Report on the

56th Stock Assessment Workshop/Stock Assessment Review Committee (SAW/SARC): Benchmark stock assessments for Atlantic surfclam and White hake

Prepared for:
The Center for Independent Experts

Dr Kevin Stokes
Stokes.net.nz Ltd
59 Jubilee Rd
Khandallah
Wellington 6035
New Zealand
Ph: +64 (04) 973 7305
E-mail: kevin@stokes.net.nz

EXECUTIVE SUMMARY

The CIE independent report shall be prefaced with an Executive Summary providing a concise summary of the findings and recommendations, and specify whether the science reviewed is the best scientific information available.

The 56th Stock Assessment Workshop/Stock Assessment Review Committee (SAW/SARC) meeting took place at the NEFSC, Woods Hole, MA, from 19th to 22nd February 2013. The review was hosted by NEFSC. Two stocks were considered at SARC 56: surfclam (*Spisula solidissima*) and white hake (*Urophycis tenuis*).

In my view the meeting was too short and reports and presentations inadequate to allow a thorough review of either assessment, especially as both assessments used new models. For white hake in particular the assessment report lacked sufficient detail and problems with specific terms of reference for each stock (on BRPs for white hake and stock definition for surfclam) required extra time. The SARC requirements for reviewers to write two SARC reports, contribute to two Assessment Summary reports, and write individual CIE reports, created considerable difficulties, especially as materials relevant to white hake BRPs required revision and eventual circulation after the review. I have made various recommendations in this report; unusually, some of these refer to SARC processes which I think could be improved.

Surfclams are an economically valuable fishery and an impressive amount of research, including extensive cooperative survey work, has enabled significant improvements in key parameter estimation. Those parameters were used as inputs to a new assessment model implemented in SS3 which used age composition data for the first time. The motivation for the new model approach was to use age information to improve recruitment estimation and hence projections, as well as providing a more flexible modeling framework for the future. In my view the new model is not yet fully explored and there are some difficulties that need to be resolved. Nevertheless, because surfclam are only very lightly exploited within a small area relative to their total distribution I accept that the ratio BRPs derived are reasonably robust and can inform management. I am not convinced by some of the BRPs which I think need to be better justified and more fully explored. More generally, I am unconvinced that the modeling approach taken is necessary. It would be possible to provide similar robust advice using very simple methods (or the older biomass dynamics model) and to undertake smarter management-focused modeling using management strategy evaluation. A major issue in the surfclam assessment this year was the consideration of stock definitions. The SAW was not able to resolve differences of opinion about whether the stock should be defined as one or two stocks (or possibly more) and inappropriately referred the decision to the SARC Panel. The SAW undertook separate stock assessments for two areas (Georges Bank as a single area and all other regions as a separate area) and combined results to enable advice at either the single or separate stock levels. Regardless of definition, surfclam appears not to be overfished or experiencing overfishing at this time and is unlikely to become so in the short-term under realistic catch scenarios despite apparent declines in biomass due probably to environmental conditions.

When last assessed at GARM III, the stock of white hake was assessed to be overfished and subject to overfishing. At SARC 56 a new assessment (ASAP) model has been use for the first time. In addition, catch streams were revised, pooled (as opposed to annual) ALKs were used and selectivity estimation was different. The new model was not explored in depth to test for the impact of these changes but they were considered by continued work on the GARM III (ASPM) model. That work indicated that changes in estimated SSB and $F_{MSYproxy}$ are likely due to changed selectivity estimates as well as the revised catch streams and use of pooled ALKs, not to the change from ASPM to ASAP. I accept the ASAP base model as a basis for providing advice on status and future catches. This is despite my criticism of the SAW report and the brevity of discussion therein. I do not find the modelling convincing as presented and think more work could usefully be done to improve understanding of uncertainty and to improve credibility. However, the continued use of the previous model (ASPM), and investigations with that model, has helped to provide confidence that the ASAP model can be used at this stage. Despite the need for more model exploration, the conclusions that the white hake stock is not overfished in 2011, nor experiencing overfishing, appear robust. BRP development for white hake has been problematic. The SAW recommended a change in $F_{MSYproxy}$ from F40% to F35%. From first principles I have no difficulty with use of F35% for white hake and with suitable analysis would not be surprised to see a lower FX% evaluated as acceptable. I do have difficulty, however, with the basis put forward by the SAW to change from F40% to F35% at this time and consider the SAW remiss in not explaining its interpretation of risk and not reporting key numbers.

The BRP debate at SAW/SARC 56 highlights some important issues. It is not clear why there are no clear defaults for BRPs in the NEFMC region or why SAW are not issued with clear guidance on how to develop BRPs or depart from defaults. I think there is confusion as to the role of the SAW and managers in the risk management process and this confusion needs to be clarified in order that good science processes can operate. I have made a recommendation related to these issues.

BACKGROUND

*The main body of the report shall consist of a **Background**, Description of the Individual Reviewer's Role in the Review Activities, Findings of whether they accept or reject the work that they reviewed, and an explanation of their decisions (strengths, weaknesses of the analyses, etc.) for each ToR, and Conclusions and Recommendations in accordance with the ToRs.*

Atlantic surfclam

Atlantic surfclam (*Spisula solidissima*) is a large (up to 22 cm), fast growing, bivalve mollusc with a lifespan of the order of 35 years or more. Effective maturity is not known but surfclams can contribute to spawning at only a few months old. Fishery selectivity by length corresponds to 6 years in southern waters and 7 years in northern waters. Surfclams are found in a wide variety of substrates but are most densely distributed in medium-grained sand. Surfclams are distributed from the Gulf of St Lawrence to Southern Virginia/North Carolina with concentrations on Georges Bank, south of Long Island, New Jersey and the Delmarva (Delaware, Maryland, Virginia) Peninsula. Surfclams are distributed widely from inshore to offshore but are found at greatest densities in depths less than 40 m except where inshore waters are warmer (e.g., Delmarva).

For the purposes of assessment, surfclams have traditionally been considered as a single stock in US waters. Due to Paralytic Shellfish Poisoning (PSP), the surfclam fishery on Georges Bank (GBK) was closed from 1989 until 2009, effectively being at an unexploited level in 2009. Catch data from GBK are limited due to closure, and surveys have generally not concentrated on GBK. Stock assessment for the area is therefore problematic, as pointed out during SARC 49 (in 2010) and in the SAW/SARC 56. A major consideration for the SAW at SARC 56 was surfclam stock definition(s) and how to structure stock assessments.

The commercial surfclam fishery in the US EEZ has operated under a fisheries management plan (FMP) and quota system, implemented through ITQs, since 1990. Total catches are estimated using allowances for incidental mortality and, historically, for discarding. Landings have varied very little since ITQs were introduced, averaging close to 20,000 mt meats per year. The fishery operates only within a small area of the total surfclam distribution and within that small area is highly concentrated on specific high value areas, determined not just by surfclam catch rates but also by economic factors related primarily to positions of ports and processing plants.

The stock of white hake was last assessed in 2010 using a relatively simple delay-difference biomass dynamics model (KLAMZ). In 2013 a new statistical catch-at-age model has been implemented in SS3. In 2010, the stock was estimated to be neither overfished nor subject to overfishing. With the development of a new model, including new data on survey size selectivity and dredge efficiency as well as age and length compositions, there is a need to establish new reference points to guide status determination.

White Hake

White hake (*Urophycis tenuis*) is a demersal gadoid distributed from Newfoundland to North Carolina, with greatest abundance in the Gulf of Maine. White hake live to slightly over 20 years of age and mature early, with 50% maturity for males and females at less than 3 years. It is recognized that there may be more than one white hake stock and that structure within and between putative stocks is possible. For the purposes of assessment, however, and in the absence of data to inform stock structure or disaggregated assessments, a single stock assessment for white hake in US waters is undertaken. The stock status of white hake in Canadian waters is not considered.

Landings (state unclear) of white hake of the order of 17,000 mt per year are reported from 1893, with a high of almost 22,000 mt in 1898. By the 1920s annual landings had reduced to the order of 10,000 per year and by the 1950/60s to 2-3,000 mt per year. Since 1964 landings have varied considerably from a low of 1,147 mt in 1967 to over 8,000 mt in 1985 and over 9,500 mt in 1992. In more recent years, landings have fluctuated in a lower range, generally less than 2,000 mt but near 3,000 mt in 2011. Catches in US waters have primarily been from the Gulf of Maine but with Massachusetts contributing high proportions in some periods, including over 80% in 2011. The majority of catches in recent decades have been by otter trawlers. Precise definition of catch history is complicated as landings are generally not of whole fish, requiring conversion using landed lengths to whole fish lengths and weights, and a mixed landings category for red hake and white hake. There is a relatively small recreational catch.

The stock of white hake was last assessed in 2008 at GARM III. At that time, SSB was estimated as 19,800 mt, well below the SSB_{MSY} reference point of 56,300 mt (the ratio of the point estimates being 35%, below the overfished threshold of 50%SSB_{MSY}). The fishing mortality rate in 2008 was estimated to be below 0.15, above the defined F_{MSY} of 0.125. At GARM III, therefore, white hake was assessed to be overfished and subject to overfishing.

The SAW in 2008 and the GARM III Review Panel had a number of reservations about the stock assessment. In SAW/SARC 56 these have been addressed and there is a new stock assessment, resulting in revised reference points and a revised determination on stock status. The key issues in the new stock assessment, covered in the main part of this report, are the use of pooled Age Length Keys in new statistical catch-at-age models and revised catch streams (both identified for further work at GARM III). Given the new stock assessment there is a need to establish new reference points to guide status determination.

REVIEW PROCESS

Reviewers shall provide a critique of the NMFS review process, including suggestions for improvements of both process and products.

The 56th Stock Assessment Workshop/Stock Assessment Review Committee (SAW/SARC), considering benchmark stock assessments for Atlantic surfclam and white hake, took place at the NEFSC, Woods Hole, MA, from 19th to 22nd February 2013. The review was hosted by NEFSC.

Participants in the review are listed in Appendix 3. The SARC Review Panel comprised an appointed Chair (Houde; MAFMC SSC) and three CIE reviewers (Cryer, Smith, and Stokes). Rapporteurs for all sessions were drawn from the NEFSC Population Dynamics Team. The Panel was tasked with providing separate SARC Reports for Atlantic surfclam and white hake. The Panel was also tasked to work with the Stock Assessment Working Group (*via* the Chair, Lead Analyst, and other members present) to develop an agreed Summary Report for each stock. CIE participants are further tasked with providing independent reports (of which this is one).

The SARC 56 review meeting (see Appendix 3) included numerous staff from the NEFSC and a wide range of industry, academic and other participants. I am not aware of any problems with notification of the meetings and interpret from the presence of stakeholder representatives and the public, and lack of explicit complaint, that notification was appropriate. All participants were able to participate throughout the meeting and opportunity was explicitly given for input. Many non-Panel participants contributed usefully to discussion and I believe that all were provided appropriate opportunity for involvement both during the Panel meeting and during extra-mural discussions. I note that in advance of the meeting, the Panel Chair refused a request to make a presentation on surfclam stock definition but noted the opportunity to use public comment sessions for this purpose; during the meeting that opportunity was taken up.

