well-organised and conducive to the reviews. The NEFSC teams were extremely helpful and responsive to our requests, and presented the results of their hard work in an open, non-defensive manner and with good humour throughout. For the most part, the stakeholder participants were also even-handed and fair with their comments and questions – there

was perhaps only one point of real contention (the use of gonad weights to derive SSB in sea scallops), but that was dealt with appropriately and in good time before the end of the week, and I think the stakeholder's concerns in that case were addressed in full.

The assessment reports were quite well laid out overall - I found the LaTeX-generated layout of the sea scallop stock assessment report in particular to be excellent, and I would always recommend the use of this document markup system for reports of this kind (Lamport 1994). However, there are several places in the documents where essential details were missed, the figure and table captions were inadequate for the reader to understand fully what was being presented, the order of figures or comments become somewhat confusing, some of the figures were presented but not referred to in the text, and in places the treatment became slightly cursory. For example, the underlying principles and specific implementations of both the CASA and the ASAP models could have been explained in more detail and clarity. The brevity that was shown in some points in the reports would be fine in an update assessment for which a detailed stock annex exists, but in these cases these documents are all there is so they do need to be complete.

I would also recommend strongly that the figures are cited in the text in the same order that they appear in the report, as this is a real help to the reviewer. That this isn't the case with the LaTeX-generated sea scallop report is particularly strange, as LaTeX will do that automatically. It also should be said that several of the Figures (particularly in the sea scallops section) were too small, with almost illegible text. Finally, I should note that two of the reports' appendices were not available to the reviewers until very late in the week prior to the meeting, which made reviewing the relevant aspects more difficult.

Comments regarding each ToR: Sea scallops

Introduction

The opening section of the report gave a good and comprehensive overview of stock distribution, recent population and assessment history, and life-history parameters to be used in the assessment. For the most part this was clear and very useful, although I did raise some points with the stock assessor that required clarification.

The stock distribution indicates that there may be more than two discrete stock areas, and I asked whether consideration had been given to assessing the population as more than two separate stocks (as we would with *Nephrops* Functional Units in Europe), given that biological linkages between areas may be small. This has not been done, or considered very deeply, but it may be an issue to raise at the next benchmark. A related point pertains to the stock in Canadian waters, which is not spatially separated from the contiguous US stock (there is evidence of larval exchange), and which is likely to be part of the same population. This area is not yet included in the assessment, but could be in the future. Similarly, the dense population of slow-growing (non-fished) scallops near Nantucket Lighthouse are not yet included in the assessment at all, but must eventually reach fishable size and will then need to be considered as part of the full population.

The assessment is based on size data (shell heights), although scallops are generally relatively easy to age, and in Europe we would normally run age-based assessments. When asked about this, the response was that the first one or two ages can be difficult to distinguish with these scallops (so that growth is modelled as increments rather than absolute ages). Although this may be true, I think it would still be worthwhile to look further into ageing issues, as an age-based assessment is generally more reliable and less reliant on assumptions. We were also told that shells on which to base ages (and sizes) were not available from the Mid-Atlantic (MA) area from 2015 onwards, as they had not been collected by VIMS (who are the responsible body in the area). However, it subsequently became apparent that shells had been collected by VIMS since 2015, but that there had been a lack of communication of this. This is an important aspect to rectify for next time.

Regarding growth, it is clear that growth rates have increased recently. The hypothesis presented was that this is a fishery effect: high fishing effort focusses on fast-growing fish, leaving only slow-growing smaller scallops, with both spatial and (less so) selectivity effects contributing. However, the conclusions on "growth periods" are a little *ad hoc* and were estimated mostly by eye at the SAW meeting – this analysis could use some more rigorous changepoint or cusum analysis. The key question is whether management advice is sensitive to these essentially *ad hoc* decisions. While the choice of growth periods should not be influential if the CASA fits the length distributions correctly, this is not (currently) always the case so it is an issue to be considered.

The stock is currently benefiting from two very large year-classes (2012 (GB) and 2013 (MA)) – normally one scallop per square metre would be a high density, whereas the density of these year-classes was around 100 or more per square metre. The data provided to support the existence of these year-classes was very impressive, particularly the HabCam and drop-frame camera images. The scallops are growing slowly, but numbers remain good although they are being thinned out by natural mortality (the area concerned has been closed and the scallops are too small to land commercially, so there should be no fishing mortality). Small scallops can swim away from sea stars, but will clam up in front of crabs and hope for the best – so areas high in crabs will have correspondingly high natural mortality for young scallops.

Finally, Figure A4.1 would be improved with notation showing where the assessment areas are – currently this needs to be inferred by the reader.

ToR A1. Estimate catch from all sources including landings, discards, and incidental mortality. Describe the spatial and temporal distribution of landings, discards, and fishing effort. Characterize the uncertainty in these sources of data.

This ToR was fully completed, and I concur with the conclusions of the SARC panel.

The management of this fishery is quite complicated, and I found it difficult to follow precisely who was able to fish where and when from the text. It would have been helpful to summarise the current regulations in a short table. I noted that the number of days at sea have declined to 24 in open areas by 2018. While most of the landings comes from licensed vessels fishing in closed areas (as I understand it), an extreme limit of this kind on effort can have perverse effects (for example, on market prices), and may *in extremis* lead to safety problems.

There appears to be some text missing from the final paragraph of the "Fishing Effort and LPUE" section, where it discusses VMS data.

I appreciated the work done on discard and incidental mortality, as these are significant and it was useful to get a firm overview of how they were estimated. Equation A5.3 looks incorrect, however – is this an error, or have I misunderstood what was intended?

ToR A2. Present the survey data being used in the assessment (e.g., regional indices of relative or absolute abundance, recruitment, size data, etc.). Characterize the uncertainty and any bias in these sources of data.

This ToR was fully completed, and I concur with the conclusions of the SARC panel.

The assessment team were keen to point out that the available survey data for this stock were excellent, and for the most part I would agree with them. The Habcam and drop-camera surveys in particular seemed very clear and performed well, although there remain issues with detection of live (as opposed to dead) scallops that may be causing errors and which need to be addressed. The longer-running dredge survey does appear to run into difficulties when scallop densities reach high levels, and as this is certainly currently the case that survey should be treated with some caution. This was addressed to a certain extent through a decision to reduce the assumed efficiency of the dredge survey in the high density areas (Elephant Trunk and Nantucket Lightship) by a factor of three during 2015-17, but this appeared quite *ad hoc* and was not well supported in the report. When questioned

about this, the response was that this was the best that could be done – but in this case, I would like to have seen analysis outlining how sensitive the assessment and advice were to this assumption (through profile likelihood modelling or similar).

A start has been made with automated image analysis for the camera surveys, although the impression I got was that the framework and data pipeline were in place without yet any completed analysis tools. Given our experience in Scotland with automated image analysis for discard monitoring (see French et al. 2015, and https://www.uea.ac.uk/computing/graphics-vision-and-speech/automated-image-analysis-for-fisheries), I would strongly urge the direction of research activity in this area, and I would be happy to consider any collaboration suggestions. If a human can learn to differentiate between live and dead scallops from video footage or stills, then so can a deep neural net system, and the efficiency and accuracy savings can be impressive.

ToR A3. Summarize existing data, and characterize trends if possible, and define what data should be collected from the Gulf of Maine area to describe the condition and status of that resource. If possible provide a basis for developing catch advice for this area.

