CENTER FOR INDEPENDENT EXPERTS (CIE) INDEPENDENT PEER REVIEW OF THE BENCHMARK STOCK ASSESSMENT FOR SEA SCALLOP AND ATLANTIC HERRING (SAW/SARC65)

Dr Cathy Dichmont

Cathy Dichmont Consulting
cathydichmont@gmail.com

Disclaimer

The author(s) do not warrant that the information in this document is free from errors or omissions. The authors do not accept any form of liability, be it contractual, tortious, or otherwise, for the contents of this document or for any consequences arising from its use or any reliance placed upon it. The information, opinions and advice contained in this document may not relate, or be relevant, to a readers particular circumstances. Opinions expressed by the authors are the individual opinions expressed by those persons and are not necessarily those of the publisher, research provider or the funding agency.

Contact Details

Name: Dr Catherine Dichmont
Address: 69 Headsail Drive, Banksia Beach, 4507, Queensland, Australia
Email: cathydichmont@gmail.com

Center for Independent Experts (CIE) independent peer review of the benchmark stock assessment for sea scallop and Atlantic herring (SAW/SARC65)

1 CONTENTS

2 Executive Summary 10

2.1 Sea scallops 10

2.2 Atlantic herring 12

3 Background 14

4 Description of the Individual Reviewer's Role in the Review Activities 15

4.1 Sea scallop 15

4.2 Atlantic herring 17

5 Summary of Findings for each ToR in which the weaknesses and strengths are described 19

5.1 Scallops 19

5.1.1 ToR 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. 19

5.1.1.1 Overall response 19

5.1.1.2 Background 19

5.1.1.3 Landings and size-frequency 19

5.1.1.4 Discards 20

5.1.1.5 Uncertainty 20

5.1.2 Tor 2: 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. 21

5.1.2.1 Overall response 21

5.1.2.2 Background 21

5.1.2.3 Dredge survey efficiency 21

5.1.2.4 Habcam 23

5.1.3 Tor 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............................... 24

5.1.3.1 Overall response 24

5.1.3.2 Background 24

5.1.3.3 Assessment and projection options 24

5.1.4 Tor 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. 26

5.1.4.1 Overall response 26

5.1.4.2 Background 26

5.1.4.3 Environmental and ecological factors 26

5.1.5 ToR 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. 27

5.1.5.1 Overall response 27

5.1.5.2 Background 27

5.1.5.3 CASA in general 27

5.1.5.4 Growth 29

5.1.5.5 Conversions 30

5.1.5.6 Natural mortality 30

5.1.5.7 Alternative approaches 31

5.1.6 ToR 6: 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. 31

5.1.6.1 Overall response 31

5.1.6.2 Background 31

5.1.6.3 General comments 31

5.1.7 ToR 7: 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. 33

5.1.7.1 Overall response 33

5.1.7.2 Background 33

5.1.7.3 General comments 33

5.1.8 ToR 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 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. 34

5.1.8.1 Overall response 34

5.1.8.2 Background 34

5.1.8.3 General comments 35

5.1.9 ToR 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. 37

5.1.9.1 Overall response 37

5.1.9.2 Background 37

5.1.9.3 General comments 37

5.2 Atlantic herring 39

5.2.1 ToR 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. 39

5.2.1.1 Overall response 39

5.2.1.2 Background 39

5.2.1.3 General comments 40

5.2.2 ToR 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. 40

5.2.2.1 Overall response 40

5.2.2.2 Background 40

5.2.2.3 General comments 41

5.2.3 ToR 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. 42

5.2.3.1 Overall response 42

5.2.3.2 Background 42

5.2.3.3 General comments 42

5.2.4 ToR 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. 43

5.2.4.1 Overall response 44

5.2.4.2 Background 44

5.2.4.3 General comments 44

5.2.4.4 Natural mortality and retrospective patterns 44

5.2.4.5 Alternative assessments 46

5.2.5 ToR 5: 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. 46

5.2.5.1 Overall response 46

5.2.5.2 Background 47

5.2.5.3 General comments 47

5.2.6 ToR 6: 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. 47

5.2.6.1 Overall response 47

5.2.6.2 Background 47

5.2.6.3 General comments 48

5.2.7 ToR 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 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. 48

5.2.7.1 Overall response 49

5.2.7.2 Background 49

5.2.7.3 General comments 49

5.2.8 Tor 8: If possible, make a recommendation about whether there is a need to modify the current stock definition for future assessments. 49

5.2.8.1 Overall response 49

5.2.8.2 Background 50

5.2.8.3 General comments 50

5.2.9 ToR 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.. 51

5.2.9.1 Overall response 51

5.2.9.2 Background 51

5.2.9.3 General comments 51

6 Conclusions and Recommendations in accordance with the ToRs. 53

6.1 Sea scallops 53

6.1.1 High priority recommendations... 53

6.1.2 Medium priority recommendations.. 54

6.2 Atlantic herring ... 55

6.2.1 High priority recommendations ... 55

6.2.2 Medium priority recommendations ... 56

7 References ... 58

8 Review process ... 59

9 Appendix 1: Bibliography of materials provided for review ... 60

9.1 Scallop references ... 60

9.2 Herring references ... 61

9.3 Working Papers ... 61

10 Appendix 2: A copy of the CIE Statement of Work ... 62

11 Appendix 3: Panel Membership or other pertinent information from the panel review meeting... 79

2 EXECUTIVE SUMMARY

The review workshop for the sea scallop and Atlantic herring took place in Woods Hole, MA on June 26-29, 2018 with review panel members Drs Sullivan (SARC chair), Dichmont, Needle and Tingley, the lead scallop (Drs Hart and Chang) and herring assessors (Dr Deroba), industry members and other scientists involved in the stock assessment, data collection and management. The review documents and background information were provided from a web-based server before and during the review. Several very insightful presentations were provided during the review, with questions well received and responded to by the teams. These discussions greatly contributed to the reviewers' knowledge base. The assessors are thanked for their high level of professionalism during the review.

Review documents and presentations were structured to address each Term of Reference (ToR) in turn, which greatly enhanced the review process.

2.1 SEA SCALLOPS

All the ToRs were fully met.

Three different models were applied for the full benchmark assessment. Much of the need for the three types of model arises from the unique nature of the resource where, on the one-hand, management is highly spatial given the nature of the resource and, on the other hand, there is a policy and legislative need to estimate equilibrium-based reference points based on robust stock and recruitment derived parameters. Each of the three models are used for separate well-defined purposes:

- the CASA model estimates historical biomass and fishing mortality rates at a regional scale;
- the SYM model estimates regional and whole stock biological reference points based on CASA outputs;
- the SAMS model forecasts future abundance, biomass and landings at a finer spatial scale to address day-to-day management needs.

The important issues are:

a. Whether each model is coherent with the other models and correctly implemented. Although some inconsistencies are highlighted, the structure of each model is similar and is coherent where required. Thus, this approach is reasonable and supported.

b. Whether a fully spatial assessment model could be produced that would satisfy each of the above three purposes in a single model. At this stage, this is not available and would require a concerted research effort, which may not be entirely successful. However, given the nature of the data and the length of time this collection has been undertaken, a way forward of potentially bringing these different approaches into a single framework using a stepwise approach is discussed.

This review concludes that the implementation of the assessment models is of a high standard and well implemented. The new aspects to the models have been effective at enhancing the models and their associated results. These can be directly used for management advice.

It is agreed that the healthy stock biomass levels in recent years have been due to two very large cohorts (the 2012-year class on Georges Bank, primarily located in the Nantucket Lightship Area, and the 2013-year class in the Mid-Atlantic, much of which is in the Elephant Trunk rotational area off Delaware Bay). When these dynamics are projected by SAMS, forecasts are uncertain, since these very high densities of scallops have rarely been observed. Despite this uncertainty, the projected total biomass and landings is correctly predicted to decline as these strong year classes are fished or die naturally.

The assessment generally indicates that a) there are high levels of unaccounted for uncertainty in the dredge surveys; b) there may be some biases in the optical surveys when there are high densities of small scallops, c) there may be density dependent juvenile mortality for some regions, d) there is still model mis-specification in the mid-Atlantic given runs of model residuals, and e) there is some confounding between natural mortality, observation error and all forms of fishing mortality. Whether this is a feature of the assessment, the data or the resource are key future research needs. Suggested ways forward are provided.

For the first time in the scallop assessment, an index of spawning stock biomass, gonad weight, was introduced. This was proposed to be used in conjunction with stock biomass estimates based on meat weight. Although the concept behind the introduction of gonad weight to describe spawning stock biomass has merit, the full implications of using this approach have not been fully investigated. It is recommended that there be further development of the gonad-based spawning stock biomass metrics before full implementation is undertaken. These include updating the shell height to gonad weight relationships (especially for areas where these are not available or out of date, e.g., the southeast Nantucket Lightship area) and evaluating the relative impact of these changes by region.

Three potential assessment options were reviewed for the Gulf of Maine. These are discussed with some additional approaches suggested.

2.2 Atlantic herring

All ToRs were fully met.

The previously used Age Structured Assessment Program (ASAP) assessment has been enhanced to address past retrospective patterns. These are generally much reduced for this benchmark assessment. The analysis was thorough and well set out - several models were explored, and different types of retrospective and sensitivity analyses were conducted anticipating many of the questions likely to be raised during the review process. Diagnostic tests and profile likelihoods were also provided. The impact of the removal of key data sets was also explored. These thorough tests explored the strengths and weaknesses of ASAP's application to Atlantic herring. The assessment is of a high quality, suitable to be used for management advice.