Notification of the meeting and dissemination of papers followed fairly closely the schedule laid out in the CIE Statement of Work (see Appendix 2) but white hake documents were unavoidably received late due to the lead analyst suffering a broken leg; the documentation was nevertheless received in advance. For both white hake and surfclams, main assessment reports and draft summaries, together with past reviews and some background papers, were provided in advance *via* a dedicated web link (see Appendix 1). No presentations were made available in advance. Overall, administration of the review was sound. Other regions use different methods for making information available. The use of ftp is common and it is not unusual for materials to be made available at least to reviewers also using, e.g., Dropbox, Google Drive, or similar cloud products. All of these methods allow much simpler synchronization of materials before, during and after meetings. The weblink method used by SARC was frustratingly cumbersome. I would **recommend** that SARC considers the use of ftp or cloud-based alternatives.

While papers were received in advance, it must be noted that the quality of the white hake report in particular was not of the highest standard with respect to a number of ToR, including new model development. I am hesitant to criticize given the lead analyst was indisposed but see the

report production as a SAW responsibility which should not be dependent entirely on a single person; certainly, I would expect sections of the SAW to be written by separate individuals. In my opinion, the report required more considered explanations in a number of areas, including model development, with explanations of modeling choices and reasons for decisions. Just one sentence conveyed that over 30 model configurations were explored before the paragraph continued to outline final model configurations. As documentation of new assessments and subjects of review, I would have expected more discussion on model development with selective tabular and graphical backup to explain SAW decision making, build confidence in the base run, and explore sensitivities. Even if not reported by the SAW, the copious technical materials that support model development should have been available through SAW working documentation available electronically and/or as hard copy during the meeting. As the assessment was new in 2013, and subject to review, a more detailed report would have been expected; for simple updates, the expectation would of course be lower. In general, across ToR, more information on SAW work and decision making would have been helpful. As a lead-in to review the documentation was arguably insufficient. The surfclam SAW report, in contrast, was more fulsome in coverage of the new model testing and development: explanations were overall more complete with discussion on initial model building and testing, including consideration of growth estimation, weighting of survey indices versus composition data, and an investigation of dredge efficiency. The surfclam report, however, did not contain sufficient detail to allow the Panel fully to consider ToR 3 (stock definition) and unhelpfully avoided meeting the ToR by passing the task to the Panel. This was inappropriate. The surfclam report also did not fully consider key information for the new assessment (age composition quality).

The problem of poor reporting (particularly for the white hake stock assessment) and lack of available working documentation (for both stocks) was compounded during review by the very short time available for review and tendency for rapid presentation. The scheduling of a full day for each stock was also unhelpful in that it precluded making early requests for additional work (a half day on each stock on day one would have allowed requests to be made for the first afternoon and evening). The presentations generally contained little that was not in the main SAW reports and did not aid the Panel by providing further insight in to key issues (especially reference points for white hake, and stock definition for surfclams). The SARC 56 review allowed effectively just three days for two stocks, with a fourth day scheduled for SARC report writing. Within the three day block, SAW Summary reports also had to be worked on. Overall, less than one and a half days were available for review of each stock, including writing of SAW Summary reports, with little time and opportunity for discussion or for requests to be made and responded to. I note also that the NEFSC SAW Chairman made clear at the outset that requests should be limited and not cause undue work or late nights.

With regard to presentations, some statistics are useful. For white hake, a total of 175 slides were provided on the second day. Of these, 129 were primarily on data issues but only 46 covered all of modeling, reference points, status and projections. None were available in advance. A new model was developed, an old model was used for bridge-building, and other models were used for data usage exploration, but only a shorter presentation was available for these, contentious

reference point issues, status determination, and projections. 175 slides in one day is too many by any standard, especially when many of those slides require careful scrutiny and when consideration of them leaves no time for delving in to issues not adequately covered in the SAW report. The balance of slides for presentation was wrong – more careful exploration of the modeling, reference point definition and projections was necessary. For surfclams, the total number of slides presented initially, again not available in advance, was 135, but in this case with the majority (77) focusing on modeling rather than data. This was a better balance, although the number was again arguably too high given the limited time available and the need to deal explicitly with a major point of contention (ToR 3 on stock definition). I may be in a minority, but it is my opinion that presentations are often far too long during reviews, too repetitive of materials already presented in reports, and do not take the opportunity to explain in more depth. I was struck during 2011 while participating in multiple STAR reviews that the most useful reviews were focused around shorter, smarter presentations, with analysts willing to explore options and undertake requests. This allowed more confidence to be built in models and advice.

I believe there was consideration of review processes at a national meeting of SSC in 2010 or 2011. It may therefore be redundant but I would **recommend** SARC processes be evaluated in relation to equivalent processes elsewhere in the US and to the regional management context, with a view, if necessary, to revision. I note issues raised at white hake ToR 5 are pertinent in this respect, notably a further recommendation to NEFMC regarding development of guidance to SAW on issues pertaining to risk and BRP. I would also note that as model usage moves more towards integrated statistical approaches implemented in, e.g., SS3 and ASAP, there is a need for more consideration and elaboration on model diagnostics and fuller exploration of uncertainties (affecting status definition, reference point choice, and projections).

Because of problems with reporting of work on white hake reference points, requiring substantial time on the third and final days, the Panel did not complete the SARC report draft as intended, causing some strain on meeting reporting deadlines. The report writing therefore continued as an after-meeting task, with reviewers spread across multiple time zones. The difficulties of white hake reference point definition are covered under ToR 5 for white hake, but it is worth mentioning at this stage that the problems arose despite careful SAW report wordsmithing but because of a failure to include complete information in support of arguments made (wordsmithing is important but is not a substitute for substance); due to failure of the SAW in any case fully to consider model uncertainties (or at least report such work); frustrating performance during the review meeting with white hake analysts unable quickly to provide simple, definitive results on risks associated with reference points; external inputs by e-mail (from non-participating SAW members *via* a meeting participant); and late realization by NEFSC that already slow responses to requests on risk levels associated with reference points were erroneous and in need of correction. This last issue may have caused frustration but it was nevertheless good to see NEFSC managers being upfront and transparent and, despite deadlines and time constraints, insisting on numbers being carefully checked.

The issue of white hake reference point definition is clearly contentious and hindered somewhat in the northeast USA by apparent lack of clear guidance to SAW on default reference points, acceptable levels of risk, agreed methods for determining risk, etc. In the absence of management-set standards, the SAW took on the role of risk manager (not just risk analyst) and adopted a working risk criterion against which to derive reference points. In the absence of appropriate management input, I have no problem with this, or particularly with the methods and risk criterion chosen. I would comment, however, that in the absence of standards (often absolute), the SAW approach of working around a low central tendency of risk associated with a range of plausible scenarios is sensible. These issues are covered more fully under ToR 5 for white hake. My problem with the white hake reference point issues that arose at SARC 56 is not fundamentally to do with the substance of what the SAW attempted; it is to do with incomplete reporting by the SAW and the Review presenters/analysts, and subsequent messy process.

The Terms of Reference (ToR) for the review are given in Appendix 2, Annex 2. Often, reviews including CIE experts focus on a particular phase of the stock assessment process – either the data inputs or the assessment *per se*, and often deal only with a single stock. The ToR set for the SAW/SARC 56 review is very wide, spanning for each of two stocks, data quality (including collection and analysis), the stock assessment, status advice and projections. I have noted in previous CIE reports that it was not always possible to devote as much time as would be desirable to every issue area when two stocks and wide ranging issues are to be considered. This was a particular problem in SAW/SARC 56 given the constrained time available.

My overall impression of the SAW/SARC 56 process is that there is much that could be done to improve it. My interpretation and **recommendation**, admittedly from a limited view, is that a fundamental review is needed by the relevant Councils, with input from SSC, with a view to providing guidance to SAW on reference point setting. Further, as **recommendations** to the NEFSC/SARC organizers, that the standard and substance of SAW reports needs to be consistently high, especially for new assessments; that if reviews are to be of value, the total time, scheduling, and materials available need to be reviewed; and finally, some thought needs to be given to ensuring that presentations are smarter and provide opportunity for in-depth discussion and exploration during meetings.

REVIEWER'S ROLE IN THE REVIEW ACTIVITIES

The main body of the report shall consist of a Background, Description of the ***Individual Reviewer's Role in the Review Activities****, Findings of whether they accept or reject the work that they reviewed, and an explanation of their decisions (strengths, weaknesses of the analyses, etc.) for each ToR, and Conclusions and Recommendations in accordance with the ToRs.*

The role of the reviewer is set out in the CIE Statement of Work, Attachment A, attached here in Appendix 2, Attachment A. CIE reviewers are tasked with producing an independent report to the CIE. The reviewers are additionally tasked with writing the SARC Panel Reports for each of

surfclams and white hake. The reviewers are further tasked, together with the SAW present, to develop the SAW Summary reports for each stock.

In addition to *becom(ing) familiar with the draft stock assessments(s) and background materials*, I participated in all discussions and contributed to the SARC reports (leading on white hake, primarily editing on surfclams). The surfclam report was finalized reasonably quickly but the white hake report was problematic, primarily due to late receipt of definitive numbers and confusion during the review leading to confusion amongst CIE reviewers. The three CIE reviewers worked collaboratively by e-mail to ensure the SARC reports were finalized; this was complicated by the reviewers and chair working across three time zones spanning 18 hours. The white hake SARC report was left with the Chair for final editing only at close on 6th March and was sent to the SARC Chair on 8th March. Editing of the white hake SARC report was a painful exercise that required far more of my time than anticipated and which interfered with writing of my individual CIE report. During the meeting, along with other Panel members, I participated in development of the SAW Summary reports. The surfclam report was finalized but the white hake report was awaiting final modification pending white hake reference point results and SARC Panel reporting. The final SAW Assessment Summary report for white hake was sent from the Panel to the SARC organizer on 9th March. This report was due to be submitted to the CIE on 8th on 9th March but due to difficulties with other reporting was delayed. A final draft was sent on 9th March to allow CIE staff to take an initial look. The final version (this one) was sent on 10th March.

FINDINGS BY STOCK AND ToR

Atlantic Surfclam

The main body of the report shall consist of a Background, Description of the Individual Reviewer's Role in the Review Activities, ***Findings*** *of whether they* ***accept or reject*** *the work that they reviewed, and an explanation of their decisions (strengths, weaknesses of the analyses, etc.)* ***for each ToR****, and Conclusions and Recommendations in accordance with the ToRs.*

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

Landings data for surfclam are considered to be precise due to the strict cage tagging system in place for ITQ monitoring since 1990. The landings data are tagged by ten minute square (TMS) and are thus also reasonably well spatially resolved although analyses generally are at the regional (i.e. Delmarva, Southern New England, etc) level. The SAW report provides conversions for cages to bushels, bushels to cubic feet, lbs meat weight ("meats"), and kg meats. These factors seem to be used to convert landings to precise meat equivalents. My understanding, however, is that bushels and perhaps also cages may have a number of volumetric interpretations and hence landings by meats may not be as precise as presumed. The conversion from bushels to meats must also presumably vary depending on fishery size structure variations by region, depth, substrate, season, etc and the impacts of repeated fishing on key grounds. Figures A14-A18 are hard to interpret visually but there do appear to be regional and annual differences in length compositions. I accept that industry fishing practices are aimed at capture of consistently high surfclam size but do not see that this should counter the variability that needs to be reflected in landings data. I **recommend** the SAW investigate this matter and note a previous research recommendation (number v) against which no progress has been made. If conversions factors are not stable then this will need to be reflected in the assessment which currently treats catches (landings plus other components) as exact. Potentially, if changes have been systematic, the total landings stream may need to be revised.