This ToR was fully completed, and I concur with the conclusions of the SARC panel.

The Gulf of Maine scallop fishery and stock are relatively data-poor, and it proved difficult to determine how to approach the assessment from the perspective of a review (although we are expected to comment as it is specified in a ToR). Certainly the area is not as important economically as the MA and GB areas, but there is still an increasing LPUE in recent years and the question is: is the Gulf of Maine sufficiently important to consider resourcing more extensive data collection? The relative landings proportions in the three areas would suggest probably not, unless it is critical for local coastal communities, and we were told that the area is an important consideration at the relevant Management Council. In terms of importance to the stock and the economy, however, I would conclude that it would be more beneficial to seek to incorporate the Canadian part of Georges Bank into the overall assessment.

The Gulf of Maine seems to be very data limited and there is no assessment yet. It is a patchy area for scallops, with a very heterogeneous distribution, although density seems high (compared to Georges Bank levels) where there are scallops. The area is not regularly surveyed and there are no discard observations, although there are quite detailed VMS data, with pings every 30 minutes for all vessels. I also note that Appendix B3 is well-written and contains many impressive data summaries (although derived from relatively few data).

The key point of the ToR was to look for a ranking of the three management approaches used or proposed recently:

- Assessing using CASA, estimating YPR, and determining reference points.
- Expanding the SAMS model to include the Gulf of Maine area.
- Other methods: catch-based advice, depletion modelling.

Of these, probably only the third has much chance of success, given the current level of data availability. There are quite detailed data on scallop distribution and biology from the area (and the relevant Appendix is one of the longer ones as a consequence), but such data are generally from snapshot surveys that are not then continued, and there is unlikely to be sufficient data yet to inform either of the first two approaches.

Whether the fishery is deemed important enough to warrant increased resources for data collection, collation and assessment is a matter for fisheries managers, and that decision will need to be taken before the required data gathering process can begin in earnest.

ToR A4. Investigate the role of environmental and ecological factors in determining stock distribution and recruitment success. If possible, integrate the results into the stock assessment.

This ToR was fully completed, and I concur with the conclusions of the SARC panel.

This was an interesting section in both the report and the SARC meeting, and I learned a lot about American scallops that I hadn't previously known. Small scallops can be swallowed whole by the American starfish *Astropecten americanus*, while crabs are another important predator, and *Didemnum vexillum* (or "D. vex", also a big problem in Scotland) can make areas uninhabitable for scallops.

In terms of integration into the stock assessment, the main route that I could find for this was through the use of area-specific growth relationships for the different discrete stock regions in the SAMS forecast model, thus accounting for the influence of location (and temperature) on growth (scallops appear to be directly limited by cold temperatures). The different growth periods built into the CASA model (as noted in my Introduction section above) should integrate changes in the environment, as well as shifts between shallow and deeper water, but this is not done explicitly. In any case, I would tend not to encourage the direct use of environmental factors into assessment and advice, as this is too often done without knowledge of causal mechanisms (or with mechanisms that break down).

ToR A5. Estimate annual fishing mortality, recruitment and stock biomass for the time series, and estimate their uncertainty. Report these elements for both the combined resource and by sub-region. Include retrospective analyses (historical and within-model) to allow a comparison with previous assessment results and previous projections.

This ToR was fully completed, and I concur with the conclusions of the SARC panel.

While I contributed in full to the discussion panel on the scallop stock assessment, I was less experienced in the area of length-based assessment than (certainly) two of the panel members. The following comments are therefore more qualitative in most cases, and I have not attempted to discuss the finer technical details of the CASA model used. An excellent Appendix (A6) covered the implementation of the CASA model in considerable detail, which was a great help to the reviewers.

Generally speaking (and as mentioned above), it seems that sea scallops are organised into several discrete areas on the basis of substrate and other environmental conditions, and I queried whether there had been consideration about treating them as more than two stocks (similarly to the way in which European *Nephrops* stocks are assessed as independent (and sometimes quite small) Functional Units). How linked are each of the subareas? It seems that the limitation is more one of data availability at the appropriate spatial scale than anything more fundamental – it is also likely that several small assessments would probably give the same overall perception as one large assessment. Having said that, local assessments would be beneficial in helping to prevent localised depletion, so there may be merit in considering the issue further.

On a related note, age-based assessment is more commonly used in Europe, because the models tend to be more robust and there are fewer concerns in Europe about the aging of scallops. No such age-based model was presented to the SARC, but it would be very interesting to do so for the next assessment.

The assessment for Georges' Bank is split between open and closed areas, on the basis that fishery impacts (and hence estimated selectivity) should be quite different in the closed areas. The closed areas are discontinuous and lie within the open area, however, so the distinction doesn't make much sense from a biological point of view.

There was considerable discussion, both in the SARC meeting and intersessionally between the panel members, on whether natural mortality M could actually be estimated at the same time as fishing mortality F without overparameterisation. The CASA model estimates juvenile mortality in two of the stocks (MA and GB open) which are not fished at those sizes, and for which therefore M cannot be confounded with F. The GB closed area also doesn't have much fishing, and there is an independent estimate of scale from surveys. However, some unreported landings will be interpreted in the model

as natural mortality, as will errors in survey selectivity estimates or divergence between survey and fishery catchability. So when the report talks about estimating natural mortality, it is really more like the ICES concept of "unaccounted removals", which is probably a more explicit and honest term.

On a related note, the report concluded that large recruitments which subsequently enter the fishery as moderate (or even small) year-classes indicate high natural mortality in the intervening years. However, it is also plausible the surveys are overestimating strong recruitment – this is a common feature in European trawl surveys at any rate, and plots comparing survey indices with final assessment (e.g., Figure A9.37) show that the surveys do tend to estimate larger year-classes as being very large indeed.

A number of potentially useful plots were missing from the report, including retrospective plots with confidence intervals and survey residual plots, but these were provided in good time during the review meeting following panel requests.

One of the main issues raised by the panel was the fact that the fitted survey data don't fit the observations very well (or at all) in some of the early years. It is not very clear why this should be – presumably the model is following the extant catch data for these years and essentially ignoring some of the survey data. This may have no implications for current management advice, but it is worth raising as a potential issue with the model (or with the survey data) that would benefit from reconsideration. More important for management purposes may be the poorly-fitting years for some of the shorter, more recent survey series (an example is given in Figure A9.28) – if these are so un-influential on the model fit, what is the purpose in using them?

Page 76 of the report discusses the very tight confidence intervals (CIs) about the model estimates for F and SSB (Figure A9.61). The text states that all models are simplifications, but while this is true, it doesn't explain why the CIs are so unusually tight (in comparison with most other stock assessment models I have seen). A hypothesis is provided for this, but not a real explanation, and I would be concerned that this may be indicative of an underlying problem with the model data or assumptions.

Finally, spatio-temporal abundance model maps through time would be more useful for managers than some of the current, more standard time-series analysis plots, given the way the stock is modelled and managed.

ToR A6. State the existing stock status definitions for "overfished" and "overfishing". Then update or redefine biological reference points (BRPs; 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. Comment on the scientific adequacy of existing BRPs and the "new" (i.e., updated, redefined, or alternative) BRPs.

This ToR was fully completed, and I concur with the conclusions of the SARC panel. It is certainly the case that the stock is not overfished, nor is overfishing occurring. However, there are some points that are worth raising.