Great thought had been put into how best to address the appearance or not of past retrospective patterns. It is supported that consideration of the presence and absence of retrospective patterns is an important aspect of this assessment. The assessors treated this benchmark assessment as a means to revisit most of the key model structures and decisions. This assessment therefore has major changes that include a) addition of an acoustic series, b) running the age composition to 8+ rather than 9+ given reliability issues of older ages in previous assessments, c) implementing the initial abundances at age differently so that a likelihood penalty as well as both age- and time- invariant natural mortality could be removed, d) treating selectivity differently, e) removal of age 1 catches from some surveys, and f) modelling annual recruitment deviations unconstrained. Many of the retrospective patterns seem to have been removed by the changes, although the stability of this benchmark base model still needs to be proven. These changes are supported.

Retrospective patterns in the new benchmark assessment were reasonable, thus obviating the need for time varying natural mortality. Furthermore, unlike previous assessments, consumption calculations were not used to scale natural mortality as the results did not support changing the natural mortality parameter based on consumption information. No Mohn's Rho adjustment was undertaken of the results, because the values fell within the 80% confidence intervals. These decisions are supported.

Of importance is that the catchability estimated between the two recent periods within the model is quite different to the field derived calibration values previously used. This could be a sign that another factor is being aliased in this catchability term. This issue needs to be addressed for the next benchmark assessment to avoid the potential return of residual patterns.

New reference points were provided in a Kobe plot and are supported. However, the projections and reference points assume mean recruitment would re-establish whereas recent runs have estimated low recruitment (albeit the last two years with large uncertainty). Given the present assessment results, the current stock status may be influenced in the future with the low recruitments and poor year classes moving through the system over time.

A two-stock Stock Synthesis model was developed to address potential stock structure issues. Another model, Stock Assessment Model (SAM) was also developed, but was in its initial stages of development. This work has merit as alternative assessment tools are important in assessments where strong retrospective patterns have (or do) existed. Undertaking a single-stock Stock Synthesis model run is also recommended, given length information can also be included in this model.

3 BACKGROUND

Sea scallop are caught on the Georges Bank and the Mid-Atlantic Gulf of Maine, with a small but locally important fishery in the Gulf of Maine. A key characteristic of scallops is that they are spatially variable in terms of, for example, their recruitment, growth and natural mortality.

Three separate, but interlinked models are used for the assessment. Each of these fulfill specific roles. The first is a forward-projecting size-based assessment (a.k.a. CASA) with time varying growth and natural mortality, which are used to estimate historical recruitment, stock size, fishing mortality and various dynamic parameters such as natural mortality. The model is applied to three geographical regions, Georges Bank open and closed, and the Mid-Atlantic. The second model (a.k.a. SYM) uses output from CASA to estimate a stock-recruitment relationship and biological reference points for each of these three regions, and to determine stock status. The third model (a.k.a. SAM), is more spatially explicit and uses the most recent survey abundance maps to forward project the impact of fishing over time. It can also evaluate the effect of area closures and re-openings over time.

Several survey indices could be applied to the assessment, the main being a lined scallop dredge survey, a drop camera survey, and a towed camera (Habcam) survey. At-sea and dockside observers also collect numerous forms of data from the fishing vessels and from the surveys.

The Age Structured Assessment Program (ASAP) model was applied to the Atlantic herring resource. Herring are caught on the Gulf of Maine and Georges Bank. This complex is composed of several spawning aggregations where the fishery may be catching fish from a mix of spawning aggregations. The different gear types are combined into two, being mobile and fixed gears. Consumption information is used to derive a predation index, but this index is sensitive to the predators included in the analyses. The scale of herring consumption is used to determine whether model estimated mortality is within the right order of magnitude.

Several abundance indices are used to tune the model, most of these from various bottom trawl surveys. An acoustic index is applied for the first time.

4 DESCRIPTION OF THE INDIVIDUAL REVIEWER'S ROLE IN THE REVIEW ACTIVITIES

The SAW/SARC65 review of the sea scallop and Atlantic herring benchmark assessments took place in Woods Hole, MA from June 26-29. In attendance were review panel members Drs Sullivan (SARC chair), Dichmont, Needle and Tingley. Also present were the stock assessment teams, industry and other scientists involved in the stock assessment and management were in attendance. The review was undertaken in a very co-operative light with requests for additional work met. The teams are thanked for a very constructive meeting.

The panel members were provided with material pertaining to the assessments with several Appendices that relate to various aspects of the assessment, and several relevant items of background material (see Appendix 1: Bibliography of materials provided for review, for a list of documents provided). Presentations were also provided covering each Term or Reference. Public comment was also provided by fishers and scientists who gave valuable background and inputs on the various assessments.

Several additional sensitivity tests, further plots, recalculating the biological reference for sea scallops using meat weight rather than gonad weight and prioritizing the research recommendations were undertaken and provided on request.

A panel report was written, and panel views were discussed during the review. An individual reviewer's report was provided addressing each Term of Reference (ToR), being:

4.1 TERMS OF REFERENCE FOR THE SEA SCALLOP ASSESSMENT

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. 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.

4.2 Terms of Reference for the Atlantic herring Assessment

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.

5 Summary of Findings for each ToR in which the weaknesses and strengths are described

5.1 Scallops

5.1.1 ToR 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.

5.1.1.1 Overall response

This TOR was fully met.

5.1.1.2 Background

The assessment model relies heavily on landings data. These are divided into the regions used in the model, being the mid-Atlantic (MA) and the Georges Bank open access (GBA) and closed (GBC) areas. Although Landings per unit Effort (LPUE) are calculated, these are not used in the assessment, but are used to calculate total fishing effort. Discards are not modelled directly in the model, but included as a general (flat percentage) incidental mortality term.

Although uncertainty is discussed, uncertainty is not described numerically. Based on the discussion during the review where sources of uncertainty were further discussed, this is likely not a major issue and therefore this component of the ToR is met. Some of this uncertainty is likely to fall within the natural mortality term. As a result, some suggestions are provided below to further expand investigation of commercial data uncertainty.

5.1.1.3 Landings and size-frequency

Two forms of landings are excluded in the assessment – those in the Gulf of Maine and in Canadian waters. The Gulf of Maine landings are small and, although significant for the area, are not significant enough to affect the assessment if they were included in the other regions (beyond adding to uncertainty). The Gulf of Maine is treated separately, which is appropriate (see ToR 3).

A source of uncertainty not fully discussed (it was not part of the ToR and so appropriate) was the effect of the fishery in Canada on the Georges Bank stock. In some years, this catch was a large proportion of the total catch. Despite the practicalities of this being managed and fished differently, ideally, one would assess these as a single region or as an additional region. This should be investigated at the very least as a sensitivity test in future assessments.

Recommendation (Medium) 1. Investigate the impact of the Canadian sector on Georges Bank as a sensitivity test in the CASA and SYM models.

5.1.1.4 Discards

Discards are not directly modelled in CASA, but included in a broadly defined incidental mortality term (which normally only includes mortality from the interaction with the fishing gear). However, SYM and CASA explicitly model discarding. Smaller scallops are discarded compared to the fishery (Fig 5.6a). CASA, and SYM and SAMS therefore model discards differently. For example, as an ad hoc way of accounting for this model structure difference, the value of the fraction of scallops that suffer incidental mortality (c in Equation A5.1) is a different value in CASA, compared to that used in SYM and SAMS models. Given the low level of discards, this is probably not a major influence on the model results, but should be corrected for the next assessment. It is better to set up all the different (sequential) models consistently. The approach used in SYM and SAMS is more intuitive.

Recommendation (Medium) 2. Include discard mortality as a separate term within the CASA model with a lower selectivity as shown by the relative size frequency distributions. Set these up similarly to those in SYM and SAMS.

5.1.1.5 Uncertainty

Uncertainty in the commercial data was not numerically presented. It is likely that quantification is not possible in its entirety. However, several sources of uncertainty were described, for example poaching and highgrading. The scale of these issues was not fully discussed and could have been expanded upon further through investigation of apprehension data. Clearly this is important enough that, in the GBC areas, unaccounted for mortality such as poaching was used as a motivation to decrease the natural mortality index from 0.23 to 0.2. Although acknowledged in the methods, these different sources of uncertainty, not accounted for in the model, are aliased into the natural mortality term, which in the longer term is not the best approach. Similarly, any annual changes to incidental mortality would be treated as natural mortality. To counteract this aliasing, several *ad hoc* adjustments to, for example, natural mortality or the fraction of scallop caught by the gear are undertaken.

Recommendation (Medium) 3. Investigate alternative data sources such as apprehensions to define the scale of poaching and highgrading and test how this affects estimates of natural mortality regionally.

Recommendation (Medium) 4. Investigate all major sources of fishing mortality that are aliased as not natural mortality and either a) address these through model structure changes or b) provide evidence that the present ad hoc adjustments are sufficient.

5.1.2 Tor 2: 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.

5.1.2.1 Overall response

This TOR was fully met.