Discards are reported in the SAW as a high but declining proportion of landings from 1982 to 1993 and as zero thereafter due to removal of a minimum size regulation. It is not clear from the report if zero discards are assumed, observed or estimated. From discussions during review there were no indications that discards could be above zero. In the context of the assessment (very low F) this does not seem to be an issue but clearer reporting would be appreciated.

For many years, the assumed catches have been defined as landings inflated by 12% to account for "potential incidental mortality". The explanation given for this in terms of fishery area relative to the resource distribution, and regulatory issues, is not clear to me. Incidental mortality is usually a factor related to the way gear impacts non retained individuals. The SAW seems to interpret incidental mortality as due to activities beyond the fishery (often referred to as "other mortality"). This needs to be clarified. If there is no additional mortality on surfclam fishing grounds due to surfclam fisheries then incidental mortality is zero. If there is mortality of

surfclams away from surfclam fishing then this is other mortality and needs to be recorded appropriately. For such a high value ITQ species it is hard to see how such other mortality would be tolerated. I suggest the SAW make clear exactly what is assumed. It is also unclear how the 12% figure was derived. Again, this needs to be made explicit.

The SAW provided a good description of spatial and temporal patterns in landings and a simple temporal description of discards. The spatial and temporal description of landings extends beyond region to include important TMS. Effort changes by region and year are well described. Additional economic and fishery information is provided as background but is not used subsequently in the assessment.

Overall, although the explanations and references are lacking in the SAW report, there seems to be a reasonable accounting for removals but a lack of characterization of uncertainty in these.

2. *Present the survey data being used in the assessment (e.g., regional indices of abundance, recruitment, state surveys, age-length data, relevant cooperative research, etc.). Investigate the utility of commercial LPUE as a measure of relative abundance. Characterize the uncertainty and any bias in these sources of data.*

Although landings and effort patterns are well described at ToR 1, attention to the utility of LPUE is very brief (with comment of relevance on p 10 and specific comment on p23). The SAW did not report any LPUE standardization or analyses to investigate utility but limited its comments to i) noting that LPUE is unlikely to provide a good biomass index for sessile species subject to highly targeted fishing, and ii) while referencing the previous assessment, noting that trends in LPUE are highly correlated with survey trends in DMV and NJ where fishing has been concentrated and is widespread. Regardless of (i), given (ii), it is disappointing that the SAW did not attend more fully to this aspect of the ToR (*Investigate the utility of commercial LPUE as a measure of relative abundance*). The SARC Panel also noted in the SARC report under ToR 4 that patterns of LPUE in some southern regions were similar to estimated trends and that LPUE could therefore potentially serve as a useful index of abundance. I note this because LPUE are annual indices while survey indices are generally triennial with plans to change to a rolling survey covering different regions in different years. Where LPUE do appear to reflect biomass trends they could provide considerable information to the assessment model. Their inclusion would also change the balance of index weighting with composition data (see ToR 4).

The SAW provided an extensive and detailed description of survey design and methodology, current technical testing (e.g. on determination of time fishing by examination of dredge critical angle) and improvements, and data provision of indices and length and age compositions for the assessment. The SAW covered in great detail cooperative experiments to estimate dredge efficiency and size selectivity, key inputs to the assessment model with dredge efficiency a primary determinant of biomass scaling. The SAW reporting of the impressive work on NEFSC and cooperative surfclam surveys is extensive, occupying almost one half on the substantive report. The work reported is of a high quality with fine attention to detail; this and the extent of

cooperative research, involving government, industry and academia, reflects the economic importance of the fishery and serves to illustrate the potential for cooperative research generally at a time when there is considerable national (USA) interest.

Work on technical aspects of survey measurements is well described and builds confidence not just in primary survey results but also in the experimental work to estimate dredge efficiency and selectivity. The SAW provided comprehensive descriptions of the NEFSC survey results by region and year for indices and composition data.

The dredge efficiency work stands out as key in SARC 56 as it leads directly to a major model input, the prior for efficiency/catchability (see ToR 4). The depletion experiments are conceptually simple but estimation of uncertainties in efficiency estimates of the survey gear is problematic and was previously underestimated. The work reported approached this problem by using information from repeated stations (between survey years) and from new depletion experiments, and by integrating all available information on survey dredge efficiency to develop a lognormal prior. The prior had a median for dredge efficiency similar to the equivalent estimate used in previous assessments but with a much larger CV (allowing the assessment model the possibility of estimating higher catchability). The work reported is of a high standard and the analyses are important. As noted by the SARC Panel, the work has highlighted uncertainty in dredge efficiency estimation and strengthens the intention to move towards use of commercial dredges for future surveying. I would encourage this as no matter how excellent the work to estimate a prior on survey dredge efficiency, the issue of uncertainty will remain. Given the lack of information in other data with which to inform the model about efficiency (see ToR 4), this is important in the current modelling paradigm (but see ToR 3 and Conclusions).

Work on size selectivity for the survey and the commercial (in survey mode) fishery gear was described. I am, however, slightly confused by the description that says on p18 that GLMM were used for this assessment while the description of current selectivity estimates on pp19-20 describes the use of GAMs. From the formulation provided the model used appears in fact to be a GAMM. Regardless, the modelling confirms previous selectivity analyses that suggest selectivity for the survey dredge is domed, with maximum selectivity in the region of 100mm to 150mm shell length. This differs substantially from fishery selectivity which is estimated to be asymptotic from about 120mm shell length. The analyses show clearly that the domed survey selectivity, though variable in detail by area appears to be robust across areas/stations. I note that the selectivity estimates derived were used in bridge building with the KLAMZ model as well as provided as fixed parameters in the new model implemented in SS3.

While the selectivity work is well reported, the description of age and growth is only briefly dealt with even though it is important for data provision and direct parametization of the new assessment. Given that the new model uses survey age data (allowing YC estimation) for the first time and that this is given as the primary reason for developing the new model, I would have expected a greater consideration of the quality of survey age data sampling and of ageing itself. Only two brief sections make mention of the sampling and ageing and neither discusses issues of

relevance to potential uncertainties that may be important in the modeling. During the review meeting the Panel requested information on ageing and length/age sampling and was provided with some references and links to work on ageing precision and validation but this was too late and in any case only partially relevant.

I find it unsatisfactory that a report can fail completely to include information on key aspects of assessment, especially when the assessment is new and the stated driver is the ability to use such information. Reporting needs to be focused and relevant.

In terms of the ToR requirements, therefore, while accepting the SAW met most of its responsibilities, I think it fell short of characterizing uncertainty and biases at least in survey age and composition data. This is important and I **recommend**, if the new model is to be used in future, that the SAW review these fully and take account in future modeling of any uncertainties.

3. *Evaluate the current **stock** definition in terms of spatial patterns in biological characteristics, population dynamics, fishery patterns, the new cooperative survey, utility of biological reference points, etc. If appropriate, recommend one or more alternative stock definitions, based on technical grounds. Integrate these results into TOR-4.*

The ToR requires a consideration of alternative stock structures but the SAW apparently concentrated considerable efforts on looking at the pros and cons of arguments and evidence in support of just two predefined alternatives: a single stock (as previously modeled) or two stocks (the northern or GBK region as a single stock and all other regions combined in to a southern stock). The main body of the SAW report contains some general information pertaining to the issue but the bulk of the SAW commentary is contained in two tables structured in relation to the components listed in the ToR (biological characteristics, population dynamics, etc). The SAW did not reach a conclusion as to how to proceed and explicitly left the task to the Panel, apparently trying to mitigate dissention and heated debate. In my opinion this was less than helpful and an avoidance of responsibilities. The SAW should contain the relevant expertise and have the available time made available to address all ToR. If the time cannot be allocated, the ToR should be reconsidered; it is a matter of planning. The Panel had very little time to review the entirety of the SAW work and could only work from the limited evidence presented to it. I have no problem with a SAW stating that it has not been able to reach a conclusion and asking for input and advice. I also have no difficulty with Panels disagreeing with SAW and even over-riding SAW recommendations. I do have a problem with SAW passing on a problem in so blatant a manner and then bringing obvious heat in to the review without adding anything of major substance. The process could and should have been better managed.

The arguments set out in the tables are not detailed and evidence is by reference only and therefore hard to follow. There are some key issues of obvious relevance (spatial patterns in life history characteristics, population dynamic variations, oceanography, etc). It does appear that there is a potential difference in some of these between the GBK and other regions but it is

unclear from the presented arguments (in the SAW report or during the review) if this is enough to justify splitting the GBK area and defining it as a "stock". It is also unclear if there is further structure in the southern area that could lead to defining other "stocks". The only possibly compelling evidence for distinction is the referenced work on larval drift, but even this is unclear for reasons provided in the report (Table 17). I note that stocks can be defined on a purely genetic basis but this is seldom satisfactory. Consideration of spatial patterns and oceanography is a good basis for functional separation of assessment and/or management units regardless of genetics.

What is not clear to me is why redefining stocks is necessary in this case, apart from issues to do with running the stock assessment(s). Nor is it entirely obvious what the legal and economic ramifications of redefinition are and why the issue is therefore so heated. Although not strictly related to the SAW or Panel remit, I am of the opinion that these issues must be taken in to account in any redefinition of stocks and/or management units. Fisheries management is not just about biology. Rather than debating a limited set of stock definition alternatives it would make more sense to consider from a management perspective the risks and benefits of alternative spatial management approaches given many uncertainties about surfclams, including stock issues. This seems to me to be a fundamental issue that needs consideration.

This ToR relates to the splitting of assessments to provide standard outputs for the management system. The whole assessment approach is designed to provide specific outputs but it is questionable whether or not this is worthwhile. Why is it necessary to use a state-of-the-art, data-intensive assessment to provide outputs of little relevance to the fishery? Simple calculations show clearly the approximate status of the surfclam stock(s) (see other sections) and it is not obvious that more refined estimates of fishing mortality and summary biomass in relation to proxy and debatable BRPs (see also ToR 5) adds anything to the management process. Those simple methods would probably suffice to provide *pro forma* advice on stock status. More detailed modeling of how fishing strategies focused on small areas may or may not lead to local or stock level risks would appear potentially far more useful and could inform smarter, reactive management accounting for biological, changing environmental and economic factors. I note that depending on how the fisheries are prosecuted, but given the generally low exploitation rates focused on key areas, even the survey timings and extent could be reconsidered if informed by appropriate management strategy evaluation. The main point here is that if the modeling approach (not the assessment *per se*) were smarter then there might not be a specific need to predefine stocks as required by this ToR.