Firstly, there is a comment towards the end of the section (when talking about the basis of the YPR estimation) that "the reference point calculations are based… on assumptions that are never even approximately true" – so why present them? The assessment team have had to use several fairly intricate tricks to get the analysis to work, and even the YPR curves are both extremely uncertain and very flat – the SYM model estimates F(max) at the upper bound of the analysis on many occasions, for the Mid-Atlantic area at any rate. We have to ask whether an equilibrium method like this is a reliable basis for management? We have had similar experience at ICES recently with SPiCT, which is a stochastic production model that often turns out to be unreliable. It should be possible to use various size-based indicators instead for sea scallops, to indicate likely stock status in relation to MSY proxies – at least in a confirmatory sense. The law says that MSY *must* be estimated, so I can appreciate why it is done, but it would help to support management conclusions if a simpler, more appropriate method were used in parallel.

Secondly, it would also be helpful to estimate a CV on the estimated reference points – this would facilitate interpretation and a better understanding of where the stock really is in relation to MSY.

Thirdly, there is generally no fishing on ages 1 and 2, and hence reference point estimates are based on recruitment at age 3 (through another intricate modelling trick). It would also have been possible to run the whole assessment assuming recruitment at age 3, in fact – as then there would be no need to consider estimates of M for younger ages which are largely inferred from older ages, as well as the difficulty in detecting the younger ages well in shells anyway. The disadvantage of this approach would be that there would be a lag in management advice, but it would at least improve the consistency between the assessment and the reference point estimation.

Finally, the report uses gonad weight as a proxy for reproductive potential, rather than an SSB estimate based on meat weight as was done previously. This is probably more relevant biologically, although it is simply stated in the report that "gonad weight is a better indicator of fecundity than meat weight" without any references or further justification. It does also introduce many management issues which led to the most contentious discussion of the SARC review meeting. This will covered in more detail under the next ToR.

ToR A7. Make a recommendation about what stock status appears to be based on the existing model (from previous peer reviewed accepted assessment) and based on a new model or model formulation developed for this peer review.

a. **Update the existing model with new data and evaluate stock status (overfished and overfishing) with respect to the existing BRP estimates.**
b. **Then use the newly proposed model and evaluate stock status with respect to "new" BRPs and their estimates (from TOR-5).**
c. **Include descriptions of stock status based on simple indicators/metrics.**

This ToR was fully completed, and I concur with the conclusions of the SARC panel.

The report stock summary tables for this ToR were all based on the gonad-weight measure of SSB (used for B(msy)), while meat weight is used for the estimation of MSY and F(msy). As mentioned above, biologically this may be appropriate (although that would need to be demonstrated), but for a fisheries manager this is quite confusing. The panel recommended that both gonad- and meat-weight based biomass summaries be included throughout the report, and this was achieved by the assessment team with a good deal of hard work through the SARC review week – this was very much appreciated by all on the panel.

This decision followed some vigorous representation by the NE Fisheries Council officials who were present, including their contracted lawyer. The legal document enshrining the regulations for this fishery requires stock summaries in meat weights, so this needs to be included in the assessment report for the time being. While this may appear quite legalistic, it is worth noting (from a personal viewpoint) that the discussion on this issue would have been much more prolonged and confrontational in Europe, with 28 countries, co-decision between Council, Commission and Parliament, and requirement for agreement with Norway and other relevant non-EU coastal states.

On a side note, that second paragraph in this Section includes the line "there is essentially no chance that the stock is not overfished" – the "not" needs to be removed here to avoid the wrong conclusion.

ToR A8. Develop approaches and apply them to conduct stock projections.

a. **Provide numerical annual projections (through 2020) and the statistical distribution (i.e., probability density function) of the catch at FMSY or an FMSY proxy (i.e. the overfishing level, OFL) (see Appendix to the SAW TORs). Each projection should estimate and report annual probabilities of exceeding threshold BRPs for F, and probabilities of falling below threshold BRPs for biomass. Use a sensitivity analysis approach in which a range of**

assumptions about the most important uncertainties in the assessment are considered (e.g., terminal year abundance, variability in recruitment).

b. **Comment on which projections seem most realistic. Consider the major uncertainties in the assessment as well as sensitivity of the projections to various assumptions. Identify reasonable projection parameters (recruitment, weight-at-age, retrospective adjustments, etc.) to use when setting specifications.**

c. **Describe this stock's vulnerability (see "Appendix to the SAW TORs") to becoming overfished, and how this could affect the choice of ABC.**

This ToR was fully completed, and I concur with the conclusions of the SARC panel.

Projections were carried out using the Scallop Area Management Simulator (SAMS). SAMS allows for spatially-differentiated F and should therefore be ideal for scallops, which are clearly subject to this type of mortality. However, the text (p. 169) mentions the F is assumed to be "spatially uniform" (apart from closed areas). The reason for this is that the ToRs ask for fishing at a constant F(msy), so this has been interpreted as constant in time and space. This may not be the most useful interpretation of the ToR, and for future projections it may be worth considering a more realistic model of F variation.

Another issue is that the medium-term projections don't have management feedback loops, so the assumed F is fixed throughout each simulation. This won't really happen, so is this projection informative? In this case, the biomass looks to remain high anyway, but it might be appropriate to consider the use of a model that incorporates feedback loops (a non-spatial example would be the tools available in the FLR library – see http://www.flr-project.org/). However, management of the stock does not appear to be conducted through formal harvest-control rules in any case, but rather through *ad hoc* responses to the current and projected status, so a formal MSE may not prove fruitful.

ToR A9. Review, evaluate and report on the status of the SARC and Working Group research recommendations listed in most recent SARC reviewed assessment and review panel reports. Identify new research recommendations.

This ToR was fully completed, and I concur with the conclusions of the SARC panel.

The research recommendations were presented in two lists. The first covered previous research recommendations and included notes as to whether (and how) they had been addressed since the last assessment meeting. The degree of success varied – some recommendations had not been addressed at all, whereas some had progressed significantly. One area that I am particularly interested in is the suggestion for automated image analysis for the HabCam survey (#5), and I followed this up with the presenter. The ongoing image analysis project (VIAME) is getting 50% agreement with human viewers now, so it is not yet ready to be implemented. The key issues are that it is hard to tell dead from living scallops, differentiation with sand dollars can be problematic, and sediment cover causes difficulties. This bears many resemblances with work we are doing in Scotland on discard monitoring, so it would be worthwhile following up potential collaborations.

The second list presented future research recommendations, which look to be useful and comprehensive. My only comment would be that there is a danger that this becomes a broad wish-list with little chance of success – it is important to consider the resource implications (in terms of money and people) for any recommendation, and only include those that are likely to be addressed.

B. Atlantic herring

ToR B1. Estimate catch from all sources including landings and discards. Describe the spatial and temporal distribution of landings, discards, and fishing effort. Characterize uncertainty in these sources of data. Comment on other data sources that were considered but were not included.

This ToR was fully completed, and I concur with the conclusions of the SARC panel.

The report began with a useful and interesting summary of the history of fisheries management for this fishery: for example, I was surprised to read that most herring are destined for use as lobster bait, which seems a waste of a species that is the basis for a large human-consumption fishery in Europe and elsewhere. I also found it interesting that quotas and allocations are set only every three years, and I think this approach is responsible for some of the issues arising with the stock forecast and advice. It is stated that the "best available information" is used to determine quota allocation between areas, but despite following up on this during the meeting, I never got a clear answer on how this is done.