5.1.2.2 Background

Several sources of survey indices are available – the most important being the NEFSC dredge, VIMS dredge using commercial vessels, SMAST drop camera video and the NOAA/NEFSC Habcam surveys. The (single vessel) NEFSC dredge surveys have been undertaken by two different vessels over time (the changeover being in 2008). Studies were used, where appropriate, to adjust for differences between the vessels or gear type. There were strong bases for these adjustments. There has been good agreement between the three surveys in the Georges Bank and Mid-Atlantic until recent years. In some models, GAM models were used to interpolate to low surveyed strata.

Uncertainties in the survey indices are well described.

5.1.2.3 Dredge survey efficiency

To address the discrepancy between the optical surveys and the dredge surveys from 2015 to 2017, dredge abundance was inflated by a factor of three in the high-density areas based on, amongst other factors, paired Habcam/dredge tows. However, the plot of dredge catch or dredge swept area relative to the Habcam density (Figure A6.6) does not provide comprehensive evidence of the value of a three-fold increase. From this plot, at the highest densities, values of 2 to 2.4 can be inferred. The basis for a value of 3 for the high-density dredge efficiency adjustment is not well substantiated in the report. In reality, this is much more likely to be uncertain and as such, sensitivity tests of this parameter should be undertaken in the assessment.

Recommendation (Medium) 5. Undertake sensitivity tests within CASA of different dredge efficiencies until more studies can be used to derive this value with greater certainty.

Despite these adjustments, there are still signs of issues, mainly with the dredge survey. For example, on many occasions, the CASA assessment results show that the dredge index may have large observation error (see ToR 4 for a further discussion on this). Although this may be a result of several issues, two main reasons should be addressed:

Firstly, the stratified random survey is analysed according to classic data-based approaches given its design. It uses the original strata that may over time have become less relevant or new data have become available that would allow for much better strata allocations today. Although restratification is not to be desired, there are occasions where this should be considered. This seems to be the case here, where the dredge survey results are uncertain and may also be differentially biased. Restratification, but consistently undertaken, should be considered for both the historical and futures surveys. This restratification should apply to both the indices and size-frequencies.

However, an alternative and more favoured approach would be to develop model-based methods to developing the indices and size-frequencies. Already used model-based approaches to the data on time and space analyses of, for example, growth or index interpolation show that there are sufficient data to undertake this type of analysis. Several forms of analyses are available, including GAMs and kriging. The most appropriate model should be tested, both on the data themselves (using training and test datasets), and through simulations. This model-based approach should be compared to the present stratification and restratified results.

It is recommended that the historical dataset is investigated in the first case. The reason is that these results would help define how best to move forward. For example, the results will show whether the original sites can be maintained with restratification, or whether there is a need to add more sites in specific regions, statically or adaptively.

Recommendation (High) 1. Undertake an extensive historical analysis of the dredge survey index and size-frequency data investigating the potential use of a) restratification and b) model-based approaches.

Recommendation (High) 2. Use the results from the historical analyses as a basis for redesign, if required, of the dredge survey.

Secondly, the dredge efficiency adjustment may not be appropriate or the relationship between density and efficiency is more nuanced relative to density. The reality is that both the optical surveys and the dredge surveys are uncertain and, to different degrees, biased. This means there is no "truth". Consequently, some work on the Habcam as suggested below and further field work on survey relative efficiencies especially at high densities are indicated.

Recommendation (High) 3. Undertake further field work on dredge efficiency and its impact on the index of abundance and size frequency data.

5.1.2.4 Habcam

The approach of expanding the sub-samples of each tow to large scales is well described in the background material. Several approaches were tested on field data and simulations. The reviewer supports the final approach used.

As discussed during the review, the optical surveys may have different selectivities with varying size-specific densities. For example, at very high small scallop densities the reader may get a biased search image for smaller scallops. To address this issue, a) protocols and QA/QC procedures should be developed that address this issue. Furthermore, given the importance of this survey index, b) a percentage of the photographs should be double counted and double measured. Finally, c) improved training of annotators on count, shell height and dead animals should be undertaken.

Recommendation (Medium) 6. Develop Habcam analysis protocols and QA/QC protocols that are aimed at reducing the risk of density dependent biases in the estimates of density and size-frequency.

Recommendation (High) 4. Consider double counting of a percentage of the Habcam video photographs for abundance and size-frequency.

Recommendation (High) 5. Improve training of annotators of shell count, size and dead scallops especially at high and low densities.

5.1.3 Tor 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.

5.1.3.1 Overall response

This TOR was fully met.

5.1.3.2 Background

The Gulf of Maine (GoM) is a patchy resource area that does not have a consistent scallop dedicated survey time series. The landings in the area are not large relative to the other main scallop areas, but these catches are regionally important. A few approaches were used to set past TACs, including using historical catch, exploitation rates applied to survey estimates and forward projecting survey data.

5.1.3.3 Assessment and projection options

Several GoM scallop surveys (dredge and drop camera) have been undertaken since 2009, but these have concentrated on different regions over time. Only two areas have been reasonably consistently sampled, but not in 2017. Although these different surveys would have had merit for biological sampling and understanding where the resource can be found, this does not assist in developing a consistent index of abundance for an assessment.

Three options for setting TAC were proposed for possible use in the GoM:

a. expanding the CASA model to include the GoM area, and estimating Yield per Recruit (SYM);

b. expanding SAMS to cover a portion of the GoM region; and

c. other approaches such as depletion analyses.

Given the lack of a survey index time series, CASA would not be a recommended option unless there is large overlap with the data rich resource dynamics and recruitment between the regions (which seem at this stage unlikely). For this reason, the better option is to consider what additional inputs one would need to run the SAMS model to assist in identifying and prioritising what information is required. CASA is more likely the best longer-term option.

Recommendation (High) 6. Focus on initially using SAMS to prioritise data and information needs.

Given that SAMS requires a biomass distribution map or index, a socio-economic cost-benefit analysis of survey requirements per region within the GoM is recommended. The ideal would be regular and consistent whole of GoM surveys using the best (for scallops) survey technique (most likely the optical surveys) to develop an index of abundance time series, and use the industry platforms to obtain key biological data.

Recommendation (High) 7. Undertake a cost-benefit analysis of the survey requirements for the GoM with a focus on creating a tool to set TACs. This should include considering the ideal option which would be regular GoM wide surveys.

Another option that should be considered in the cost-benefit analysis is to undertake regular, but not annual, scallop specific surveys within the GoM and use this to project the TAC for the years where no survey will be undertaken. This means that the interval between the surveys would need to be considered in terms of adjusting for increased uncertainty. This option is not ideal, but may be cost-effective.

In the short term, there are several data moderate approaches that could be considered in place of CASA. These would initially need to rely on fishery dependent information for the longer-term index of abundance (LPUE) with the recent surveys providing additional information. Given the data rich nature of neighbouring regions, Bayesian hierarchical models are likely to perform well, e.g. a multi-stock Bayesian biomass dynamic hierarchical model (Zhou et al., 2009) or a Robin Hood approach modified for hard to age species (Punt et al., 2011). These are the recommended approaches as a first step towards ultimately running CASA.

Additionally, catch only (e.g., Carruthers et al., 2014) approaches that include survey data could be considered, but these usually have wide confidence intervals and can be quite uncertain.

Recommendation (High) 8. Investigate the use of data moderate models. Amongst these, concentrate on Bayesian approaches where hyper-priors can be informed by the data rich regions.

Similar to the other regions, model-based estimates of biomass indices and maps are likely to be beneficial.

Recommendation (Medium) 7. Move towards model-based estimates of biomass indices and maps as an option. This may be best to be either undertaken at the same time as the other regions or thereafter and so will need to be determined.

5.1.4 Tor 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.

5.1.4.1 Overall response

This ToR was fully met.

5.1.4.2 Background

Several environmental and ecological processes are described. These include scallop food supply, effects of temperature and pH, role of predators and competitors, and parasites and diseases that affect scallops. Two major studies are included that use space-time modelling to obtain scallop growth, and shell height to gonad/meat weight conversion parameters. These analyses showed that depth is a major factor. Impacts of climate change were also discussed. Evidence of a relationship between scallop recruitment and *Astropectin* biomass is provided.

5.1.4.3 Environmental and ecological factors

Environmental and ecological factors are well considered in the work and, where appropriate, used in the assessment. Examples were, a) the fact that phytoplankton supply decreases with depth is used to include depth (beyond location) in the growth and conversion models; b) the different growth periods in the CASA model address potentially different environmental periods; c) the addition of juvenile mortality to account for periods of high predation; d) explaining increased natural mortality in the 2011-2013 Georges Bank closed areas; and d) excluding the Southern Nantucket Light (SNL) area due to its unusually slow growth.

Most of this type of information was not directly included in the model, but mainly used to explain and support CASA model results. This approach is supported.

The use of space-time models to develop growth and conversion parameters is one step further in that these are directly used within the assessment. This approach is also supported.

Given the successful implementation of the space-time models for growth and the conversion factors, recommendations are made in ToR 5 and 9 on further improvements to SAMS which affects the use of CASA. However, in addition to expanding model-based methods to survey design and analyses, it is recommended that these approaches start investigating additional predictive environmental space-time correlates beyond depth and location.

Recommendation (Medium) 8. Investigate as part of model-based space-time analyses of survey biomass indices and maps whether the inclusion of environmental correlates add value to the predictions.

5.1.5 ToR 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.

5.1.5.1 Overall response

This ToR has been fully met.