Regardless, in relation to the ToR the SAW addressed the issue but focused too narrowly on a definition of one *versus* two stocks. From the limited evidence of the SAW report and discussion during the review meeting it appears the SAW considered a wide range of information and I have no doubt the discussions were in-depth. My interpretation from the review meeting is that consensus could not be reached on technical grounds (as required by the ToR) but also due to consideration of wider implications. That no consensus was reached should not matter as the

approach taken (see ToR 4, 5, 6, and 7) to separately model (and then add outputs) provides a sufficient basis for decision making.

4. Estimate annual fishing mortality, recruitment and stock biomass (both total and spawning stock) for the time series (integrating results from TOR-3), and estimate their uncertainty. Include a historical retrospective analysis to allow a comparison with previous assessment results. Review the performance of historical projections with respect to stock size, recruitment, catch and fishing mortality.

The SAW has reasonably well described its model development and fitting process and has provided extensive diagnostics for the final model run.

As noted at ToR 2, the driver for the new model was that use of age data allows the estimation of annual recruitments and that SS3 allows more flexibility in future assessments than the previously used KLAMZ model. As noted at ToR 2, if age data are to be used then the SAW has a responsibility to explain and evaluate those data before proceeding to modeling. It is important to understand potential uncertainties in the age data in order to make judgments about data weighting. I am not convinced that the SAW has not over-weighted the age data and would like to have seen a fuller analysis.

Notwithstanding, the SAW approach to model development seems generally sound, with considerable effort expended on initial growth estimation. I note, however, that growth, selectivity, and ultimately catchability estimation are all related, and that initial growth estimation is therefore a critical step if the parameters are to be kept constant thereafter as model exploration proceeds. The SAW appears to have decided upon the specific growth parameterization using an otherwise consistent model and with reference to its perceived credibility of the resulting growth estimates. What is not clear is how those alternative growth estimates affected other parameter estimation or if alternative model weightings were also explored as part of the exercise. It is common to try to pin one parameter set early in the development phase but as this model effectively has only to balance how it fits age and size compositions with growth and prescribed selectivity (and how to balance that with any tension from competing index information) it would have been useful and possible to explore all at the same time while varying weightings. It would in any case also have been useful to make initial explorations also under alternative M assumptions (see also comments on M and selectivity, below).

I am concerned also that even though the model wanted to fit broader domed survey selectivity when allowed (p26, para 2), though it is unclear under what growth assumptions, that selectivity was nevertheless forced, with constant M, leaving only fishery selectivity and growth (and recruitment) estimable. This is for the southern area model. For the northern area only recruitment was estimated. As the SAW notes, growth in the southern area has changed through time and as survey selectivity estimation could therefore be less precise than indicated through the GAMM, a fuller exploration, or exposition, of the model fitting would have been useful.

Three things most concern me about the model. First, is the estimation of q, as seen through Table A27. The prior presented to the model has a wide CV but is highly peaked with a mode little over 0.1 and a mean of 0.23. The model estimate is 0.33, I suspect at about the 90th percentile of the distribution. This implies that the model either has a great deal of information on q and/or that the likelihood surface is very flat, potentially allowing multiple interpretations of the data depending on prescribed parameters (survey selectivity and M) and data weightings used. Table A27 suggests the primary information used by the model to shift q is the fishery and particularly survey lengths. The survey ages effectively have no influence. It is not obvious that the length data should have this effect. At a minimum, it would have been useful during review to have explored an extended likelihood profile with looking at how q estimation interacts with survey selectivity and growth assumptions. My overall impression is that q is ill-defined and that absolute biomass estimates should be treated with great caution. I accept that ratio BRPs are less sensitive but absolute numbers matter in determining catch options.

My second concern is the lack of exploration of M. The previous SARC panel and this one were concerned that M may overstate productivity. I agree with this. I note the SAW tried to improve the lack of fit to larger commercial sizes by allowing M to be freely estimated on older ages. What the SAW did not do was explore a lower constant M or lower M with free estimation for older ages. In order to improve the fit, the SAW adopted a domed shaped selectivity for the southern area assessment base case (but asymptotic for the northern area), arguing that the domed selectivity represented both true selectivity (asymptotic) and availability (reduced for larger clams through exploitation). This seems to me to be wrong. Selectivity is a distinct, physically and behaviorally determined process. Availability may be biologically determined and can in that case be folded in to selectivity, but where variation in availability is caused by the fishery (as assumed by the SAW) this needs to be factored in to estimation of F. Use of domed selectivity in this case is wrong.

My third concern is data weighting. The large composition sample sizes can lead to composition data swamping index data. In the case of this assessment it is clear by comparison to the KLAMZ model, most easily seen through the historical projections at Fig A83, that use and high weighting of the age data, leads to a major change in perception of trends for the southern+GBK models, driven especially by the trends estimated for the southern area. There is also an overall scaling difference driven by the q estimation. As noted above, I think there is need for further exploration to balance data weights and investigate alternative M, growth and survey selectivity. Work to understand uncertainties in length and age sampling and in ageing would be useful in guiding weighting decisions. I note also that the LPUE are not used as indices but as commented on at ToR 2 may be useful indices, providing annual information on trends.

Overall, the new model generally fits composition data well but it seems to miss trends as seen through survey indices (modeled) and LPUE (not modeled). The change in estimated trends from previous modeling is driven by the inclusion of age data which may be over-weighted. The scaling of the new model is apparently driven by commercial and fishery length frequencies but it is unclear why this should be so and I would expect an interaction with the way in which

fishery selectivity and growth are modeled. There would be interactions with all modeled components if M were varied. Although the model is Bayesian, it has been run only to MPD. Uncertainty in estimates of B and F does not take account of variability in survey selectivity, trends and fluctuations in growth, and misspecification of M.

I am highly impressed at the efforts of the SAW and of the research programs that have fed in to the assessment, especially to estimate dredge efficiency and size selectivity. The SAW has made a major effort to explain how it developed the new model. In doing so it has exposed problems. I am not convinced the new model(s) are a step forward from the old KLAMZ models and think more work is needed to provide confidence in their outputs. At ToR 3, however, I commented that alternative modeling approaches might be more useful to inform management of surfclams than either KLAMZ or the new model. I am therefore reluctant to recommend anything specific with respect to the new assessment model. I would suggest that if the model is to be used in future, then a more detailed examination is needed. At this time, however, I accept the model gives credible ratio BRP estimates sufficient to inform management (see ToR 5, 6, 7).

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. This should be carried out using the existing stock definition and, if possible, for the recommended "alternative" stock definitions from TOR-3. If analytic model-based estimates are unavailable, consider recommending alternative measurable proxies for BRPs. Comment on the appropriateness of existing BRPs and the "new" (i.e., updated, redefined, or alternative) BRPs.*

The SAW carefully explained that while the assumed M was 0.15 in both the old KLAMZ biomass model and the new model implemented in SS3, and was used with both as an $F_{MSYproxy}$, the interpretation had to be treated with care. The old $F_{MSYproxy}$=0.15 refers to all surfclams above 120+ mm shell length whereas the new $F_{MSYproxy}$=0.15 refers to surfclams of 160/170+ mm shell length depending on area. The fundamental difference is that the new $F_{MSYproxy}$ is more conservative than that used previously because it references a smaller biomass component (of larger surfclams).

The SARC 49 Panel commented that M=0.15 may overstate stock productivity of surfclams and during SARC 56 all Panel members were concerned that this might be so. I think the SAW should have explored further model sensitivity to M, especially in light of its adoption effectively of a different M (as it refers to a different population component when using it as a basis for BRPs). The SAW presentation during review included some simple catch curve analyses for the southern area suggesting Z might be of the order of 0.16. It also presented simple catch over minimum swept area biomass ratios which suggested exploitation rate had increased from about 0.05 after 1995 to near 0.15 in 2011. These simple calculations strongly suggest that M is overstated. During the review the Panel noted that analyses looking at proportions at age from surveys could be performed on the GBK potentially to obtain a better estimate of M given the long history of closure, or from non-exploited small areas from the southern region. The SAW

responded by providing Z estimates by survey year for the GBK. The estimates of Z varied considerably between years with a mean and median just under 0.10. Again, the indication is of an M lower than the assumed 0.15.

It is worth noting that while the $F_{MSYproxy}$ refers to a population component of 160/170+ mm shell length and above, the summary biomass measure used for biomass reference points refers to biomass above 120mm+ shell length. This is potentially confusing and should be made clear in all advice. As the summary biomass already indexes only a subset of the SSB it is difficult to make simple sense of the BRP set.

The SAW explained the continued use of B_{1999} as a proxy for virgin biomass and the use of 50%B_{1999} as a $B_{MSYproxy}$ for the southern area. This is superficially consistent with previous usage of B_{1999} as a reference year but is different in that previous usage was for the whole area stock. There is a lack of robust justification for the use of B_{1999} as a reference point. No justification for previous use is available and the justification for continued use for the southern area is that with the new southern area base case model, B_{1999} just happens to be near the (poorly) estimated virgin biomass level. The SAW should provide a firmer justification (but see below).

The precise choice of biomass reference point does matter at a time when the assessed southern area biomass is declining (likely due to environmental conditions) and when the current assessment suggests a small possibility of being overfished, which could grow in the near future. The SAW presented the distributions of B_{2011} and $B_{THRESHOLD}$ and there is a nominal 15.5% probability of overlap for the southern area. As discussed by the SAW, however, this probability is technically overstated as the two biomasses are correlated.

The SAW did not attempt to develop a biomass reference point for the GBK/northern area due to lack of information and argued that after a long period free from exploitation, the area should be near virgin. In fact, as seen in Fig A77, estimated biomass in the GBK has declined consistently from early in the closure period, suggesting a response to environmental conditions. This raises the question of what is meant by "virgin" in relation to reference points and why, for example, 1999 was argued to be a good basis for biomass reference points for the southern area.

If biomass is declining across the whole surfclam range due to changing environmental conditions, there is a need to reconsider with some urgency biomass reference points.

Overall, therefore, while I accept that the SAW has provided existing and new reference points as a basis for stating whether the stock is overfished and experiencing overfishing (at ToR 6), I think a clearer and firmer foundation for the reference points is required. If traditional stock assessment is to be used to determine stock status for surfclams (whether for one, two or more stocks), and standard type reference points are used, I **recommend** the SAW reconsider how to set biomass reference points and also explores ways of making F reference points relate to equivalent biomass components, as well as examining the credibility of M assumptions. I note, however, the issue of what constitutes appropriate modeling approaches needs to be considered

as does the issue of appropriate M assumptions; these are considered at ToR 3 and the conclusion.

In terms of existing or alternative (two stock) stock status, the SAW adopted a simple summation approach to the separate northern and southern assessments to derive biomass and fishing mortality estimates for the whole stock. This is a sensible, pragmatic approach.