I asked two further questions about the management of the fishery during the review meeting. Firstly, I enquired about the possibility of illegal or unreported landings. This would have been a difficult question to answer directly with the US herring industry representatives in the room, but the response was that, while previously VTR and dealer reports were not very consistent, they are much better now. The issue had also been raised during the SAW assessment, but mostly in relation to foreign fleet landings from the period prior to the implementation of the EEZ – these can be problematic and are considered to be relatively uncertain.

Secondly, I asked how mixed is the fishery – in other words, is it a clean fishery catching only herring and mackerel? It seems that there are small amounts of bycatch in relation to the catch of these species, but that the stocks involved (river herring or alewife blueback, and haddock) are sensitive to this bycatch so limits are imposed.

Catch data were presented in an appropriate amount of detail, for the most part. Catches consist mostly of landed fish – there are discard estimates from self-reporting, but while it is not clear how reliable these are, they amount to less than 1% of catch (generally) so their veracity is probably not critical. As an aside, I do find it surprising that discards are so low in an essentially demersal fishery for which TACs (in some areas) can be restrictive. There is a limited observer scheme, but only those vessels intending to fish in areas closed to demersal fishing are required by law to carry observers so this does not form a key part of the data collation exercise.

Catch-at-age data raising accounted for gear type, but not for spatial considerations (see p. 20, 3rd paragraph). However, the Tables referred to in order to support this decision don't have the data necessary to do so – this may just be a table numbering issue. I would have found a map of the relative locations of the US and Canadian fisheries towards and within New Brunswick waters, as without one it was more difficult to evaluate what the likely implications would be. A Figure (B1-7) is given showing the history of landings for fixed and mobile gear, but it would also be useful to see a table and figure giving the historical landings by country and year (the first Figure citation on page 21 is perhaps misleading). I also queried why fixed gear seems to catch only age-2 herring – the response was that it is due to both selectivity or area: younger fish tend to be distributed more inshore, but the gear selects against them as well.

On a more minor point, it is not clear from Tables B1-3 to B1-5 whether the samples mentioned are from port sampling or onboard observers. Overall, the tables of input data are well laid out, and would be very helpful for other researchers wishing to attempt to replicate or check the results. The Figures are also helpful – the one format that I would query would be the bubble plots in Figures B1-8 and B1-9, as I consider bivariate scatterplots by year-class to be more informative of the ability of a data series to track year-class strength. These were provided on request for the survey data, but I would recommend that they be included as standard in the catch diagnostic plots as well. Also, the weights-at-age time series plot (Figure B1-10) needs a key for the ages.

Finally, the FLDRS programme sounds like a promising approach to collecting and interpreting fishery-dependent data, and I would encourage the implementation and use of this in the assessment and advisory process. The FLDRS vessel coverage by 2016 was up to 40%, which is significant, and the programme provides a good basis for further development. Although it wasn't covered in any detail during the SARC week, the notes in Appendix B6 were very useful. There are always issues with the

use of “sentinel” fleets of this kind, principally how to ensure normal fishing operations while being monitored, and how does one ground-truth the information coming from the programme, but it remains a very promising start.

ToR B2. Present the survey data being used in the assessment (e.g., regional indices of abundance, recruitment, state surveys, age-length data, food habits, etc.). Characterize the uncertainty and any bias in these sources of data.

This ToR was fully completed, and I concur mostly with the conclusions of the SARC panel.

Unlike the catch data, here there are no tables summarising the relevant survey data for this stock. This makes subsequent reconsideration and checking of the data very difficult, and also means that other scientists cannot attempt to check the results.

As mentioned for catch data above, I tend to find bubble plots for survey data (Figures B2-4) quite difficult to interpret. I requested additional plots (bivariate scatterplots by year-class), which were speedily provided by the assessment team, and which showed reasonable (although not great) cohort tracking. In general, ICES assessments would usually include many more diagnostics plots of the basic input data (catches and surveys) to enable an evaluation of whether the assessment is likely to be representative, and some more of these summaries would be useful here. It is good to see uncertainty bounds on survey plots, though, and this is always to be encouraged. I was also pleased to see concerns expressed over the calibration between indices from successive research vessels, and sensitivity analyses to determine the impact of splitting the survey series at the point the vessel changed.

This assessment (for the first time) included a relative abundance index derived from an acoustic survey. However, this was not a directed acoustic survey with regular transects and fishing on marks to verify acoustic estimates, but rather acoustic signals from a standard demersal trawl from which a standard age-based RV index is also derived. For me this is an improvement, but I find it difficult to understand the use of a demersal trawl survey index in the assessment of herring. I queried this during the meeting, and was told that the herring are thought to be largely hard to the bottom during the survey, which if true would mean that a demersal trawl could be representative. However, herring are a fundamentally pelagic, midwater species, and a directed acoustic survey should still be the best fishery-independent means of evaluating abundance and distribution. This may be a very ICES view, however, as my colleagues on the review panel had no issues with it. I note that the acoustic survey is presented in the report in considerably more detail than the demersal trawl surveys, although that may be because this is the first time it has been used.

ToR B3. Estimate consumption of herring, at various life stages. Characterize the uncertainty of the consumption estimates. Address whether herring distribution has been affected by environmental changes.

This ToR was fully completed, and I concur with the conclusions of the SARC panel.

The assessment for this stock uses a time- and age-invariant natural mortality rate M of 0.35, based on a combination of the Hoenig and Lorenzen methods (although the report is not very clear on this point). The aim of the first part of this ToR was to determine whether this estimate was at about the right level, by estimating predation consumption rates, rather than necessarily using the consumption analysis to determine new M estimates. It would have been useful to see more details about how this was done: for example, how are herring detected in stomach samples, and if it is through otoliths, how rapid are digestion rates? The report concludes that the level is about right, although Figure B3-3 would suggest that the estimates are very different. In any case, the consumption estimates are missing data from very important predators such as marine mammals, seabirds and large finfish, so it is perhaps not surprising they are not consistent with other estimates. The analysis method is well described and seems plausible.

For the second part of the ToR, the centre-of-gravity plots (Figure B3-4) don't show much change through time, implying that stock distribution has not really changed.

ToR B4. Estimate annual fishing mortality, recruitment and stock biomass (both total and spawning stock) for the time series, and estimate their uncertainty. Incorporate ecosystem information from TOR-3 into the assessment model, as appropriate. Include retrospective analyses (both historical and within-model) to allow a comparison with previous assessment results and projections, and to examine model fit.

This ToR was fully completed, and I concur with the conclusions of the SARC panel.

The principal assessment was conducted using the ASAP model, which is an implementation of a Fournier-Archibald type forwards-iterating statistical catch-at-age model. The assessment team were all very familiar with the model (one of them had been a co-author on the original paper), and they were able to explain clearly and concisely what had been done and why. This was helpful, as the report itself was not always quite so clear, and there were a few points in the text that required further clarification. In my view, it was sometimes not very obvious exactly what has been used in the final assessment. I also note that no output tables were given – F and N at age, for example – and I would recommend that these be included next time.

The main alternative (exploratory) model presented was SAM, a state-space model developed by Anders Nielsen in Copenhagen. While the SAM model fits look reasonable, SAM itself is very much a black art (in my experience) and most of the SAM assessments currently used in ICES require significant input from Anders to set up and maintain. The model is certainly worthy of further exploration, but it will not necessarily "fix" issues that might exist with the data. A Stock Synthesis III (SS3) model was also implemented, but this remains at a quite early stage of development.