5.1.5.2 Background

The assessment uses a forward projecting size-structured assessment model, CASA, based on that from Sullivan et al. (1990). Three separate regional CASA models were developed for each of GBC, GBA and the MA zones. Key to this assessment is that several survey indices are treated as absolute indices of abundance. A modification to previous assessments was the inclusion of time- and size-varying juvenile and adult natural mortality for the Georges Bank open and Mid-Atlantic areas. The inclusion of juvenile natural mortality was not required for the Georges Bank closed area. The use of the CASA model and its present formulation is supported. The southeast corner of southeast Nantucket Lightship was not assessed, since growth rates and potentially other life history parameters do not fit historical patterns.

5.1.5.3 CASA in general

There were several new aspects to the benchmark assessment. These have been well implemented and innovative. These include how natural mortality and growth were implemented.

- Natural mortality was estimated by year and by size (Georges Bank open and Mid-Atlantic), and for juveniles and adults separately (Georges Bank open and Mid-Atlantic).
- Growth included individual random effects on the growth rate (K) and asymptotic length at which growth is zero (L_∞). The results showed temporal patterns in growth rate deviations, which were included in the model as low to high growth rate time blocks.

A feature of the results is that the survey biomass indices are affected by the patchiness of the resource in some years, which the assessment at times interprets as being observation error beyond the calculated uncertainty. This effect is particularly noteworthy in the Georges Bank Closed and Open zones. Compared to the previous assessment, the inclusion of time-varying juvenile natural mortality has in part resolved this issue. These changes were mainly applied to address past underestimation of the survey indices in the model. The assessment was able to explain some increases in survey indices and subsequent substantial decreases through increased juvenile mortality (as supported by the size frequencies and indices in subsequent years; and extraneous information on predation) and therefore adequately fit the indices for these years.

However, the inclusion of variable growth and natural mortality was only partially successful in addressing this underestimation. There are periods in the time series when the model biomass estimates are below the survey observations, particularly in the Mid-Atlantic and Georges Bank Open zones. Residual patterns were therefore observed and are not ideal. The main reason for this is that observation error, natural mortality, and fishing mortality are confounded. Generally, the model allows the survey indices to have elevated levels of observation error (i.e., it underestimates these due to error in the survey index). Despite this potential for large observation error within the dredge surveys (see ToR 2), in some years the correlated deviations suggest some component of mortality is missing from the model for these years. It is unclear whether this is due to underestimation of natural mortality, fishing mortality, or both. *Ad hoc* adjustments to address this potential aliasing of other forms of fishing mortality into natural mortality has not fully addressed this issue.

Apart from these runs of residuals discussed, patterns were also observed in the Georges Bank assessment. These could be adequately explained by very noisy and (particularly) conflicting survey indices, and are therefore defendable.

Despite the above, generally, the assessment produced reasonable model retrospective patterns, although still an important aspect to consider for the next benchmark assessment. The worst of these, GBA, can be explained when the peels move over the noisy increases in survey indices in the last few years.

Although the results above regarding the confounding is an important issue to address further within the model, there is a greater need to address the impacts of density dependence within the surveys themselves. There may be some coherency with the assessment treating the dredge indices as including unaccounted for error. This is in particular due to having high levels of unexplained error resulting from size frequencies that do not always follow an explained sequence of cohort patterns post large survey indices. This, in turn, can be explained by large levels of natural mortality. Furthermore, the 2016 GBC dredge index declined while the optical indices increased – and this change occurred despite changing the dredge efficiencies in the last three years of the series. Addressing these issues with the surveys is a high priority and discussed in ToR 2.

The results of the CASA regional assessments support various environmental and ecological studies that indicate that natural mortality on juveniles increases at high densities. Work on this should continue.

Recommendation (Medium) 9. Continue studies on linking environmental and ecological research to density-dependent dynamics important to the assessment.

As a minor issue, the report often has very small plots that are difficult to read. Furthermore, size models usually have residual plots for the size frequencies (which were provided during the review).

Recommendation (Medium) 10. As a standard practice, provide residual plots in absolute and relative sense for both biomass/abundance fits and size frequencies. Each provide different and important information.

5.1.5.4 Growth

A new growth model that assumes individual variability in L_∞ and K are implemented in this assessment using shell increment data. These analyses show that growth is time varying. Several growth matrices are produced that reflect high to low growth rate and are applied to various time periods within the assessment. This approach to implementing growth within the assessment is supported.

Given the importance of applying time varying growth and for the size-frequency estimation component of the model, shell data need to be consistently collected and analysed. For example, no shell data were collected or analysed in MA since 2013. This means that the appropriate growth block and rate had to be inferred, which is not ideal. The growth analyses will have to be updated with the new data for each benchmark assessment.

Recommendation (High) 9. Given the importance of the inclusion of time-varying growth in these assessments, shell data should be regularly collected and analysed. The growth analyses would therefore need to be updated for each benchmark assessment.

The growth blocks were based on the incremental year effects of the analyses. The blocks were mainly based on visual inspection of the results. There is therefore a high chance that a different user may produce different blocks. Several approaches can be used to analytically develop or at least inform the creation of these blocks. These include change point (which can be Bayesian or frequentist in design) or cluster analysis. Some expert override may still be needed to consider the number of data in each block.

Recommendation (Medium) 11. Undertake change point and cluster analyses on the growth increment year effects to better describe appropriate growth blocks analytically.

5.1.5.5 Conversions

Landings are recorded as meat weight, one of the major currencies used within the model. The size-frequency of the landings is also included in the model and it therefore relies on a shell-height to meat weight conversion. Given that growth and the shell height – meat weight relationship has time and space varying components in the parameters, it is important that shell weight and height data are consistently monitored.

Recommendation (Medium) 12. Ensure that shell height and meat weight relationships are regularly updated over space and time.

5.1.5.6 Natural mortality

The CASA model calculates annual estimates of additional natural mortality that cannot directly be accounted for by fishery landings. Most of this mortality is due to time-varying natural causes (principally predation and disease), but there remains a small proportion that may be due to unaccounted for fishing-related mortality. For brevity, the additional mortality is included in the natural mortality calculations. Suggestions are made in ToR 1 regarding obtaining further discard information to partly resolve this issue.

With regard to underestimation of the survey indices, there is a penalty in how much the natural mortality (M) deviations can vary over time. There is obviously a trade-off between a fully flexible M and a highly restricted M deviation. One would expect this issue to be higher in regions where several high M deviations could be predicted over the time series as the survey indices are more variable (e.g., GBA). Sensitivity tests should be undertaken to investigate this trade-off.

Recommendation (Medium) 13. Undertake sensitivity tests on the trade-off between allowing greater flexibility in the natural mortality deviations and how this affects the residual patterns and survey noise.

5.1.5.7 Alternative approaches

This benchmark assessment only considered one alternative approach, a Beverton-Holt equilibrium length-based estimator. Further examples are also discussed in ToR 3, which could be applied for these regions as a check.

5.1.6 ToR 6: 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.

5.1.6.1 Overall response

This TOR is fully met.

5.1.6.2 Background

A second model, SYM, is applied to estimate regional biological reference points. A stock-recruitment relationship is estimated for each region using output from the respective CASA models. These are used to undertake per recruit calculations.

5.1.6.3 General comments

An important aspect of sequential, but linked models is to ensure that parameter values (such as selectivity, mortality), uncertainty and errors in variables are correctly addressed in the subsequent model. Based on the descriptions provided, these have been correctly implemented.

Interestingly, CASA is set up to undertake the per recruit analyses as well. It is unclear why SYM is used beyond that this model has been previously reviewed and used. Ideally, per recruit analyses should be undertaken within a single framework – mainly to avoid possible translation errors. Using both SYM and CASA as has been undertaken in this assessment as checks can also have benefits.

An important difference between the recruitment terms in CASA and that used in SYM is that these start at size 1, age 1 in CASA, whereas recruitment in SYM is from 3-year old animals. CASA aims to introduce juvenile mortality and uses all size classes to implement this. However, juvenile mortality is difficult to predict as CASA shows this to be more sporadic and density dependent, and so this is best not included in SYM. Therefore, using age-3 as an index of recruitment in SYM is supported.

A particular innovation of using gonad weight to calculate Spawning Stock Size (SSB) was proposed. The original meat weight unit was applied to stock biomass and yield per recruit analyses. In other words, the two would be used in conjunction. However, the analyses of gonad weight to shell height were not adequately demonstrated in terms of sample size, spatial coverage and comprehensiveness. Furthermore, this change variously emphasised different regions compared to the original meat weight biomass approach. This change was not properly discussed. Finally, gonad weight was not available for all regions, for example the SNL area.

Although the concept behind the introduction of gonad weight to describe spawning stock biomass has merit, the full implications of using this approach have not been fully investigated. The innovation in this approach is also highlighted. As a result, further work on consolidating these gonad-based spawning stock biomass metrics before full implementation through a) updating the shell height to gonad weight relationships (the SNL area), and b) evaluating the relative impact of these changes by region is recommended (see ToR 9).

Recommendation (High) 10. Undertake additional analyses and monitoring on developing the gonad based spawning biomass metrics.

During the review, both meat weight and gonad weight were therefore requested and provided. Both should be reported in the final benchmark assessment, but the stock biomass metrics based on meat weights are recommended for use as the criterion for determining stock status within this 2018 assessment. Further research on developing the gonad weight based spawning stock biomass index is recommended in ToR 9.