6. Evaluate stock status with respect to the existing assessment model and with respect to any new assessment model. Determine stock status based on the existing stock definition and, if appropriate and if time permits, for "alternative" stock definitions from TOR-3.

a. When working with the existing model, update it 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).

This ToR refers to "existing" assessment, meaning the KLAMZ model, and "new" assessment, meaning the new statistical model implemented using SS3. The ToR also refers to alternative stock definitions as considered at ToR 3. The SAW undertook both KLAMZ and SS3 model runs for southern and northern (GBK) stock components and considered the sum of these for each model. The SAW, however, only reported results from the new model implemented in SS3. The SAW did report on model comparisons and did include details of the KLAMZ modeling results at appendix A5, but it did not make explicit comment on stock status using the "existing" model. I have gone through the KLAMZ results and it is difficult to read directly or to infer stock status. Appendix 5 is somewhat terse and the results, displayed graphically, are not well described.

With reference, therefore, only to the new model, and with the SAW-selected new reference points, the SAW provided detailed statistics on estimated summary biomass and fishing mortality for the whole stock, and separately for northern and southern areas, and related these to the SAW selected BRPs. For the whole stock, or for the northern area lone, there appears to be no possibility that the stock (or area) is overfished or experiencing overfishing. For the southern area alone, although overfishing is not occurring, there appears to be a 15.5% probability of being overfished. As noted at ToR 5, that probability is technically overstated.

I accept the SAW conclusions insofar as I think the assessment is meaningful, that M has been appropriately explored, and that BRPs are appropriate (see Conclusions).

7. Develop approaches and apply them to conduct stock projections and to compute the statistical distribution (e.g., probability density function) of the OFL (overfishing level) and candidate ABCs (Acceptable Biological Catch; see Appendix to the SAW TORs).

a. Provide numerical annual projections (3-5 years). 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.
c. Describe this stock's vulnerability (see "Appendix to the SAW TORs") to becoming overfished, and how this could affect the choice of ABC.

The SAW provided projections according to the ToR. These were reported in the SAW report and updated and extended projections were provided on request during the review. During review there was discussion about alternative states of nature to be included in projections as opposed to just alternative catch/F scenarios applied to the base case model. I would have liked to see alternative model runs with lower M and also to have had the opportunity to explore alternative BRPs (see ToR 5). A single attempt to run the model at fixed lower M during the review was not successful and more work was needed before any credible alternative state of nature could be considered. I think such work should be done if the current model is to be used as a basis for future advice.

Accepting the base run as given and as the basis for projections, the provided projections appear sound. The projection basis is clear and consistent with assumptions made elsewhere with respect to landings being inflated by 12% to reflect "incidental mortality" (though see ToR 1) and OFL being defined by $F_{MSYproxy}$=M=0.15. Setting the landings for GBK required an assumption of a constant catch (1,000,000 bushels =7,710 mt) to be applied to two constant catch scenarios (SQ and quota under the FMP). Given market limitations and the quota limits under the FMP, it appears that the SQ catch option is the most realistic. The OFL projections appear unrealistic. Based on the SQ scenario, results presented to the review suggest small probabilities of the whole stock or southern area component being overfished in the next five years. The SAW did not calculate similar probabilities but as $B_{2016/17}/B_{threshold}$ for GBK is projected lower than for the southern area, and because uncertain in $B_{threshold}$ and $B_{2016/17}$ is greater than for the southern region, the probability of being overfished would presumably be higher.

Recalling ToR 4 and 5, It is worth noting with respect to the projections that the assumed high productivity (i.e. M=0.15) is optimistic and needs further exploration. The projections could therefore understate the probability of becoming overfished. On the other hand, i) the assumed incidental mortality is pessimistic; and ii) results depend on the estimated selectivity with a large amount of biomass not accounted for. If selectivity is reasonably estimated then summary biomass will under-represent SSB and stock status could appear unduly pessimistic. It is also worth noting that use of B_{1999} as basis for B_{MSY} and hence $B_{threshold}$ is not well justified and the basis for the overfished definition may in any case be flawed.

Overall, Although I have reservations about the BRPs in use and consequent indications of status against those BRP, and also about the lack of exploration of the model at lower M, I consider the projections sound and that they provide a good indication of likely stock and fishery trajectories in the short-term.

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

Previous research recommendations were listed and usefully commented on by the SAW. A number of areas have seen no progress, some have been noted as relevant perhaps for academic research and some have had progress made. The SAW was not explicit about whether any existing research recommendations should be carried forward or deleted except for a recommendation relating to habitat on GBK. The SAW made seven new research recommendations (below). The listed recommendations were provided by the SAW without comment or justification. This is not uncommon practice amongst SAW in many regions but is unhelpful in priority setting or funding allocation decisions. If recommendations cannot be ranked by utility, feasibility, and cost it is hard to see value in making them.

i) Biomass reference points need to be reconsidered.
ii) Has surfclam biomass shifted offshore into deeper water over time?
iii) Look into a better way to implement regime change into the SS3 model. Look into patterns which may match other species and climate indices.
iv) Determine the best spatial and temporal distribution to use for surfclam assessment models
v) Look at habitat on GBK
vi) Given the increasing importance of GBK re-evaluate the optimal sampling design for the survey.
vii) Look into area specific recruitment streams for SS3 and how to accommodate the 2012 and 2013 surveys.

The first SAW recommendation relates to recommendation (8) of this report. I support it but note it could usefully be expanded. Recommendation (ii) could be evaluated using surveys and might be a consideration for new survey design. Generally, regardless of the question and means of gathering relevant information, I would suggest the use of information would best be through planned MSE rather than in the traditional assessment framework (see recommendation (i.e. encouragement) (7) of this report). Recommendation (iii) has two components; one is a software and methodological issue while the other is a matter of data analysis and perhaps collection. I would encourage the data analytic part of the recommendation but have no comment on the data collection implications. With respect to modeling I would repeat my comments at 2 and refer to Recommendation (7) of this report. I am not convinced that progress will be made trying to include regime change or distributional changes in to the new SS3 model which already faces estimation problems. It would also raise major issues regarding BRP definition. Recommendation (iv) is a result of lack of progress at ToR 3. The note by the Panel is reasonable but it is not obvious what research *per se* is being suggested rather than a continued evaluation as *per* ToR 3. I would note my previous comments at ToR 3 and at Recommendation (7). I would also note the Panel conclusion that assessment can reasonably be split and results used separately and in combination to inform management. Recommendation (v) is a continuation of a previous recommendation to determine how much of GBK is good surfclam habitat. As reported to the review, exploratory video work is already underway but future funding seemed to be in doubt. It is unclear to me how the work will feed in to stock assessment or other (e.g. MSE) approaches to informing management so I have no further comment. Recommendation (vi) makes sense and

seems to fit with general intentions to review survey design. I would note that survey design evaluation could usefully be part of any MSE work that is planned. Recommendation (vii) has two components. The first is software (SS3) related but depends on what is intended. SS3 is already adequately structured to permit "area specific recruitment streams"; it is what has been done in the new assessment, albeit sequentially. If the recommendation is to allow joint assessment with shared parameter estimation (or borrowing) but separate recruitment stream estimation then I am surprised SS3 cannot already accommodate this and agree it would be useful/efficient. The second component of the recommendation relates to a change in scale and possibly size-selectivity due to changing survey gear (to a commercial dredge). This is a far different issue which can be approached external and/or internal to the modeling. With respect to working internally, I would make the same comment as for recommendation (iii). It might be more fruitful to concentrate on ensuring data collection is sufficient during the surveys to allow splicing of time series developed using old and new gear based on measured selectivity and efficiency.

Only some of the SAW-provided research recommendations flow naturally from the SAW report. I prefer to see clear rationale and justification for any research recommendations to allow, as noted above, an idea of utility, feasibility and cost to be derived.

Atlantic Surfclam Conclusions and Recommendations

The main body of the report shall consist of a Background, Description of the Individual Reviewer's Role in the Review Activities, Findings of whether they accept or reject the work that they reviewed, and an explanation of their decisions (strengths, weaknesses of the analyses, etc.) for each ToR, and ***Conclusions and Recommendations*** *in accordance with the ToRs.*

Conclusions

Generally, the SAW reported surfclam assessment work with fine attention to detail; this and the extent of cooperative research, involving government, industry and academia, reflects the economic importance of the fishery and serves to illustrate the potential for cooperative research generally at a time when there is considerable national (USA) interest. I am very impressed by the cooperative experimentation and analyses

The SAW has well described the derivation of inputs to the model, especially dredge efficiency and survey size selectivity. I find it unsatisfactory that it has not included evaluation of age and length composition data, key aspects of the assessment, even though the assessment is new and the stated driver is the ability to use age composition data.

I am relaxed about the stock definition issue and think the discussion could be obviated if modeling was focused on supporting management relevant to the directed fishing in localized areas within a widely distributed stock (or stocks). Concentration at appropriate scales could allow better integration of biological, economic and environmental information in a way the

traditional stock assessment cannot. Given the size of the surfclam resource and given the low overall catches, simple approaches to informing stock status should be sufficient and suitable data are available for that purpose. It is not clear to me that the new SS3 model adds any clear benefits over the old KLAMZ model or even simpler approaches. I realize estimation of recruitment can potentially help to improve projections, but given the low overall exploitation rate and management to quotas under the FMP, it is not obvious that this is necessary.

My overall impression and conclusion is that it would be more valuable to adopt a procedural approach rather than an assessment one and would strongly **encourage** plans to use management strategy evaluation looking at finer scale spatial management.

In terms of the assessment model, while accepting that ratio BRPs are robust and can inform stock status determination, there appear to be many aspects in need of further exploration. I am unconvinced that all informative data are yet included in the model (e.g. LPUE) or that appropriate weighting of data sets has been pinned down. More importantly, I think further exploration of M and of balancing survey selectivity and growth estimation is necessary. This is all possible but my sense of the model is that the data are generally uninformative in some key respects and that rather than estimating q using the prior while holding survey selectivity and many growth parameters constant, it might be more profitable to approach the problem by looking in detail at the likelihood components at fixed q's while exploring selectivity and growth under alternative M assumptions, or even of using a q fixed at the median of the estimated prior and performing restricted sensitivity tests.

Recommendations

Under Process

1) I **recommend** that SARC considers the use of ftp or cloud-based alternatives. This would considerably help document control before, during and after review meetings.
2) I **recommend** SARC processes be evaluated in relation to equivalent processes elsewhere in the US and in the regional management context. My overall impression of the SAW/SARC 56 process is that there is much that could be done to improve it. I recognize this recommendation may be redundant as I am unsure of discussions on review processes at a joint SSC meeting in 2010 or 2011. Other recommendations here are relevant to this general one.
3) I **recommend** to the NEFSC/SARC organizers, that the standard and substance of SAW reports needs to be focused and strengthened, especially for new assessments; that if reviews are to be of value, the total time, scheduling, and materials available need to be sufficient; and finally, some thought needs to be given to ensuring that presentations are smarter and provide opportunity for in-depth discussion and exploration during meetings.