The report contains very little on stock structure, and I asked whether this was something that the SAW WG had considered. The response was that yes, they had, in 2012 – however, there are no tagging data, and not much work has been done on DNA, morphometrics or otolith microchemistry. Real stock structure should be accounted for if possible, but you do need to be careful what you ask for – we have been going down this path with West of Scotland herring for several years, and as the data are not quite sufficient for a clear determination of sub-stock structure, the attempt to do so tends to lead to more management problems than we had before.

Assessments generally in the US tend to have more retrospective noise and bias than we would have in ICES assessments, and this herring assessment is no exception. In the previous assessment, this had been "fixed" by increasing M by 50%, but this strikes me as extremely *ad hoc* and I am pleased to see this hasn't been done in the assessment. High retrospective bias tends to lead to the widespread use of such *ad hoc* fixes here, but it would be helpful to understand why the bias happens and what (if anything) can or should be done about it. Obviously this has been a key topic of research for many years, but it seems no closer to a solution. Discussion among the panel led to the suggestion that US assessments may have more retro bias because the vessels used for survey series tend to change more rapidly than would be the case for ICES assessments.

I found the use of only ages 7 and 8 in the estimation of mean F to be a strange choice. The text states that these are the first ages that are fully selected by the dominant mobile gear fishery, but it still seems odd to only use two ages for the summary mean - the mean will then be subject to high uncertainty which would be evened out if more ages were included. My personal preference would be for a wider age range to minimise the impact of assessment uncertainty.

Regarding the outcomes, it is clear that the last 5 recruit estimates are all poor. Even if the last 2 estimates were towards the upper ends of their wide confidence intervals, they would still be below the long-term geometric mean, and the short-to-medium term prognosis for the stock is not good as a consequence. The revision in perception in the new assessment is also considerable, although as the reference points have also been revised this may not be quite so impactful.

Overall, I found this to be an excellent stock assessment chapter once a few clarifications had been made. One further suggestion I would make would be to improve the figure captions – they are often very short and don't tell the reader what the plot is about.

ToR B5. State the existing stock status definitions for "overfished" and "overfishing". Then update or redefine biological reference points (BRPs; 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. Comment on the scientific adequacy of existing BRPs and the "new" (i.e., updated, redefined, or alternative) BRPs.

This ToR was fully completed, and I concur with the conclusions of the SARC panel.

It did not seem to be possible to fit a Beverton-Holt model as part of the standard reference-point estimation for this assessment, so $F_{40\%}$ was used as an F_{msy} proxy. To me, this doesn't have a strong justification, but it does seem to be a standard approach in this part of the world (and in the experience of the other reviewers). Although this is a reasonable approach, I would argue for exploration of something rather more qualitative – elsewhere in this review I have discussed the use made in ICES of the stochastic production model SPiCT (if it works) or length-based proxies (which I would prefer), and there would be a good argument for these to be considered for this herring stock also.

There also seems to be no real justification for the 80% confidence intervals on p. 181 in the report. Indeed, throughout the report, 80% is used in some places, while in others it is 90%, and it would be beneficial for the managing authority to stipulate what is required (and ask for it to be applied consistently).

I note that the new reference point estimates are considerably different from the old. While this may be accounted for by the rescaling of the stock estimates in the new assessment, it is difficult to tell if stock perception has changed as there is no comparison of F/F(msy) and B/B(msy) between old and new. This would be a useful check to include.

ToR B6. Make a recommendation about what stock status appears to be based on the existing model (from previous peer reviewed accepted assessment) and based on a new model or model formulation developed for this peer review.

a. **Update the existing model with new data and evaluate stock status (overfished and overfishing) with respect to the existing BRP estimates.**
b. **Then use the newly proposed model and evaluate stock status with respect to "new" BRPs and their estimates (from TOR-5).**
c. **Include descriptions of stock status based on simple indicators/metrics.**

This ToR was fully completed, and I concur with the conclusions of the SARC panel.

For me, the main thing missing from the report section was time series plots of estimated F and SSB with both confidence intervals and the new reference points. These were quickly provided during the SARC meeting.

There is no management action planned because ½ B(msy) < B < B(msy), and action is only triggered currently if B < ½ B(msy). It might be fruitful to consider alternative harvest control rules (HCRs), including such approaches as hockey stick rules.

ToR B7. Develop approaches and apply them to conduct stock projections.

a. **Provide numerical annual projections (through 2021) and the statistical distribution (i.e., probability density function) of the catch at FMSY or an FMSY proxy (i.e. the overfishing level, OFL) (see Appendix to the SAW TORs). Each projection should estimate and report annual probabilities of exceeding threshold BRPs for F, and probabilities of falling below threshold BRPs for biomass. Use a sensitivity analysis approach in which a range of**

assumptions about the most important uncertainties in the assessment are considered (e.g., terminal year abundance, variability in recruitment).

b. Comment on which projections seem most realistic. Consider the major uncertainties in the assessment as well as sensitivity of the projections to various assumptions. Identify reasonable projection parameters (recruitment, weight-at-age, retrospective adjustments, etc.) to use when setting specifications.

c. Describe this stock's vulnerability (see "Appendix to the SAW TORs") to becoming overfished, and how this could affect the choice of ABC.

This ToR was fully completed, and I concur with the conclusions of the SARC panel.

The report contains no equivalent to a catch-option table as we would always have in ICES – all options are given for fixed F. This is what is asked for in the ToRs, so it is correct, but I think it a missed opportunity to demonstrate to managers the likely outcomes of a wide range of potential actions – from closing the fishery to doubling fishing effort. I believe these options are explored further in later stages in the management process, but it would still have been helpful to have seen them here.

Using the long-term geometric mean (GM) for recruitment for 2019-2021 implies higher recruitment than has seen for 5 years, so this may be unrealistic (although not a regime shift as such does not yet seem likely). The panel asked for alternative runs looking at a more recent recruitment GM for 2019-2021, and these were provided in good time. The conclusions were as expected – the lower recruitment assumption increased the risk of biomass declining below the reference point level.

The ASAP model does appear to have a retrospective pattern, although no correction was made for this forecast. The report expresses serious misgivings that future herring assessments will have worsening retrospective patterns, although I don't agree that it will *necessarily* worsen in future years. As I've mentioned elsewhere, retro bias does seem to be worse with North American assessments than with European ones, and there must be a fundamental reason for this.

ToR B8. If possible, make a recommendation about whether there is a need to modify the current stock definition for future assessments.

This ToR was fully completed, and I concur with the conclusions of the SARC panel. There are currently insufficient data to enable catches to be disaggregated into any likely substock components.

ToR B9. For any research recommendations listed in SARC and other recent peer reviewed assessment and review panel reports, review, evaluate and report on the status of those research recommendations. Identify new research recommendations.

This ToR was fully completed, and I concur with the conclusions of the SARC panel.

Similarly as for scallops, the research recommendations were presented in a series of lists – the first giving new recommendations, followed by several giving previous recommendations (organised in this case by the source meeting). The first list looks to be useful and comprehensive, although (as before) it is important to consider the resource implications (in terms of money and people) for any recommendation, and only include those that are likely to be addressed. The list of previous recommendations showed that most of these had not been addressed successfully, which is a little disappointing although perhaps not surprising, given how many there are.