Despite this being a data rich assessment, large amounts of uncertainty are still propagated through the per recruit analyses. This uncertainty does not seem to be over-estimated nor incorrectly implemented. These analyses show that equilibrium analyses sit uncomfortably with this species, for example:

- recruitment is sporadic,
- there is a high degree of variability in recruitment and subsequent biomass,
- mortality and growth rates vary with time, and
- juvenile mortality may be density dependent.

Thus, the legal requirement to estimate MSY and associated reference points needs to be combined with the practical day-to-day management of the resource using an adaptive recruitment-based spatial approach, which appears to be a better framework for sustainable management.

5.1.7 ToR 7: 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.

5.1.7.1 Overall response

This ToR is fully met.

5.1.7.2 Background

The output from SYM (and by inference CASA), updated with the new data and information, was used to produce stock status for each of the regions and all regions combined. The SNL scallops are added to the biomass plots, but treated separately because they are not based on CASA or SYM models, but derived directly from the surveys.

5.1.7.3 General comments

Based on the updated SYM model output, there is support for the statement that the stock is neither overfished nor that overfishing is occurring, and that the probability that the stock is overfished or overfishing is occurring is very low. It is noted that the resource is healthy, but these are based on recent good year classes.

As stated in ToR 6, the gonad-based SSB and relative reference points that were developed were not recommended for this assessment (although should be reported) until further work is undertaken on the proposed research area highlighted in ToR 6.

5.1.8 ToR 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 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.

5.1.8.1 Overall response

These TORs were fully met.

5.1.8.2 Background

A third model, SAM, was used to simulate projections for spatial management. Since area management plays an important role in sea scallop dynamics (much of the biomass during some periods located in long-term and/or rotational closures), SAM is more spatially discrete than CASA and SYM. SAM uses the most recent survey index to forward project using outputs from CASA and SYM.

5.1.8.3 General comments

One could question the use of three different models. Much of this need arises due to the unique nature of the resource on the one hand (space-time varying dynamics and management) and the need to estimate equilibrium-based reference points based on robust stock and recruitment derived parameters. Each of the three models are used for separate well-defined purposes:

- the CASA model estimates historical biomass and fishing mortality rates at a regional scale;
- the SYM model estimates regional and whole stock biological reference points based on CASA outputs; and
- the SAMS model forecasts future abundance, biomass and landings at a finer spatial scale to address management needs.

The important issues are:

c. Whether each model is coherent with the other models and correctly implemented. Although some inconsistencies are highlighted in previous ToRs, the structure of each model is similar and is coherent where required. Thus, this approach is reasonable and supported.

d. Whether a fully spatial assessment model could be produced that would satisfy each of the above three purposes in a single model. At this stage, this is not available and would require a concerted research effort, which may not be entirely successful. However, given the nature of the data and the length of time this collection has been undertaken, a way forward of potentially bringing these different approaches into a single framework using a stepwise approach is discussed in ToR 9.

Points of divergence between SAMS, and CASA and SYM are that:

a) The projections are based on the most recent survey, which is not available at the time of the assessment (and therefore not included). SAMS is used directly to assist with the implementation of the temporal and spatial management of the fishery. Given the adaptive and spatial nature of management, it is supported that SAMS uses the most recent survey and not the recent estimates from CASA (which in fact would be one year behind).

b) The area-specific recruitment is scaled to the dredge surveys and the regional stock-recruitment relationship from SYM (based on CASA output). Care should be taken in these first two steps to ensure the same survey data are being used so that biomass is consistent both spatially and temporally.

c) Area specific growth parameters were developed, but these are included as a growth transition matrix in SAMS in a similar manner to CASA. Adult natural mortality is the same in CASA and SAMS, whereas juvenile natural mortality is based on output from CASA (where these are estimated) based on an externally (to CASA and SAMS) derived linear function of juvenile mortality and area specific density (a stronger function in the MA areas is applied compared to the GB areas). These plots (derived function and associated "data") should be provided in the report.

d) A function quantifying the relationship between LPUE and area biomass is required to calculate effort and fishing mortality. The source of mean exploitable biomass is not well explained in the methods, but should again be based on similar information as above. An alternative approach modelling LPUE against economic and catch variables was attempted, but not used. Although there is merit in continuing development of this fleet dynamic model, the use of the simpler regression approach as used is reasonable. The result of a test of SAMS's historical predictive ability on the recent past is striking support this finding.

Two projection systems are provided, fishing at Fmsy and according to Framework 29. The provision of both is appropriate in the context of the needs of policy versus management.

Sensitivity tests included the impact of bootstrapping (i.e., uncertainty), adult natural mortality as per previous assessments and lower L_∞. However, no sensitivity tests were undertaken on the juvenile mortality function. For MA particularly, the function was derived from the worst-case region; Elephant Trunk. It would be important to highlight the role of this function, especially given the present high biomass.

Recommendation (Medium) 14. Include sensitivity tests on the impact of density dependent juvenile mortality.

Recommendation (High) 11. Develop a more comprehensive understanding and updated function of juvenile mortality for SAMS.

As one would expect from CASA, the healthy stock biomass levels in recent years have been due to two very large cohorts (the 2012-year class on Georges Bank, primarily located in the Nantucket Lightship Area, and the 2013-year class in the Mid-Atlantic, much of which is in the Elephant Trunk rotational area off Delaware Bay). When these dynamics are projected by SAMS, forecasts are uncertain, since these very high densities of scallops have rarely been observed. Despite this uncertainty, the projected total biomass and landings is correctly predicted to decline as these strong year classes are fished or die naturally.

An important feature to note about spatial management compared to regional assessments is that the fishing mortality from CASA assumes the mortality is spread over the whole assessed region, whereas in reality this mortality is concentrated in open access areas. As a result, the reported fishing mortality from the assessments would underestimate the fishing mortality where fishing occurs (and of course imply fishing mortality in areas that are closed). Thus, it is possible that areas open to fishing can be locally depleted despite overfishing not occurring on the whole stock.

5.1.9 ToR 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.

5.1.9.1 Overall response

This ToR was fully met.

5.1.9.2 Background

Responses to previous research priorities were provided in the report. Also, included were new research. During the review workshop, the panel asked the assessors to prioritise this list. These were provided and formed the basis of this response.

5.1.9.3 General comments

The process of commenting against previous research recommendations should be updated. It would be much more useful if advice was provided on a) whether a recommendation can be removed (i.e., it is either completed or not relevant) and b) which are highest priorities of the full research list. In the future, this should be part of the SAW process.

Four high research priorities were provided:

1. “Investigate changes in dredge efficiency and saturation due to high scallop densities or high bycatch rates.
2. Develop a spatially-explicit methodology for forecasting the abundance and distribution of sea scallops by incorporating spatial data from surveys, landings, and fleet effort (a.k.a. GEOSAMS).
3. Analyze past juvenile scallop mortality events and develop better methods to model time-varying mortality in the assessment models.
4. Investigate methods to better estimating biomass and abundance variances from Habcam optical surveys and other resource surveys including development of Bayesian geostatistical methods, model-based dredge estimates or re-stratification to address dredge survey variability.”

In the above context, a comprehensive package of linked projects should be developed. These include the work described above (and suggested), but should also apply to the size frequency analyses. In addition to restratifying the size frequency information, consider reweighting the survey size frequencies by survey catch (a common approach is sqrt weighting, which downweights large catches relative to smaller catches). This will help balance the information content of the different sites, data sources and account for spatial heterogeneity inherent in the survey observations.

Given the survey data are now an extensive dataset, both temporally and spatially, ultimately moving towards providing more cohesion between CASA, SYM and SAMS should be possible. As a first step, undertaking extensive space-time modelling of surveys and commercial VTR data would be highly beneficial and is a high priority. Methods could be drawn from generalised additive modelling, geostatistical and lattice approaches. Nested within this work is the need, as stated above, to understand dredge and optical survey efficiency and saturation effects in the length frequencies, and adult and juvenile densities so that all the surveys can be combined in this approach. Initially, this work would result in a combined index of abundance for use in CASA and the same, but spatially explicit, in SAMS. Similarly, the size-frequencies would be more consistent and cohesive. Dynamic space-time auto-regressive models could be used to forward project.

As an interim as well, given that CASA can forward project as well as SYM, this model could be removed. However, having two separate models for estimation and projection is not uncommon.

While this space-time modelling is underway, restratification of the surveys are essential, especially the dredge survey. Even if the assessment is only partially correct, the high levels of unaccounted for dredge observation error is not ideal for such a valuable fishery and shows that much work is needed on this survey's design. Potential redesign or addition of geostatistically derived sites should be considered, although without breaking the historical time series.

Ultimately, based on the strength of the space-time models, removal of CASA or using it as an alternative assessment may be possible.

Given that the gonad-based index of SSB has merit as a biological reference point, undertaking information gathering on shell height to gonad weight relationships in all areas is essential. The regional impact of using a gonad-based SSB should also be comprehensively analysed.

In the Gulf of Maine several alternative approaches are suggested (ToR 3). The key to this region is to collect information required to ultimately undertake a SAMS model. Investigating what data are needed would help prioritise the data needs. A single survey is required and suggestions are made in ToR 3 on how to approach this going forward, which includes undertaking a cost-benefit analysis of survey scope, regularity and scale as a priority.

The full set of recommendations provided within each ToR is provided in Section 6.1.

5.2 Atlantic Herring

5.2.1 ToR 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.

5.2.1.1 Overall response

This ToR was fully met.