Under ToR 1

4) I **recommend** that the SAW provide a fuller description of conversion factors that lead to estimates of landed weight, and better characterize uncertainty in annual landings. Similarly, a better justification and possible reevaluation of the 12% incidental mortality is necessary. If conversions factors are not stable then this will need to be reflected in the assessment which currently treats catches (landings plus other components) as exact. Potentially, if changes have been systematic, the total landings stream may need to be revised.

Under ToR 2

5) I **recommend**, if the new model is to be used in future, that the SAW review fully age and length sampling and ageing and better characterize uncertainties in age and length composition data presented to the model and as used externally for estimating e.g selectivity.

Under ToR 4

6) I would strongly **encourage** plans to use management strategy evaluation looking at finer scale spatial management.

Under ToR 5

7) I **recommend** the SAW reconsider biomass reference points and also explores ways of making F reference points relate to equivalent biomass components, as well as examining the credibility of M assumptions. I note, however, the issue of what constitutes appropriate modeling approaches needs to be considered as does the issue of appropriate M assumptions; these are considered in the conclusion.

White hake

The main body of the report shall consist of a Background, Description of the Individual Reviewer's Role in the Review Activities, ***Findings*** *of whether they* ***accept or reject*** *the work that they reviewed, and an explanation of their decisions (strengths, weaknesses of the analyses, etc.)* ***for each ToR****, and Conclusions and Recommendations in accordance with the ToRs.*

1. Estimate catch from all sources including landings and discards. Describe the spatial and temporal distribution of fishing effort. Characterize the uncertainty in these sources of data. Analyze and correct for any species mis-identification in these data. Comment on the consistency of the approach to identify the catch of white hake with respect to that used in the red hake assessment.

This ToR has five parts.

i) Estimate catch from all sources including landings and discards.

The SAW considered historical catches, conversion factors (as fish are landed headed and gutted), catch composition (red and white hake aggregate category), recreational catches, and discards. The report commentary is not extensive on these various issues. Historical catches have been changed since the previous assessment by inclusion of new data from States and by revised/corrected application of conversion factors. The details of these are lacking and the changes by year are not explicit. It is also unclear if a common conversion factor has been applied to all data and whether this is appropriate as landings by State and season have changed through time. There is clearly the possibility of substantial error in estimating catches given the application of conversion factors. It would be interesting to see more detail on this and evaluation of uncertainty in catch streams (and compositions). The issue of dealing with red and white hake identification was addressed by the SAW. Previously the assessment of white hake used numbers estimated using ratios of the species by area in the NMFS surveys. Consistent with practice adopted in the red hake assessment (SAW/SARC 51), and because of poor species spatial representation in the survey, the SAW/SARC 56 used nominal landings in this assessment. I have no opinion on the best way of dealing with this issue and am content to see consistent application. Recreational catches were examined and reported by the SAW but the numbers are low and highly variable and are not used in the assessment. Not using the recreational catches will add to a variable bias but this should be very small compared to other errors in catch streams and compositions. It is appropriate to ignore recreational catches. The SAW estimated discards for all fleets and total, by year, using standard methods used in the region. The estimates have high CVs and have generally declined through time. The percentage of discarding is not high but the declining percentage makes it worthwhile including the estimates in the catch stream. [NB. Some text on discard CVs seems to have been appended accidentally to the final paragraph on catch compositions on p12 of the SAW/SARC 56 report.]

The SAW has simply described the catch component data used and has fully tabulated and graphed numbers. It has shown graphically (Fig. B49) how the total catch stream differs to that used previously; for this assessment total catch is generally lower in the series used since *circa* 1995 and higher prior. The SAW did not provide sufficient detail to enable consideration of

plausible alternative catch streams (and compositions). Given the numerous revisions, corrections and changes of approach, as well as the uncertainties associated with catch sampling, conversion factors, discard estimation, etc, a fuller exploration of possible variation in catch streams would have been useful for use in model sensitivity testing. I note that the only evaluation of the potential importance of the changes in catch streams was through application of the previously used ASPM model. The SAW apparently concluded from that work (p23) that the modifications to the annual catch data are important in determining changes from the previous assessment but this is unclear in the text and I cannot find relevant information in the SAW or presentations made during review.

ii) Describe the spatial and temporal distribution of fishing effort.

The SAW provided extensive plots showing distributions of landings and effort by gear type. The summary text is brief but carefully and usefully describes spatial and temporal changes in the fisheries. In my view the SAW fully addressed this ToR and the summary text presented in SAW/SARC 56, p 9, is sound.

iii) Characterize the uncertainty in these sources of data.

As noted above. The SAW has not well characterized uncertainty in basic catch information, either by placing overall confidence bounds on annual catches or by providing alternative catch streams for sensitivity testing. Given the number of assumptions in data building, and especially the number of changes from previous practice, this would have been a useful exercise to allow examination of possible alternative states of nature when determining stock status and making projections. This is not to say that alternative catch streams would lead to important sensitivities, just that an examination would have been worthwhile, especially at a time of new model development and when new BRPs are being proposed.

iv) Analyze and correct for any species mis-identification in these data.

v) Comment on the consistency of the approach to identify the catch of white hake with respect to that used in the red hake assessment.

These ToR components are covered at (i), above.

Overall, on ToR 1, the SAW has provided sufficient background on the catch estimation to enable understanding of the basic approach. The SAW has not provided sufficient background to allow a deep understanding of the complex issues surrounding the many issues involved in developing annual catches. A wider examination of potential uncertainties and errors would be useful.

2. Present the survey data being used in the assessment (e.g., regional indices of abundance, recruitment, state surveys, age-length data, etc.). Investigate the utility of commercial or recreational LPUE as a measure of relative abundance. Characterize the uncertainty and any bias in these sources of data.

This ToR requires only presentation of survey data used in the assessment but an investigation of the utility of commercial LPUE. I note these (presentation and evaluation) are quite different and that a more generally useful approach (e.g. as used during SEDAR processes) would be to present and evaluate all potential data inputs to the stock assessment model and to provide direct guidance on their possible utility, taking account of uncertainties.

With respect to LPUE, the SAW has briefly described a *pro forma* GLM approach to LPUE modeling, split by gear type (trawl and gillnet) and for the main gear, trawl, using either all data or only data in which a given percentage of the catch is white hake. The SAW report included standardized outputs (I think from SAS) as well as some simple figures showing nominal and standardized LPUE. The review presentation repeated some of these figures. Without seeing much fuller analyses and useful diagnostics of fit and influence it is hard to judge whether the fits to LPUE are acceptable or not and whether there are potentially useful alternative analyses that might be undertaken. The variables explored are year, quarter, tonnage class and area (five main subareas on GoM or GBK, all others not included according to the outputs in referenced tables) but no discussion is provided on the relative influence of these variables or whether there are other possible ways of exploring the data. The LPUE results were simply shown and the SAW did not report any analyses to investigate the utility of LPUE as inputs to the assessment model (as required by the ToR). In this regard the SAW failed to address the ToR. I am not sure if evaluations have previously taken place which might have relevance or if the approach reflects a general reluctance to consider fishery dependent LPUE given that long-term fishery-independent surveys are available. However, some comparison of the survey results in relevant areas (511-514 and 525) might be appropriate. Presenting the model with LPUE to see "ghost fits" might be instructive as to potential conflicts or agreements with survey indices and as a means to exploring utility. At a minimum, the SAW should explain why LPUE were not used in any model fits; it did not do so. I note that the SAW may have considered LPUE and other indices in exploratory model fitting but this is not reported anywhere and is unclear.

With respect to survey data, the SAW provided an extensive description of NMFS and State surveys. I have no problems with these descriptions but again note that while described, the three State surveys are not used in the modeling and no explanations are provided for this. It is unclear why the ToR did not require the SAW to investigate the utility of State surveys as measures of relative abundance.

I **recommend** that in SAW reports descriptions of all data should be provided, together with descriptions of past and present considerations and conclusions about data utility within the modeling process. If data are excluded from modeling *a priori*, explanations should be given. If data are explored in the modeling process but ultimately down-weighted or excluded, explanations should similarly be given.

3. Evaluate the utility of pooled age-length keys for development of a stock assessment model.

At the previous assessment (GARM III) the ASPM (now called SCAA) model was used for final stock status definition due to problems in configuring the ASAP model which was then under development. In particular, the use of a common (or "pooled") ALK to provide age information in the early period (1963/4 to 1985) in order to help convergence and improve diagnostics of the ASAP assessment was not deemed satisfactory by the GARM III Review Panel. At SAW/SARC 56 the SAW therefore investigated the use of annual and pooled ALKs for either or both survey and commercial catch at age data. The SAW used two models for this, both a traditional VPA/ADAPT and the new ASAP model.

The SAW examined how the two models reacted to the use of annual only ALKs, pooled ALKs for surveys or commercial catch compositions while using annual ALKs for the other, and only using annual ALKs. The SAW concentrated on SSB, F and YC estimation and considered retrospective patterns and BRP definitions, and looked in some detail at model fits (though comparison e.g. of LLs is not straightforward). Overall, the SAW took a practical approach to determining the utility of pooling ALKS and its conclusion that use of pooling with ASAP (as used for the base case run) appears robust. The conclusion was based largely on pooling not degrading recruitment pattern estimation, a major concern at GARM III.

The SAW noted the need for further work in this area, including investigating alternative pooling years, simulation studies, and presenting the model with length composition data and a single growth curve. Another option might be to use the available data to estimate growth within the model.

Overall, the SAW appears to have spent considerable effort on this ToR in an attempt to address GARM III concerns and develop the ASAP model as the basis for stock assessment. From the initial testing it appears that the ASAP model configured using pooled ALKs is acceptable, at least in the sense that the derived quantities of interest to management are likely insensitive to whether pooling is used or not. It is less than satisfactory, however, to move to a new model when the underlying compositional data are not compelling. I would prefer to have seen a more thorough analysis of the ASAP model, looking also at weightings and detailed diagnostics of the composition data, exploration of alternatives under different assumptions, etc. In this respect, while accepting the use of pooling for the SARC 56 assessment, I would note (see ToR 4) that a more comprehensive model development phase (or better description thereof) would have been useful. My sense is that work on pooling could usefully have taken place within a more rounded development process rather than taking considerable time, apparently as a separate exercise.

4. Estimate annual fishing mortality, recruitment and stock biomass (both total and spawning stock) for the time series, and estimate their uncertainty. Include a historical retrospective analysis to allow a comparison with previous assessment results. Review the performance of historical projections with respect to stock size, recruitment, catch and fishing mortality.

The SAW reported only briefly on new ASAP model development and gave a brief description of final model configuration. The presentation provided for review did not provide additional information. That the SAW adopted the base run by consensus and that the base run is "run 60" suggests the SAW considered extensive model development runs and diagnostics. This was and remains unclear. It is comforting, however, that the ASAP model was not developed *ab initio* but was a further development from GARM III, that a major concern at GARM III (the use of pooled ALKs) was well explored, and that extensive work on the ASPM model used at GARM III was undertaken during SAW/SARC 56. Without that extensive work on ASPM and on pooled ALKs it would be difficult to have confidence in the new ASAP model base run.