Appendix 1: Bibliography of materials provided for review and references

References

French, G., Fisher, M., Mackiewicz, M. and Needle, C. L. (2015). Convolutional Neural Networks for Counting Fish in Fisheries Surveillance Video. In Amaral S. Matthews, T. P. S. M. and Fisher, R. (Eds.), *Proceedings of the Machine Vision of Animals and their Behaviour (MVAB)*, BMVA Press, 7.1-7.10.

Lamport, L. (1994). *LaTeX: A document preparation system* (2nd edition). Addison Wesley.

Materials provided for review

Working Papers

Working Group, Stock Assessment Workshop (SAW 65) (2018). Stock Assessment Report for Atlantic Sea Scallop. Working Paper A1 SAW/SARC 65. June26-29, 2018, NOAA Fisheries, Northeast Fisheries Science Center. Woods Hole, MA.

Working Group, Stock Assessment Workshop (SAW 65) (2018). Stock Assessment Report for Atlantic Herring. Working Paper B1. SAW/SARC65. July26-29, 2018, NOAA Fisheries, Northeast Fisheries Science Center. Woods Hole, MA.

Working Group, Stock Assessment Workshop (SAW 65) (2018). Stock Assessment Summary Report for Atlantic Sea Scallop. Working Paper A2. SAW/SARC 65, June 26-29, 2018, NOAA Fisheries, Northeast Fisheries Science Center, Woods Hole, MA.

Working Group, Stock Assessment Workshop (SAW 65) (2018). Stock Assessment Summary Report for Atlantic Herring. Working Paper B2. SAW/SARC 65, June26-29, 2018, NOAA Fisheries, Northeast Fisheries Science Center, Woods Hole, MA.

Presentations

Sea Scallop Working Group, Dvora Hart, Jui-Han Chang, Burton Shank (2018). Sea Scallop Assessment (SARC 65). SAW/SARC 65, June26-29, 2018, NOAA Fisheries, Northeast Fisheries Science Center, Woods Hole, MA.

Atlantic Herring Working Group, Jon Deroba (2018). Atlantic Herring Assessment (SARC 65). SAW/SARC 65, June26-29, 2018, NOAA Fisheries, Northeast Fisheries Science Center, Woods Hole, MA.

Background papers

Noel Cadigan, Martin Cryer, J.-J. Maguire (Chair), Jon Helge Vølstad, Brent Wise (2015). Summary Report of the Review of Sea Scallop Survey Methodologies and Their Integration for Stock Assessment and Fishery Management Meeting to Review Sea Scallop Survey Methodologies and Their Integration for Stock Assessment and Fishery Management. 17-19 March 2015, Waypoint Event Center at the Marriott Fairfield Inn and Suites, New Bedford, MA.

Jui-Han Chang, Burton V. Shank, Deborah R. Hart (2017). A comparison of methods to estimate abundance and biomass from belt transect surveys. Limnol. Oceanogr.: Methods 15, 480–494.

Jonathan J. Deroba (2018). Sources of variation in stomach contents of predators of Atlantic herring in the Northwest Atlantic during 1973–2014. ICES Journal of Marine Science, doi:10.1093/icesjms/fsy013.

Deborah R. Hart & Paul J. Rago (2006) Long-Term Dynamics of U.S. Atlantic Sea Scallop Placopecten magellanicus Populations, North American Journal of Fisheries Management, 26:2, 490-501, DOI: 10.1577/M04-116.1.

Hart, D. R., and Chute, A. S. (2009). Estimating von Bertalanffy growth parameters from growth increment data using a linear mixed-effects model, with an application to the sea scallop Placopecten magellanicus. ICES Journal of Marine Science, 66: 2165–2175.

Deborah R. Hart (2013). Quantifying the tradeoff between precaution and yield in fishery reference points. *ICES Journal of Marine Science,* doi:10.1093/icesjms/fss204.

Deborah R. Hart, Larry D. Jacobson, Jiashen Tang (2013). To split or not to split: Assessment of Georges Bank sea scallops in the presence of marine protected areas. Fisheries Research 144, 74–83.

Daniel R. Hennen and Deborah R. Hart (2012). Shell Height-To-Weight Relationships for Atlantic Sea Scallops (Placopecten magellanicus) in Offshore U.S. Waters. Journal of Shellfish Research, 31(4):1133-1144.

Invertebrate Committee (2013). Stock assessment for Atlantic sea scallops in 2014, updated through 2013. SAW/SARC 59.

Northeast Fisheries Science Center. (2012). 54th Northeast Regional Stock Assessment Workshop (54th SAW) Assessment Report. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 12-18; 600 p. Available from: National Marine Fisheries Service, 166 Water Street, Woods Hole, MA 02543-1026, or online at http://www.nefsc.noaa.gov/nefsc/publications/.

Kevin D. E. Stokesbury, Bradley P. Harris, Michael C. Marino II and Jacob I. Oguelra (2004). Estimation Of Sea Scallop Abundance Usng A Video Survey In Off-shore Water. Journal Of Shellfish Research, 23(I), 33-40.

Samuel B. Truesdell, Deborah R. Hart, and Yong Chen (2016). Effects of spatial heterogeneity in growth and fishing effort on yield-per-recruit models: an application to the US Atlantic sea scallop fishery. ICES Journal of Marine Science, 73(4), 1062–1073. doi:10.1093/icesjms/fsv238.

Samuel B. Truesdell, Deborah R. Hart, and Yong Chen (2017). Effects of unequal capture probability on stock assessment abundance and mortality estimates: an example using the US Atlantic sea scallop fishery. Can. J. Fish. Aquat. Sci. 74: 1904–1917. dx.doi.org/10.1139/cjfas-2016-0296.

Appendix 2: Statement of work

Statement of Work

National Oceanic and Atmospheric Administration (NOAA)

National Marine Fisheries Service (NMFS)

Center for Independent Experts (CIE) Program

External Independent Peer Review

65th Stock Assessment Workshop/Stock Assessment Review Committee (SAW/SARC) Benchmark stock assessment for Sea scallop and Atlantic herring

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. (http://www.cio.noaa.gov/services_programs/pdfs/OMB_Peer_Review_Bulletin_m05-03.pdf).

Further information on the CIE program may be obtained from www.ciereviews.org.

Scope

The Northeast Regional Stock Assessment Review Committee (SARC) 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 SARC peer review is the cornerstone of the Northeast Stock Assessment Workshop (SAW) process, which includes assessment development, and report preparation (which is done by SAW Working Groups or Atlantic States Marine Fisheries Commission (ASMFC) technical committees), assessment peer review (by the SARC), public presentations, and document publication. This review determines whether or not the scientific assessments are adequate to serve as a basis for developing fishery management advice. Results provide the scientific basis for fisheries within the jurisdiction of NOAA's Greater Atlantic Regional Fisheries Office (GARFO).