5.2.1.2 Background

Catches from various sources including Canada have been described since 1965 and discards since 1996. Evidence, based on weight-length relationships, as to which gears could be combined was also provided. Biological information, such as age and length composition by gear type, were described.

5.2.1.3 General comments

Spatio-temporal information relating to the fishery was provided, plus an understanding of the issues with these data. These were useful in understanding the behaviours of the fishery and the impact of environmental drivers. Discard information demonstrated that the lack of information prior to 1996 would not be an important source of uncertainty in the assessment given the low percentage of discarding in the fishery (where data existed). Catch biological information such as age and length structure, maturity, by gear type was well described. Spatial maps of catches were useful to understand how the fishery moves over time, and the influence of herring migration patterns. Some overlay of which region hit a cap would be informative.

There was some discussion on the influence of illegal and unreported catches when a large part of the catch was by foreign fleets. The importance of bycatch caps on fleet behaviour was also described.

There were periods where no sampling of US fixed gear age composition took place. These data had to be derived from Canadian weir data. Given that the age composition of this gear type is of a younger age group, sample sizes may not need to be high and it is recommended that, where possible, these data are collected directly from the US fixed gear fleet.

Recommendation (High) 1. Maintain a consistent sampling program for the US fixed gear, especially for age and length data.

5.2.2 ToR 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.

5.2.2.1 Overall response

This ToR was fully met.

5.2.2.2 Background

A number of different surveys and their associated uncertainties are described. Changes in gear type and vessels were described and actions taken (if any).

5.2.2.3 General comments

The appropriateness of using bottom trawl surveys to monitor a semi-pelagic fishery was fully explored and was shown to be an appropriate means of providing an index of abundance for herring. Uncertainty and biases in the indices were well described for all the surveys. These included providing confidence intervals on the plots and discussing issues in the text.

Age correlation matrix plots were provided during the review. These showed that correlations between a given age and the next age class were reasonable, but very weak at predicting subsequent ages. The fall 2009-2016 survey series was particularly weak in its predictive capability.

The approach of dividing the NMFS spring and fall bottom trawl survey into three distinct time periods is supported – the first to address a change to the trawl doors and the second a vessel change. Also, how selectivity is set up (a fixed input for the early series, but estimated thereafter) is a practical solution given the absence of age data prior to 1987. This is an improvement on previous assessments that used a calibration factor for vessel change.

The acoustic survey data were used for the first time in the assessment. This is a common survey approach used for other small pelagic species globally (more so than bottom trawls). The data here are based on acoustic data collected while the survey vessel moves between sites for the bottom trawl survey and is therefore an opportunistically developed index. The appropriateness of using a non-standard design index (from a statistical view point) should be investigated. Similar to that used in the past, a dedicated acoustic survey (designed to create an index of abundance) should be investigated.

Recommendation (High) 2. Investigate the impact on the index of abundance as used in the assessment of using a non-standard design acoustic index (from a statistical view point). Undertake a dedicated acoustic survey designed specifically to create an index of abundance.

Several other sources of indices were examined and prioritised for use. One potential index not considered was whether other sources of acoustic data (e.g. from the lobster industry) could be useful to the herring assessment.

Recommendation (Medium) 1. ***Consider other forms of acoustic data, for example from lobster vessels.***

A food habits index was developed using Generalised Additive Mixed Models (GAMM) for use in the assessment. However, the index was sensitive to which predators were included in the GAMM (e.g., spiny dogfish) and showed some retrospective pattern with updating the data to 2016. This index was not therefore directly used in the assessment, but rather as allied information to support the level of natural mortality used in the assessment. This approach is supported, but further work on these indices should be undertaken.

Recommendation (Medium) 2. ***Continue development of the predation index.***

5.2.3 ToR 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.

5.2.3.1 Overall response

This ToR was fully met.

5.2.3.2 Background

A time series of herring consumption was estimated from prey consumption data by herring predators. These data were used to develop an annual natural mortality proxy using estimates of herring consumption. When combined with predator abundance from survey or assessments, total herring consumption was calculated.

5.2.3.3 General comments

Consumption of herring by the main fish predators was documented, and uncertainty and biases in these were well described. The main potential bias discussed was the presence of mostly larger herring in the diets. As part of the discussion during the review, more information was provided about the scale of predation by marine mammals, birds and larger finfish such as sailfish and tuna. It was indicated that these forms of predation were likely minor relative to the main fish predators. Not all this information provided during the review was clearly articulated in the report and should be added in the future.

This work does not consider prey switching through, for example, the confounding of space-time impacts of other prey species also available in the area for consumption by predators.

Recommendation (Medium) 3. ***Investigate the impacts of prey switching on herring consumption indices.***

The consumption of herring by tuna was further discussed, especially given the assessed stock status of herring. It was noted that the population of tuna was small and that tuna consumption would be small compared to consumption by other predators. It was also noted that the fishery was responsible for a minority of removals compared to the natural predators and would likely have a low impact on food availability to tuna.

The herring consumption calculations were not considered for use in the assessment as an additional "fleet". Although the uncertainty in this index is such that it would have been premature for this assessment, it may be worth considering as an option for future assessments.

Recommendation (Medium) 4. ***Consider whether predator total consumption values could be added as a separate "fleet" in the assessment. Length/age data could then also be included.***

Herring changes in distribution were discussed in the report and during the review. Some more information was provided during the review (e.g., centre of gravity plots) that could be included in future reports. However, the information that was provided (e.g., kernel densities) was very informative and did support the conclusions drawn by the assessors that herring climate vulnerability is relatively low, along shelf distance has remained unchanged but there are some trends towards herring moving deeper. Despite these results, there was agreement during the discussions that further environmental and oceanographic work and how these change over time is justified.

Recommendation (Medium) 5. ***Continue investigating the influences on herring abundance, dynamics and distribution of oceanographic and environmental factors.***

5.2.4 ToR 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.

5.2.4.1 Overall response

This ToR was fully met.

5.2.4.2 Background

An update of ASAP of Legault and Restrepo (1998) was run. This was the same model as used in previous assessments. An ASAP base run was provided, plus sensitivity tests and several types of retrospective analyses. Additional model types were also explored.

5.2.4.3 General comments

The analysis was thorough and well set out - several models were explored, and different types of retrospective and sensitivity analyses were conducted anticipating many of the questions likely to be raised during the review process. Diagnostic tests and profile likelihoods were also provided. The impact of the removal of key data sets was also explored. These thorough tests explored the strengths and weaknesses of ASAP's application to herring. Although there are some concerns, mainly for future assessments mentioned below, the assessment is accepted.

5.2.4.4 Natural mortality and retrospective patterns

The previous assessment in 2012 had large retrospective patterns when data removal peels were undertaken. Based on these retrospective patterns and a comparison of consumption data, the index of natural mortality was increased by 50% which reduced these patterns. However, the re-emergence of retrospective patterns in the subsequent assessment no longer supported time varying natural mortality. Mohn's Rho adjustments were applied in the previous assessments to account for these patterns, which were appropriate when the existing model was updated through to 2017.

Consideration of the presence and absence of retrospective patterns is an important aspect of this assessment. The assessors treated this benchmark assessment to revisit most of the key model structures and decisions. This assessment therefore has major changes that include a) addition of the acoustic series, b) running the age composition to 8+ rather than 9+ given reliability issues of older ages in previous assessments, c) implementing the initial abundances at age differently so that a likelihood penalty could be removed as well as both age- and time- invariant natural mortality, d) treating selectivity differently, e) removing age 1 catches from some surveys, and f) unconstrained annual recruitment deviations. Many of the retrospective patterns seem to be removed by the changes, although the stability of this benchmark base model still needs to be proven. These changes are supported.

Retrospective patterns in the new benchmark assessment were reasonable, thus obviating the need for time varying natural mortality. Furthermore, unlike previous assessments, consumption calculations were not used to change the scale of natural mortality as the results did not support changing natural mortality based on consumption information. No Mohn's Rho adjustment was undertaken of the results, because the values fell within the 80% confidence intervals. These decisions are supported.

However, based on the report alone, these processes are not as clearly articulated as during the review. The between assessment changes mean that the rigour of the required decision processes should be clearly articulated as a series of protocols. Furthermore, there may need to be some between species standardisation of the value of the Confidence Interval criterion for the Mohn's Rho adjustment. The justification and process undertaken setting up the assessment structure and process is supported.

Recommendation (Medium) 6. ***Set up protocols describing the assessment decision process (when natural mortality adjustments are supported and when a Mohn's Rho adjustment for retrospective patterns should be applied).***

Consumption calculations were used to scale natural mortality rather than including this information directly into the model. This avoids introducing excess random variation into the model. However, a state-space approach could be applied as a sensitivity test that includes this information directly into the model while allowing for a smoothing of the process. Importantly, this test should consider that the predatory index is biased by predator species composition meaning that these analyses and their interpretation should be undertaken with care.

Recommendation (Medium) 7. ***Investigate a state-space model that includes consumption data directly in the model for use as a sensitivity test.***

During the review an additional sensitivity test was requested to see the response of the population time series to the value of natural mortality. This was noted because the value of M used in the assessment was not the minimum negative log likelihood estimate based on the profile likelihood. A reasonable justification for the M implemented was provided, but sensitivity to this parameter is often requested during reviews and should be undertaken if the value of M is not the best value that is implied from a likelihood profile.