The SAW report makes statements that raise questions. Why was a two block selectivity split made in 1997/98 with both full selectivity ages set at 6? Presumably because of gear change, but why is this not explained? If not gear or other management change, why? I note the ASPM model has three selectivity blocks. I am not overly concerned with these issues but raise them only to note that the SAW report describes but often does not explain.

The description of weighting is useful but not convincing. The report says that the method of Francis (2011) was used to adjust weights but because the ESS for the spring survey were "very small" the SAW used the average of the estimated ESS for both the fishery and survey catch-at-age. It is not clear why the SAW regarded the estimated ESS as very low – again, an explanation is required. It is also not clear why using an average ESS is appropriate. In general the quality of the two data sets can be widely different and the bits of effective information certainly are; I would not expect them to be equally weighted. The SAW also says that the base run ESS for early spring and autumn surveys "should be higher". With reference to the SAW figure B128 that is one interpretation. At least for the autumn survey, and perhaps for the spring survey, the visual inspection could also lead to the conclusion that the later ESS assumed for surveys are too high. Weighting is a difficult technical issue and it is not clear that the SAW has fully examined best weighting for the model. It may have – it just isn't clear. I would point out also that Francis does not just give technical, algorithmic advice on weighting, but also talks pragmatically of reasons why composition data in particular might contain less information than presumed/hoped, especially commercial compositional data for which autocorrelation can be a systemic problem and for which forced down-weighting is often appropriate. The SAW report at ToR 2 and 3 does not discuss fundamental issues of compositional data quality and potential uncertainties so it is hard in the modeling process to know what is best.

Apart from weighting, the only modeling issue reported by the SAW relates to choice of initial F in the start year (1963). The SAW seems to have dealt with this effectively by examining behavior of the model as initial F is profiled, and by comparison with ASPM. While the SAW may have dealt with this issue, it's inclusion as the only issue in the modeling section leaves

doubt as to how much other exploration took place. Apart from this discussion, the ToR 4 report leads straight to base model diagnostics which are described briefly with reference to plots, primarily of residuals. It would have been useful to see more tabular diagnostics, particularly looking at negative log-likelihoods by model component under different model configurations. It would also have been useful to see at least some exploration of model fitting to alternative M and to alternative catch streams. These were not developed under ToR 2 but some sense of uncertainty in catches and model fitting would have been useful as M and catch streams probably represent major uncertainties in the model. Certainly, it would have been useful to see some exploration of alternative selectivity formulations.

Regardless, the presented diagnostics do suggest an overall good model fit. The SAW reported that there are no strong residual patterns. I would like to have seen at least for the fishery composition data a higher resolution plot of residuals. More generally, I would note that the age composition residual plots are very hard to read. Where there are many years stacked on the ordinate, I would **recommend** instead plotting by cohort to facilitate reading. (It is difficult to discern cohorts on such a low degree diagonal.) Unsurprisingly given the limited data, survey indices are well fit. I note there is a consistent but small positive residual pattern from about 1986 to 2006 in the fall survey. An explanation for this would be nice. It is also notable that the estimated SSB pattern is very different to the LPUE shown earlier and that estimated F shows little resemblance to the standardized effort series; this could perhaps have been commented on at ToR 3.

The SAW reported only briefly that MCMC was used to obtain posterior distributions of relevant model outputs. No MCMC diagnostics were provided in the SAW report but they were requested and provided during review. The MCMC appeared to be well behaved with no obvious problems. The SAW used the MCMC to estimate uncertainty in SSB and F and to provide a distribution for starting projections. The SAW has thus addressed this aspect of the ToR. The uncertainty is underestimated as it does not integrate uncertainty where parameters were fixed, nor was any attempt made to characterize uncertainty by exploring alternative models with, e.g, reduced M or with the catch stream as assumed at GARM III. These would have been useful exercises and I suggest in general a wider consideration of uncertainty to feed in to BRP, status and projection ToR.

The SAW included historical retrospectives. These show that trends in SSB and F are robust while the scaling of SSB (and hence potential catches) is sensitive to the assumed catch stream.

The SAW has stated that the presented base run represents the best model with which to evaluate stock status and provide catch advice. Although I would like to have seen more in depth explanation of the model development I accept this opinion and can see no reason to reject it.

*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.

The existing definitions for stock status are based on F40% as an $F_{MSYproxy}$, using the old ASPM model with data as available at GARM III. The SAW reported those figures, which are $F_{MSYproxy}$ = 0.125 and $SSB_{MSYproxy}$ = 56,300 mt. The values of the BRPs, however, are dependent on the selectivity patterns assumed/estimated, and on the catch streams and catch composition data presented to the model. With the adoption of a new model (ASAP, new selectivity estimates, a revised catch stream, and updated catch compositions), the BRPs are expected to change in value even if the same basis (F40%) is retained.

In working with the new model, the SAW took the opportunity not just to update the F40% proxy BRPs but also considered a redefinition of the basis for BRPs. The SAW correctly observed that estimation of F*MSY* based on a stock-recruitment relationship was not possible due to difficulties in estimation of SSB in the early time-series, and lack of contrast in recruitment estimates in the later part of the series. To this might be added the lack of understanding about the lower than average recruitment over the past 15 years. The SAW appropriately, therefore, chose to reconsider BRPs from the FX% family, noting that the use of F40% as a common proxy was due primarily to the work of Clark. The SAW chose to use the approach of Clark but with white hake specific parameterization. I accept this generally as a good way to proceed.

The SAW selected a criterion of "no more than 5%" that the population would drop below 20%B_0 and used Clark's method to find the constant annual F values that would meet that criterion under three stock-recruitment assumptions. The first of these is the standard resampling method used in stock projections. The second and third assume Beverton-Holt stock-recruitment with steepness set as h=0.8 and 0.7. These are good choices for a gadoid. The SAW described these stock-recruitment choices as "plausible". I accept that these are all plausible values but note that does not indicate they are the only plausible values or that exploring other dimensions of uncertainty is not warranted. The SAW did not comment on this.

The SAW also did not explain clearly what it intended by selection of "no more than a 5% probability". This is critical in what follows. Clark simulated many scenarios and based his general recommendations not on whether all scenarios met a given absolute risk criterion but on roughly average performance, including a large degree of human integration of results. As Clark considered a generic problem, this was sensible. Since Clark, many jurisdictions and RFMOs have considered BRPs and control rules and many have adopted absolute risk thresholds to be applied case specifically. The SAW needed to be explicit as to whether its selected 5% probability reflected Clark's humanly integrated approach or that of avoiding absolute thresholds, which is now more common. The words "no more than" could be interpreted to mean the latter. The use of Clark's approach could be interpreted as implying the former.

The SAW provided F values associated with the 5% probabilities of falling below 20%B0 for the three stock-recruitment scenarios. At h=0.7, the value is F=0.22. The SAW then provided F35%

and F40% values as 0.2 and 0.24 respectively. The SAW commented simply that the risk levels of these two reference points do not differ greatly and recommended the use of F35%. The SAW did not present the actual probabilities associated with F35% and F40% under the three stock-recruitment scenarios but it is obvious that the probability associated with F35% is greater than 5% under the h=0.7 scenario (because 0.24 is greater than 0.22).

In my view the SAW was remiss in i) not making explicit how it would use the selected 5% probability criterion (as an absolute threshold on any plausible scenario, or as a rough average across scenarios [implying model averaging influenced by the choice of range]), and ii) not providing the actual probabilities for the three scenarios when applying both F40% and F35%. Despite careful wordsmithing, not including this information has made interpretation and subsequent discussion very difficult.

During the SARC review (Wednesday 20th Feb), the Panel initially was willing to accept F35% as an F*MSYproxy*. However, the Panel was concerned that attempts to find out the actual probabilities associated with F35% and F40% were not well responded to. The presenters said more than once that the values were "low" and "similar" but gave no numbers. Given it was obvious that at F35% and h=0.7 the probability of falling below 20%B_0 was above 5%, the Panel requested that numbers be provided. The Panel, therefore, did not sign off on the issue until numbers could be presented and considered. At this stage, no discussion had been had on the interpretation being put on the 5% probability criterion.

On Thursday afternoon (21st Feb) numbers were provided to the Panel. For most stock-recruitment scenarios, for both F35% and F40%, the reported probabilities were 4% or less. However, for F35% with h=0.7, the probability was reported as 11%. At this stage the issue of whether the selected 5% threshold was intended as absolute or not became an issue. Panel members had various views on the 11% value with some uncomfortable about it exceeding the 5% threshold but also exceeding a more commonly applied 10% absolute threshold. However, the Panel moved towards favoring the F40% BRP option not because of the absolute value but because of the shape of the risk/response surface and uncertainty as to how the surface would look on a wider range of scenarios or with sensitivities explored (e.g. to M and assumed catch stream). A general disquiet was created by the lack of exploration of the new assessment (see ToR 4). In the absence of other information I accept this position at the time by the Panel as considered and reasonable.

Unfortunately, the issue was complicated further when overnight an e-mail was received from an industry review meeting participant, including text from a SAW member. That e-mail explained the intention of the SAW to use the method of Clark and treat the 5% probability as an average or "central tendency". On Friday 22nd February, therefore, the Panel reconsidered information previously presented in light of this explanation. The panel members did not clearly agree or disagree with the explanation provided but in any case reiterated their general concerns about the response surface and need for fuller exploration. I accept the explanation of working towards central tendencies but note it raises issues about the process of defining scenarios. The issue is

less marked, however, than defining plausible scenarios for testing against absolute thresholds. As such, I am sympathetic to the approach of Clark and of the SAW as expressed in the overnight e-mail. I would note, however, that the e-mail did not come through the SAW Chair and may or may not reflect a consensus SAW understanding.

If the foregoing were not difficult enough, late in the morning of Friday 22nd Feb, the SARC Chairman and NEFSC Population Dynamics Group Head informed the Panel that the numbers presented the previous day may have been erroneous and needed to be checked. The Panel was left with no option but to await revised and checked numbers and to work remotely after the review meeting. The Panel would have liked as part of any new work to see further exploration of model sensitivities but was informed this was not possible; a request to add a scenario with h=0.6 was therefore made to provide some insight in to the response surface. A short paper with revised numbers (attached here at Appendix 3) was sent to the Panel on 27th February. As seen previously, for F35% at h=0.7 the probability (9.7% as opposed to the previously reported 11%) of falling below 20%B0 increases steeply compared to the value at F40%. At h=0.6 the increase is very marked. The Panel concluded based on these numbers that it would be appropriate to continue to use F40% and that further exploration is needed. I note that h=0.6 is not in my opinion a plausible steepness value for white hake but was included only to show the non-linearity in the response surface. Some Panel members were concerned that the probability of 9.7% was in itself sufficient to reject F35% usage; I am not of that opinion. The Panel did not reject F35% and supported the general approach taken by the SAW. In my opinion proposing continued use of F40% while further (and better) exploring the use of F35% is appropriate.