The purpose of this meeting will be to provide an external peer review of a benchmark stock assessment for **Sea scallop and Atlantic herring**. The requirements for the peer review follow. This Statement of Work (SOW) also includes: **Appendix 1**: TORs for the stock assessment, which are the responsibility of the analysts; **Appendix 2:** a draft meeting agenda; **Appendix 3:** Individual Independent Review Report Requirements; and **Appendix 4:** SARC 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 SARC 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 SARC 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 SOW, OMB Guidelines, and the TORs below. All TORs must be addressed in each reviewer's report. No more than one of the reviewers selected for this review is permitted to have served on a SARC panel that reviewed this same species in the past. The reviewers shall have working knowledge and recent experience in the application of modern fishery stock assessment models. Expertise should include forward projecting statistical catch-at-age (SCAA) models. Reviewers should also have experience in evaluating measures of model fit, identification, uncertainty, and forecasting. Reviewers should have experience in development of Biological Reference Points (BRPs) that includes an appreciation for the varying quality and quantity of data available to support estimation of BRPs. For scallops, knowledge of sessile invertebrates, length-structured models, and spatial management would be desirable. For herring, knowledge of migratory pelagic species and SCAA models would be useful.

Tasks for Reviewers

- Review the background materials and reports prior to the review meeting
- 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 SOW and TORs, in adherence with the required formatting and content guidelines; reviewers are not required to reach a consensus.
- Each reviewer shall assist the SARC Chair with contributions to the SARC Summary Report
- Deliver individual Independent Review Reports to the Government according to the specified milestone dates
- This report should explain whether each stock assessment Term of Reference of the SAW was or was not completed successfully during the SARC meeting, using the criteria specified below in the "Tasks for SARC 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 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 SARC Summary Report on specific stock assessment Terms of Reference or on additional questions raised during the meeting.

Tasks for SARC panel

- During the SARC meeting, the panel is to determine whether each stock assessment Term of Reference (TOR) of the SAW 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 SARC chair shall identify or facilitate agreement among the reviewers for each stock assessment TOR of the SAW.
- 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 SOW and Schedule of Milestones and Deliverables below.

Tasks for SARC chair and reviewers combined:

Review both the Assessment Report and the draft Assessment Summary Report. The draft Assessment Summary Report is reviewed and edited to assure that it is consistent with the outcome of the peer review, particularly statements about stock status recommendations and descriptions of assessment uncertainty.

The SARC Chair, with the assistance from the reviewers, will write the SARC Summary Report. Each reviewer and the chair will discuss whether they hold similar views on each stock assessment 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 SAW. For terms where a similar view can be reached, the SARC Summary Report will contain a summary of such opinions. In cases where multiple and/or differing views exist on a given Term of Reference, the SARC Summary Report will note that there is no agreement and will specify - in a summary manner – what the different opinions are and the reason(s) for the difference in opinions.

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

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

Foreign National Security Clearance

When reviewers participate during a panel review meeting at a government facility, the NMFS Project Contact is responsible for obtaining the Foreign National Security Clearance approval for reviewers who are non-US citizens. For this reason, the reviewers shall provide requested information (e.g., first and last name, contact information, gender, birth date, country of birth, country of citizenship, country of permanent residence, country of current residence, dual citizenship (yes, no), passport number, country of passport, travel dates.) to the NEFSC SAW Chair for the purpose of their security clearance, and this information shall be submitted at least 30 days before the peer review in accordance with the NOAA Deemed Export Technology Control Program NAO 207-12 regulations available at the Deemed Exports NAO website: http://deemedexports.noaa.gov/ and http://deemedexports.noaa.gov/compliance_access_control_procedures/noaa-foreign-national-registration-system.html. The contractor is required to use all appropriate methods to safeguard Personally Identifiable Information (PII).

Place of Performance

The place of performance shall be at the contractor's facilities, and at the Northeast Fisheries Science Center in Woods Hole, Massachusetts.

Period of Performance

The period of performance shall be from the time of award through August 17, 2018. Each reviewer's duties shall not exceed **16** 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.

No later than May 21, 2018	Contractor sends reviewer contact information to the COR, who then sends this to the NMFS Project Contact
No later than June 12, 2018	NMFS Project Contact will provide reviewers the pre-review documents
June 26-29, 2018	Each reviewer participates and conducts an independent peer review during the panel review meeting in Woods Hole, MA
June 29, 2018	SARC Chair and reviewers work at drafting reports during meeting at Woods Hole, MA, USA
July 13, 2018	Reviewers submit draft independent peer review reports to the contractor's technical team for review
July 13, 2018	Draft of SARC Summary Report, reviewed by all reviewers, due to the SARC Chair *
July 20, 2018	SARC Chair sends Final SARC Summary Report, approved by reviewers, to NMFS Project contact (i.e., SAW Chairman)
July 27, 2018	Contractor submits independent peer review reports to the COR and technical point of contact (POC)
Aug. 3, 2018	The COR and/or technical POC distributes the final reports to the NMFS Project Contact and regional Center Director

* The SARC 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

All travel expenses shall be reimbursable in accordance with Federal Travel Regulations (http://www.gsa.gov/portal/content/104790). International travel is authorized for this contract. Travel is not to exceed $12,000.

Restricted or Limited Use of Data

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

NMFS Project Contact

Dr. James Weinberg, NEFSC SAW Chair

Northeast Fisheries Science Center

166 Water Street, Woods Hole, MA 02543

James.Weinberg@noaa.gov Phone: 508-495-2352

Appendix 1. ***Stock Assessment Terms of Reference for SAW/SARC-65***

T*he SARC Review Panel shall assess whether or not the SAW Working Group has reasonably and satisfactorily completed the following actions.*

A. Sea scallop

1. Estimate catch from all sources including landings, discards, and incidental mortality. Describe the spatial and temporal distribution of landings, discards, and fishing effort. Characterize the uncertainty in these sources of data.
2. a. Present the survey data being used in the assessment (e.g., regional indices of relative or absolute abundance, recruitment, size data, etc.). Characterize the uncertainty and any bias in these sources of data.
3. Summarize existing data, and characterize trends if possible, and define what data should be collected from the Gulf of Maine area to describe the condition and status of that resource. If possible provide a basis for developing catch advice for this area.
4. Investigate the role of environmental and ecological factors in determining stock distribution and recruitment success. If possible, integrate the results into the stock assessment.
5. Estimate annual fishing mortality, recruitment and stock biomass for the time series, and estimate their uncertainty. Report these elements for both the combined resource and by sub-region. Include retrospective analyses (historical, and within-model) to allow a comparison with previous assessment results and previous projections.
6. State the existing stock status definitions for "overfished" and "overfishing". Then update or redefine biological reference points (BRPs; 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. Comment on the scientific adequacy of existing BRPs and the "new" (i.e., updated, redefined, or alternative) BRPs.
7. Make a recommendation[a] about what stock status appears to be based on the existing model (from previous peer reviewed accepted assessment) and based on a new model or model formulation developed for this peer review.
    a. Update the existing model with new data and evaluate stock status (overfished and overfishing) with respect to the existing BRP estimates.
    b. Then use the newly proposed model and evaluate stock status with respect to "new" BRPs and their estimates (from TOR-5).
    c. Include descriptions of stock status based on simple indicators/metrics.
8. Develop approaches and apply them to conduct stock projections.
    a. Provide numerical annual projections (through 2020) and the statistical distribution (i.e., probability density function) of the catch at F_{MSY} or an F_{MSY} proxy (i.e. the overfishing level, OFL) (see Appendix to the SAW TORs). Each projection should estimate and report annual probabilities of exceeding threshold BRPs for F, and probabilities of falling below threshold BRPs for biomass. Use a sensitivity analysis approach in which a range of assumptions about the most important uncertainties in

the assessment are considered (e.g., terminal year abundance, variability in recruitment).

b. Comment on which projections seem most realistic. Consider the major uncertainties in the assessment as well as sensitivity of the projections to various assumptions. Identify reasonable projection parameters (recruitment, weight-at-age, retrospective adjustments, etc.) to use when setting specifications.

c. Describe this stock's vulnerability (see "Appendix to the SAW TORs") to becoming overfished, and how this could affect the choice of ABC.