Although generally there were not large residuals or trends, that for the catch stands out and should be resolved for the next assessment. It is noted that the residuals are small.

Recommendation (Medium) 8. ***Investigate the reasons behind the sequences of positive and negative residuals in the catches.***

Of more importance is that the catchability estimated between the two recent periods within the model is quite different to the field derived calibration values previously used. This could be a sign that another factor is being aliased in this catchability term. At this stage, it is unclear given the tests provided, which model and data assumptions are causing this issue. This issue needs to be addressed for the next benchmark assessment to avoid potential residual patterns.

Recommendation (High) 3. Investigate the reasons for the difference between model estimated catchability and the field derived calibration values for the Albatross to Bigelow vessel transition.

Recent recruitment is poorly estimated with high coefficients of variation (CVs). This result influences how future recruitment is drawn.

5.2.4.5 Alternative assessments

A two-stock Stock Synthesis model was developed to address potential stock structure issues. This model is discussed in ToR 8. Another model, using Stock Assessment Model (SAM) was also developed, but was in its initial stages of development. SAM was difficult to diagnose and use as a comparison to ASAP and to derive relevant reference points. This work has merit as alternative assessment tools, especially given strong retrospective patterns in ASAP have (or do) existed.

Recommendation (High) 4. Continue development of SAM and SS assessments.

5.2.5 ToR 5: 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.

5.2.5.1 Overall response

This TOR was fully met.

5.2.5.2 Background

The existing MSY reference points were based on a model fit to a Beverton-Holt stock-recruitment relationship, which was internally estimated within the ASAP assessment.

5.2.5.3 General comments

For the updated benchmark assessment, estimating steepness was not justified. A $F_{40\%}$ proxy was therefore applied. New reference points were proposed. These new reference points are well justified given changes in the selectivity in the commercial fishery over time. The fishery is now targeting older, larger fish more strongly. It is noted that MSY and SSBmsy would change because of this update. This response is scientifically sound.

5.2.6 ToR 6: 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.

5.2.6.1 Overall response

This TOR was fully met.

5.2.6.2 Background

The use of a stock-recruitment relationship estimated within the assessment was found to be inadequate in the new assessment. Targeting practices have also meant that fishery selectivity had changed, also affecting the new reference point selection. A new approach is suggested, but means that the newly proposed reference points cannot be compared with past assessments.

5.2.6.3 General comments

Recruitment in each year was appropriately drawn for the reference points. The recruitment estimates from 2016-2017 were excluded from the distribution used to draw future recruitment values, i.e., only years 1965-2015 were used. The 2016-17 values were too uncertain. This approach is supported.

New reference points were provided in a Kobe plot. The panel requested that the time series phase plane plot be provided for reference and historical context. These were also provided and is a usual addition to future reports.

The finding that the stock is currently not overfished, and that overfishing is not taking place with at least 50% probability is supported. However, this assumes mean recruitment would be re-established.

It is noted that Management Strategy Evaluations are being undertaken to investigate alternative management procedures. This work is supported and is important given the current stock status will be influenced in the future with the low recruitments and poor year classes moving through the system over time.

Recommendation (High) 5. Continue Management Strategy Evaluation work to develop management procedures for this species.

5.2.7 ToR 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 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.

5.2.7.1 Overall response

This ToR was fully met.

5.2.7.2 Background

Short-term projections were undertaken based on the results from the ASAP model. Due to the uncertainty in the 2016 and 2017 values, forward projections were handled differently for 2018 and 2019-2021.

5.2.7.3 General comments

Although unusual, the forward projection methodology is supported. The SAW did appropriately deal with the high uncertainty associated with recent recruitment estimates when developing projections. Mohn's Rho retrospective adjusted values were (correctly) not applied given the retrospective adjusted values fell within the 80% confidence interval of the base model estimates. This follows previous protocols.

Two values of the 2018 catch were also tested – a) that equal to the predicted ABC, and b) half this value. This is due to present fishery activities tracking below those needed to achieve the ABC in 2018. Lower harvest scenarios appear to result in less pessimistic projections.

An additional sensitivity run was requested during the review where average recruitment was assumed to be ½ the mean assumed in the base projections. This would partially simulate current low recruitment levels (especially if the 2016 and 2017 low values are confirmed in subsequent assessments). These sensitivity tests were used to highlight how a run of poor recruitment would affect the short-term future. Lower recruitment scenarios seemed to better represent current stock conditions.

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

5.2.8.1 Overall response

This ToR was fully met.

5.2.8.2 Background

Based on the information provided, there is likely to be unaccounted for stock structure in the assessment. A different assessment tool, Stock Synthesis was run in an attempt to account for stock structure on a coarse level. This assessment was unable to be completed since key assumptions such as migration rates were unknown.

5.2.8.3 General comments

An additional model was developed to address potential unaccounted for stock structure. Data were presented and a Stock Synthesis III spatial model was explored, but the analysis was inconclusive given the stock information available. The information needed to advise on changes to stock structure and consequently to motivate management actions is unavailable at this time. However, aspects of stock structure are worthy of further exploration for future assessments, including genetic separation, rates of movement and distinguishing stock specific harvesting from mixed catch fisheries.

There was discussion about whether tagging studies would be effective in being able to distinguish migration rates or general movement patterns. Advice was conflicting about return rates, but seems warranted given past work where tagging studies had been undertaken. This information should be reviewed based on present modelling needs.

Recommendation (High) 6. Undertake a review of past tagging studies for their utility in present assessment models.

Recommendation (High) 7. There are likely still unanswered questions about mixing which otolith microchemistry and morphometric studies could address. These studies should be pursued.

The Stock Synthesis model was developed to address potential stock structure issues. However, it was not run simply as an alternative single stock assessment to ASAP. One of the key features of this model is that it is able to fit to length and age data. Running a Sock Synthesis model as an alternate assessment to ASAP is warranted.

Recommendation (High) 8. Apply a Stock Synthesis model as an alternate single stock model to ASAP so that size and age data could be included in the model.

5.2.9 ToR 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.

5.2.9.1 Overall response

This ToR was fully met.

5.2.9.2 Background

In the report, progress against the research recommendations from various sources was described. New research recommendations from the SAW were also added.

5.2.9.3 General comments

During the review, it was discovered that much more progress was made against the recommendations than was described in the report. Some of the recommendations were also not always technically sound. As a result, it is recommended that redundant and completed recommendations should be removed from the list (with justifications). This should be an on-going process within the SAW.

Recommendation (Medium) 9. ***Continuously remove redundant or completed recommendations (with reasons) as part of the SAW process.***

The list of recommendations was also not prioritised in the report. During the review, the assessors were asked to undertake this task and these priorities were provided to the panel. This was very helpful.

Recommendation (Medium) 10. ***As part of the SAW process, prioritise research recommendations.***

Priorities provided during the review (in italics) were:

1. *Further research on the use of acoustic technology for inclusion in stock assessment.* This priority is supported and I agree it is high; however, it is further recommended that the initial approach with the acoustic data should not be to move away from a relative index to an absolute index. The level of knowledge required on target strength, and species orientation and identification are as yet not as complete as would be required to assume absolute abundance in the assessment.

2. Evaluate data collected in study fleet program. Evaluating whether the flounder study fleet program data could be useful for the assessment is merited.

3. *Evaluate the ability of state-space models to reliably estimate observation and process error variances to estimate quantities of management interest*. This is an important priority. Given the history of this assessment, developing a structurally different assessment is essential and a high priority. Rather than dismissing the Stock Synthesis III model development, because it was not able to run a multi-stock version, it is further recommended that a single-stock Stock Synthesis model be developed given that it can also fit to length information.

Undertaking research on stock structure such as movement rates, harvesting rates from each stock and genetic separation is important. Investigation of the earlier tagging study would be an essential place to start. Furthermore, otolith microchemistry and geomorphology have been shown to be useful tools to partly address stock structure issues.

6 CONCLUSIONS AND RECOMMENDATIONS IN ACCORDANCE WITH THE TORS.

6.1 SEA SCALLOPS

The high priorities provided during the review are supported. Additional priorities that were discussed against each ToR are provided below.