9. Review, evaluate and report on the status of the SARC and Working Group research recommendations listed in most recent SARC reviewed assessment and review panel reports. Identify new research recommendations.

[a]NOAA Fisheries has final responsibility for making the stock status determination based on best available scientific information.

B. Atlantic herring

1. Estimate catch from all sources including landings and discards. Describe the spatial and temporal distribution of landings, discards, and fishing effort. Characterize uncertainty in these sources of data. Comment on other data sources that were considered but were not included.

2. Present the survey data being used in the assessment (e.g., regional indices of abundance, recruitment, state surveys, age-length data, food habits, etc.). Characterize the uncertainty and any bias in these sources of data.

3. Estimate consumption of herring, at various life stages. Characterize the uncertainty of the consumption estimates. Address whether herring distribution has been affected by environmental changes.

4. Estimate annual fishing mortality, recruitment and stock biomass (both total and spawning stock) for the time series, and estimate their uncertainty. Incorporate ecosystem information from TOR-3 into the assessment model, as appropriate. Include retrospective analyses (both historical and within-model) to allow a comparison with previous assessment results and projections, and to examine model fit.

5. State the existing stock status definitions for "overfished" and "overfishing". Then update or redefine biological reference points (BRPs; 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. Comment on the scientific adequacy of existing BRPs and the "new" (i.e., updated, redefined, or alternative) BRPs.

6. Make a recommendation[a] about what stock status appears to be based on the existing model (from previous peer reviewed accepted assessment) and based on a new model or model formulation developed for this peer review.

   a. Update the existing model with new data and evaluate stock status (overfished and overfishing) with respect to the existing BRP estimates.

   b. Then use the newly proposed model and evaluate stock status with respect to "new" BRPs and their estimates (from TOR-5).

   c. Include descriptions of stock status based on simple indicators/metrics.

7. Develop approaches and apply them to conduct stock projections.

    a. Provide numerical annual projections (through 2021) and the statistical distribution (i.e., probability density function) of the catch at F_{MSY} or an F_{MSY} proxy (i.e. the overfishing level, OFL) (see Appendix to the SAW TORs). Each projection should estimate and report annual probabilities of exceeding threshold BRPs for F, and probabilities of falling below threshold BRPs for biomass. Use a sensitivity analysis approach in which a range of assumptions about the most important uncertainties in the assessment are considered (e.g., terminal year abundance, variability in recruitment).

    b. Comment on which projections seem most realistic. Consider the major uncertainties in the assessment as well as sensitivity of the projections to various assumptions. Identify reasonable projection parameters (recruitment, weight-at-age, retrospective adjustments, etc.) to use when setting specifications.

    c. Describe this stock's vulnerability (see "Appendix to the SAW TORs") to becoming overfished, and how this could affect the choice of ABC.

8. If possible, make a recommendation about whether there is a need to modify the current stock definition for future assessments.

9. For any research recommendations listed in SARC and other recent peer reviewed assessment and review panel reports, review, evaluate and report on the status of those research recommendations. Identify new research recommendations.

[a]NOAA Fisheries has final responsibility for making the stock status determination based on best available scientific information.

Clarification of Terms

used in the Stock Assessment Terms of Reference

Guidance to SAW WG about "Number of Models to include in the Assessment Report":

In general, for any TOR in which one or more models are explored by the WG, 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 WG 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 Stock Assessment Working Group:

Anyone participating in SAW 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}

65th Stock Assessment Workshop/Stock Assessment Review Committee (SAW/SARC) Benchmark stock assessment for A. Sea scallop and B. Herring

June 26-29, 2018

Stephen H. Clark Conference Room – Northeast Fisheries Science Center

Woods Hole, Massachusetts

DRAFT AGENDA* (version: Dec. 22, 2017)

TOPIC	PRESENTER(S)	SARC LEADER	RAPPORTEUR

Tuesday, June 26

10 – 10:45 AM

Welcome/Description of Review Process **James Weinberg**, SAW Chair

Introductions/Agenda **TBD,** SARC Chair

Conduct of Meeting

10:45 – 12:45 PM Assessment Presentation (A. Scallops)

Dvora Hart **TBD**

12:45 – 1:45 PM Lunch

1:45 – 3:45 PM Assesssment Presentation (A. Scallops)

Dvora Hart **TBD**

3:45 – 4 PM Break

4 – 5:45 PM SARC Discussion w/ Presenters (A. Scallops)

TBD, SARC Chair **TBD**

5:45 – 6 PM Public Comments

Wednesday, June 27

8:30 – 10:30 AM Assessment Presentation (B. Herring)
Jon Deroba **TBD**

10:30 – 10:45 AM Break

10:45 – 12:30 PM Assessment Presentation (B. Herring)
Jon Deroba **TBD**

12:30 – 1:30 PM Lunch

1:30 – 3:30 PM SARC Discussion w/presenters (B. Herring)
TBD, SARC Chair **TBD**

3:30 – 3:45 PM Public Comments

3:45 -4 PM Break

4 – 6 PM Revisit with Presenters (A. Scallops)
TBD, SARC Chair **TBD**

7 PM **(**Social Gathering**)**

Thursday, June 28

8:30 – 10:30 Revisit with Presenters (B. Herring)
TBD, SARC Chair **TBD**

10:30 – 10:45 Break

10:45 – 12:15 Review/Edit Assessment Summary Report (A. Scallops)
TBD, SARC Chair **TBD**

12:15 – 1:15 PM Lunch

1:15 – 2:45 PM (cont.) Edit Assessment Summary Report (A. Scallops)
TBD, SARC Chair **TBD**

2:45 – 3 PM Break

3 – 6 PM Review/edit Assessment Summary Report (B. Herring)
TBD, SARC Chair **TBD**

Friday, June 29

9:00 AM – 5:00 PM SARC Report writing

*All times are approximate, and may be changed at the discretion of the SARC 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 SARC.

Appendix 3. Individual Independent Peer Review Report Requirements

1. The independent peer review 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 SARC 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 SARC 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 Statement of Work

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

Appendix 4. SARC Summary Report Requirements

1. The main body of the report shall consist of an introduction prepared by the SARC 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 SARC. Following the introduction, for each assessment reviewed, the report should address whether or not each Term of Reference of the SAW Working Group was completed successfully. For each Term of Reference, the SARC Summary Report should state why that Term of Reference was or was not completed successfully.

   To make this determination, the SARC chair and reviewers should consider whether or not the work provides a scientifically credible basis for developing fishery management advice. If the reviewers and SARC 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 SAW, and relevant papers cited in the SARC Summary Report, along with a copy of the CIE Statement of Work.

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

Appendix 3: Review Panel Membership

- Professor Patrick Sullivan, Cornell University (Council member and Panel chair)
- Dr Cathy Dichmont, independent consultant (CIE reviewer).
- Dr Geoff Tingley, independent consultant (CIE reviewer).
- Dr Coby Needle, Marine Scotland Science, Aberdeen, Scotland (CIE reviewer).