6.1.1 High priority recommendations

Recommendation (High) 1. Undertake an extensive historical analysis of the dredge survey index and size-frequency data investigating the potential use of a) restratification and b) model-based approaches...........................23

Recommendation (High) 2. Use the results from the historical analyses as a basis for redesign, if required, of the dredge survey. 23

Recommendation (High) 3. Undertake further field work on dredge efficiency and its impact on the index of abundance and size frequency data. 23

Recommendation (High) 4. Consider double counting of a percentage of the Habcam video photographs for abundance and size-frequency. 23

Recommendation (High) 5. Improve training of annotators of shell count, size and dead scallops especially at high and low densities. 23

Recommendation (High) 6. Focus on initially using SAMS to prioritise data and information needs. ... 24

Recommendation (High) 7. Undertake a cost-benefit analysis of the survey requirements for the GoM with a focus on creating a tool to set TACs. This should include considering the ideal option which would be regular GoM wide surveys. 25

Recommendation (High) 8. Investigate the use of data moderate models. Amongst these, concentrate on Bayesian approaches where hyper-priors can be informed by the data rich regions. 25

Recommendation (High) 9. Given the importance of the inclusion of time-varying growth in these assessments, shell data should be regularly collected and analysed. The growth analyses would therefore need to be updated for each benchmark assessment. 30

Recommendation (High) 10. Undertake additional analyses and monitoring on developing the gonad based spawning biomass metrics. 32

Recommendation (High) 11. Develop a more comprehensive understanding and updated function of juvenile mortality for SAMS. 36

6.1.2 Medium priority recommendations

Recommendation (Medium) 1. Investigate the impact of the Canadian sector on Georges Bank as a sensitivity test in CASA and SYM. 20

Recommendation (Medium) 2. Include discard mortality as a separate term within the CASA model with a lower selectivity as shown by the relative size frequency distributions. Set these up similarly to those in SYM and SAMS. 20

Recommendation (Medium) 3. Investigate alternative data sources such as apprehensions to define the scale of poaching and highgrading and test how this affects estimates of natural mortality regionally. 21

Recommendation (Medium) 4. Investigate all major sources of fishing mortality that are aliased as not natural mortality and either a) address these through model structure changes or b) provide evidence that the present ad hoc adjustments are sufficient. 21

Recommendation (Medium) 5. Undertake sensitivity tests within CASA of different dredge efficiencies until more studies can be used to derive this value with greater certainty. 22

Recommendation (Medium) 6. Develop Habcam analysis protocols and QA/QC protocols that are aimed at reducing the risk of density dependent biases in the estimates of density and size-frequency. 23

Recommendation (Medium) 7. Move towards model-based estimates of biomass indices and maps as an option. This may be best to be either undertaken at the same time as the other regions or thereafter and so will need to be determined. 26

Recommendation (Medium) 8. Investigate as part of model-based space-time analyses of survey biomass indices and maps whether the inclusion of environmental correlates add value to the predictions. 27

Recommendation (Medium) 9. Continue studies on linking environmental and ecological research to density dependent dynamics important to the assessment. 29

Recommendation (Medium) 10. As a standard practice, provide residual plots in absolute and relative sense for both biomass/abundance fits and size frequencies. Each provide different and important information. 29

Recommendation (Medium) 11. Undertake change point and cluster analyses on the growth increment year effects to better describe appropriate growth blocks analytically................................. 30

Recommendation (Medium) 12. Ensure that shell height and meat weight relationships are regularly updated over space and time. 30

Recommendation (Medium) 13. Undertake sensitivity tests on the trade-off between allowing greater flexibility in the natural mortality deviations and how this affects the residual patterns and survey noise. 31

Recommendation (Medium) 14. Include sensitivity tests on the impact of density dependent juvenile mortality. 36

6.2 ATLANTIC HERRING

The high priorities provided during the review are supported. Additional priorities that were discussed against each ToR are provided below.

6.2.1 High priority recommendations

Recommendation (High) 1. Maintain a consistent sampling program for the US fixed gear, especially for age and length data. 40

Recommendation (High) 2. Investigate the impact on the index of abundance as used in the assessment of using a non-standard design acoustic index (from a statistical view point). Undertake a dedicated acoustic survey designed specifically to create an index of abundance................................. 41

Recommendation (High) 3. Investigate the reasons for the difference between model estimated catchability and the field derived calibration values for the Albatross to Bigelow vessel transition. 46

Recommendation (High) 4. Continue development of SAM and SS assessments................................ 46

Recommendation (High) 5. Continue Management Strategy Evaluation work to develop management procedures for this species. 48

Recommendation (High) 6. Undertake a review of past tagging studies for their utility in present assessment models. 50

Recommendation (High) 7. There are likely still unanswered questions about mixing which otolith microchemistry and morphometric studies could address. These studies should be pursued. 50

Recommendation (High) 8. Apply a Stock Synthesis model as an alternate single stock model to ASAP so that size and age data could be included in the model. 50

6.2.2 Medium priority recommendations

Recommendation (Medium) 1. Consider other forms of acoustic data, for example from lobster vessels. 41

Recommendation (Medium) 2. Continue development of the predation index. 42

Recommendation (Medium) 3. Investigate the impacts of prey switching on herring consumption indices. 43

Recommendation (Medium) 4. Consider whether predator total consumption values could be added as a separate “fleet” in the assessment. Length/age data could then also be included. 43

Recommendation (Medium) 5. Continue investigating the influences on herring abundance, dynamics and distribution of oceanographic and environmental factors. 43

Recommendation (Medium) 6. Set up protocols describing the assessment decision process (when natural mortality adjustments are supported and when a Mohn’s Rho adjustment for retrospective patterns should be applied). 45

Recommendation (Medium) 7. Investigate a state-space model that includes consumption data directly in the model for use as a sensitivity test. 45

Recommendation (Medium) 8. Investigate the reasons behind the sequences of positive and negative residuals in the catches. 46

Recommendation (Medium) 9. Continuously remove redundant or completed recommendations (with reasons) as part of the SAW process. 51

Recommendation (Medium) 10. As part of the SAW process, prioritise research recommendations... 51

7 REFERENCES

Carruthers, T.R., Punt, A.E., Walters, C.J., MacCall, A., McAllister,M.K., Dick, E.J.,k, Cope, J. 2014. Evaluating methods for setting catch limits in data-limited fisheries,Fisheries Research,153: 48-68. https://doi.org/10.1016/j.fishres.2013.12.014.

Punt, A. E., Smith, D. C., and Smith, A. D. M. 2011. Among-stock comparisons for improving stock assessments of data-poor stocks: the “Robin Hood” approach. – ICES Journal of Marine Science, 68: 972–981.

Zhou S, Punt AE, Deng R, Dichmont CM, Ye Y, Bishop J. 2009. Modified hierarchical Bayesian biomass dynamics model for assessment of short-lived invertebrates: a comparison for tropical tiger prawns. Mar. Freshw. Res. 60(12): 1298-1308.

8 REVIEW PROCESS

The review was undertaken in a very constructive light with extensive requests met by both assessment teams. The assessors should be complimented on their open and positive approach to any comments, and their informative resolution of key questions. The background material was very comprehensive and addressed each Term of Reference. Tests were comprehensive.

The agenda did not allow for enough time to discuss each assessment during the presentation time. This was particularly notable for the sea scallop review which needed to cover three separate assessment tools. It would have been preferable that another day was added to the review or less time was allocated to other components of the agenda.

9 APPENDIX 1: BIBLIOGRAPHY OF MATERIALS PROVIDED FOR REVIEW

9.1 SCALLOP REFERENCES

Chang, J-H, Shank, B. V. & D. R. Hart. 2017. A comparison of methods to estimate abundance and biomass from belt transect surveys. Limnology and Oceanography: Methods 15, 480–494.

Hart, D. R. & P. 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.

Hart, D. R. & A. S. Chute. 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.

Hart, D. R. 2013. Quantifying the tradeoff between precaution and yield in fishery reference points. ICES Journal of Marine Science 70:591-603.

Hart, D. R., Jacobson, L. D. & J. 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.

Hennen, D. R. & D. 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.

Northeast Fisheries Science Center. 2014. 59th Northeast Regional Stock Assessment Workshop (59th SAW) Assessment Report. US Department of Commerce, Northeast Fisheries Science Center Reference Doc. 14-09; 782 p. Available at http://nefsc.noaa.gov/publications/

Stokesbury, K. D. E., B. P. Harris, M. C. Marino & J. Nogiera. 2004. Estimation of sea scallop abundance using a video survey in off-shore U.S. waters. Journal of Shellfish Research 23(1):33-40.

Scallop Survey Methodology Review Team. 2015. Summary Report of the Review of Sea Scallop Survey Methodologies and Their Integration for Stock Assessment and Fishery Management.

Truesdell, S. B., Hart, D. R. & Y. 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.

Truesdell, S. B., D. R. Hart, & Y. Chen. 2017. Effects of unequal capture probability on stock assessment abundance and mortality estimates: an example using the US Atlantic sea scallop fishery. Canadian Journal of Fisheries and Aquatic Sciences 74: 1904–1917.

9.2 HERRING REFERENCES

Deroba J. J. 2015. Atlantic herring operational assessment report 2015. NEFSC Ref Doc 15-16; 30 p.

Deroba, J. J. 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.

Northeast Fisheries Science Center. 2012. 54th Northeast Regional Stock Assessment Workshop (54th SAW) Assessment Report. US Department of Commerce, Northeast Fisheries Science Center Reference Doc. 12-18; 600 p. Available at http://www.nefsc.noaa.gov/publications/

9.3 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.

10 APPENDIX 2: A COPY OF THE CIE 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.

<u>Tasks for SARC panel</u>

- 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, <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 SOW and Schedule of Milestones and Deliverables below.

<u>Tasks for SARC chair and reviewers combined</u>:

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, passport number, country of passport, travel dates, country of citizenship, country of current residence, and home country) to the NMFS Project Contact for the purpose of their security clearance, and this information shall be submitted at least 40 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***

The 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

<u>Tuesday, June 26</u>

10 – 10:45 AM	Welcome/Description of Review Process Introductions/Agenda Conduct of Meeting	James Weinberg, SAW Chair TBD, SARC Chair	
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		

TOPIC	PRESENTER(S)	SARC LEADER	RAPPORTEUR

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)		

TOPIC PRESENTER(S) SARC LEADER RAPPORTEUR

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.

11 APPENDIX 3: PANEL MEMBERSHIP OR OTHER PERTINENT INFORMATION FROM THE PANEL REVIEW MEETING.

Panel members

Dr Patrick Sullivan (chair)

Dr Cathy Dichmont

Dr Geoff Tingley

Dr Coby Needle