In my view, the SAW report and presentations during review did not provide information necessary to make a clear, informed decision. The report was remiss in not explaining the intended use of the 5% probability threshold (as an average or central tendency) and not reporting actual probabilities associated with use of F35% and F40% under the selected "plausible" stock recruitment scenarios. The Panel was informed that the report was the result of careful wordsmithing; if this is so, I would comment that wordsmithing does not obviate the need for essential substance. The apparent initial resistance during review to provide numbers (and loose verbal comment of numbers being "similar" and "low") was frustrating and the inability to provide definitive numbers compounded that frustration. Receiving e-mails ex-process is concerning and it is in any case unclear if the SAW would generally concur with the explanations in those e-mails. Waiting almost one week after the meeting to receive revised numbers is not impressive. These are process issues that need to be considered carefully by SARC and hopefully improvements will be made. I am loathe to make a general recommendation to improve process; that is a complicated matter and would have little value without careful consideration and detailed recommendations. I do think it appropriate, however, for some evaluation of the processes (see section above on Process Issues).

There is a need regionally to be clear about default BRPs or processes by which BRPs can be changed. Science is fundamentally helpful in risk analysis, but risk analysis is just a part of risk management. Risk management needs also to set risk standards as inputs to risk analysis, and to

analyze risk analyses with a view to making informed risk treatments. The SAW/SARC 56 process appears to place the development of risk standards in a risk analysis context. As the standards are unclear it is not surprising that the results are also unclear. Attention at a regional level to the framework for setting and using BRPs would be useful. To the extent that it is a management issue, a CIE report cannot make recommendations. To the extent that it leads to poor science process, however, I consider it pertinent and therefore **recommend** that the NEFMC consider the issue of risk standards and development of guidance to SAW. I note in this respect that other FMC have clear policies and many assessment processes in other US regions have clear guidance.

At the risk of straying in to management, I would comment that the approach taken by the SAW (i.e. the use of Clark's approach) is sound and can be helpful in determining appropriate reference points. Clarks' approach was generic and use of a similar, stock specific methodology is sensible. However, the issue of how to consider average and absolute thresholds needs to be agreed and explicit and the issue of working with plausible scenarios that span carefully considered uncertainties needs careful consideration. There is some scope for prescription in dealing with these matters but in my opinion the key requirement is for good process involving reasonable people. The white hake SAW/SARC 56 seems to have made a good start but it needed to have explored more fully model sensitivities to be included in scenarios and to be more explicit about its risk criteria and approach.

The ToR for this review requires me to comment on whether I accept or reject the work reviewed. In this case it is a difficult task. The work reviewed cannot be divorced from the process. As is clear from the foregoing, I think the process was poor. I accept the method adopted by the SAW but not necessarily its conclusions. I do not, however, reject them outright. Given white hake stock characteristics, I would expect *a priori* that F35% or lower would be a suitable F*MSYproxy*. It depends, however, on details such as the estimated selectivity and on uncertainties that need to be explored. I consider it good that the SAW did not just work from *a priori* considerations but instead explored the matter. I think that exploration was not sufficient at this time to depart from the use of F40%.

6. Evaluate stock status with respect to the existing model (from previous peer reviewed accepted assessment) and with respect to a new model developed for this peer review. In both cases, evaluate whether the stock is rebuilt.

a. If possible update the ASPM with new data and evaluate stock status (overfished and overfishing) with respect to the relevant BRP estimates.

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

The SAW did not report fully on ToR 6(a). The report includes only brief text on work done using ASPM (the previous model used at GARM III) to investigate the impact of new data sources on estimated stock status. That work indicated that changes in estimated SSB and $F_{MSYproxy}$ are due to changed selectivity estimates as well as the revised catch streams and use of pooled ALKs. The latter point is in fact confusing in the text provided (p23) and could usefully

be improved. With respect to this ToR, however, the important ASPM results are not directly provided nor referenced, but are apparent only through Appendix B1 in which sensitivity analyses are provided together with an ASPM base case somewhat equivalent to the new ASAP base case assessment. The ASPM base case in fact includes an explicit stock-recruitment relationship and it is quite difficult, and care is needed, to extract the information from Appendix B1 (at Table 1.3) that includes estimates from ASPM of B/B_{MSY} and F/F_{MSY} where BRPs use the F40% proxies. These are given in the table as 0.71 and 0.75 respectively, suggesting that based on the ASPM analysis and using the GARM III proposed BRPs, white hake is neither overfished not experiencing overfishing in 2011.

The SAW base case assessment is outlined at ToR 4. SAW proposed BRPs are considered at ToR 5. The SAW proposed F35% BRP proxies. The SARC Panel had some concerns (see ToR 5, above) and recommended that retention of the F40% based proxies might be preferable at this time. Under either F35% or F40%, white hake would not be considered overfished or subject to overfishing in 2011.

While I have concerns about details of the assessment and choice of BRPs I accept the overall conclusions of the SAW and SARC Panel on stock status. The conclusion that white hake is neither overfished nor subject to overfishing in 2011 appears to be robust.

7. Develop approaches and apply them to conduct stock projections and to compute the statistical distribution (e.g., the probability density function) of the OFL (overfishing level) and candidate ABCs (Acceptable Biological Catch; see Appendix to the SAW TORs).

a. Provide numerical annual projections (3-5 years). 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.

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

The SAW report contains just four lines of text on short-term projections. The text says that projections were run but no detailed results were provided in text, tables or figures. The draft Summary Report provided for review did not contain any projections. The presentation provided during the review did contain four slides showing projections from 2012-2015 based on $F_{OFL}=F_{MSYproxy}$=F35% and at $0.75F_{OFL}$, both in tabular form and graphically. The limited results included medians, 5th and 95th percentiles of projected distributions of catch and SSB given the assumed constant Fs. The projections were based on starting distributions derived from the base run MCMC, an assumed catch in 2012 of 2,900 mt, and recruitment resampling from 1995 onwards. Although these tables and figures are not referenced in the SAW report or draft Summary Report provided for review, I presume they were available because the SAW report states that the results indicate the stock is rebuilt in 2012 and provides figures for OFLs and

TACs by year consistent with the projected median catches in tables provided in the presentation. I accept, therefore, that the SAW at least partially met this ToR and note the lack of reporting was likely due to the indisposition of the lead analyst.

In terms of the ToR, no attention was paid to items (b) and (c). My main concern with the projections is lack of attention at (a) to sensitivity testing or, in other words, to other states of nature. This concern relates to discussion at ToR 4. Also of concern is the lack of exploration of annual Fs other than F35%. I think the use of recruitment resampling from 1995 onwards is sensible given the apparent lower average recruitment since then. In general, I consider it good practice to use recent recruitment for resampling when apparent changes in recruitment level are estimated. I would note, however, that the resampling applies to 2010 onwards and that given selectivity the conclusions about rebuilding and near-term projected catch and SSB depend largely on annual recruitments estimated in the model. There are at least two issues to be considered here. First, the most recent estimated recruitments (e.g. 2008, 2009) used in projection are poorly estimated due to few data to inform estimation and are therefore effectively shrunk towards the long-term (not recent) mean recruitment. Second, and related, that retrospective patterns show a consistent overestimation of recruitment, no doubt related to the first point and arguably supporting the conclusion of a recent lower average level of recruitment. Generally, therefore, even though recruitment resampling is from 1995 onwards, the near-term projections are likely still to be optimistic. While use of recent recruitment is more "realistic" (ToR 7(b)), this is more in theory than practice and will influence projections only as fish derived from the resampling enter the fishery. As fish are only fully selected at age 6 (50% at age4) resampled recruitments will only start to affect the projection numbers 5 or 6 years after the last model estimate (i.e. 2014/15).

To be useful to the SSC and decision-making, the projections need to capture more of the uncertainty in the assessment. The MCMC-derived numbers-at-age in the starting year are just one component of this. Those numbers and other key model estimates, notably selectivity, are uncertain due to model uncertainty. An exploration of the model at other M, alternative selectivity assumptions, or to assumed catch streams would be an obvious starting place to explore states of nature. Some exploration of ASPM at higher or incremental M was undertaken and reported in Appendix B1 but this does not substitute for fuller analysis using ASAP; in any case, higher M is less interesting than lower M. The extended sensitivity analysis on ASPM does, however, give some comfort as to the likely resistance of conclusions reached using the new ASAP model.

At the end of the SARC 56 review meeting, there was an outstanding need for the SAW/NEFSC to run projections using F40% to define OFL. It was and is unclear at the time of writing if the SAW Summary report will include projections using both F35% and F40%, or just F40%. My opinion is that both should be presented. The projections should (as per ToR 7(a)) include the probability of falling below the overfishing threshold but it is unclear if this will be done or if only the 5th and 95th percentiles of projected SSB will be shown. Despite discussions during the review and comments above, I do not see how it will be possible to present alternative states of

nature unless and until the SAW more fully explores the assessment model. I am unsure if the projections to be presented to the SSC will meet formal requirements (as outlined in ToR 7) but in my opinion the projections could be made much more useful as input to a rational decision-making process.

In terms of ToR 7(c) the SAW did not comment. So long as selectivity is well-defined and if M is not assumed too high, then the current Fs are highly likely to be below any reasonable $F_{MSYproxy}$ (F35% or F40%). White hake live to 20 years or slightly more and standard considerations suggest an M in the region of 0.2, as assumed in the assessment. I would not expect M to be much higher across the age range, though it could be for lower ages, and therefore see no merit in sensitivity testing at much higher values (i.e. 0.4 as in Appendix B1); investigating age-structured M would be interesting but would confound with selectivity estimation and should not be undertaken lightly. Some exploration at lower M values might be warranted. Overall, I think the assumed productivity is reasonable. In terms of susceptibility, because white hake is largely taken in mixed trawl fisheries, it is possible that quota incompatibility, technological interactions, and potentially market drivers could create higher than intended fishing pressures on white hake. This would depend, of course, on quota monitoring and compliance and current quota levels for interacting stocks/fisheries. In the short term, given F is apparently well below $F_{MSYproxies}$, I do not see any major susceptibility issues.

*8. Evaluate the validity of the current **stock** definition, taking into account what is known about migration among stock areas. Make a recommendation about whether there is a need to modify the current stock definition for future stock assessments.*

The SAW considered a range of information sources on white hake stock structure. The Review meeting presentation provided considerable detail on the issue. Various studies suggest the existence of two, perhaps three, groups of white hake in US waters, with separate spawning locations and timing. Genetics work in Canadian waters suggests at least three distinct genetic populations, but mixing throughout their ranges. No information was available on the connection between US and Canadian stocks (if any).

Overall, it appears that white hake biological stock structure is complex with multiple spawning groups, likely spawning fidelity, but general mixing at least on fishing grounds. In US waters, it appears the large majority of white hake result from winter-spring spawning on deep waters of the continental slope from the northeast Gulf of St Lawrence to Southern New England. The ToR asked the SAW to take account of what is known about migration; the SAW did not comment explicitly on migration but did review information on distribution patterns of spawning, larvae and adult fish.

The SAW concluded (SAW/SARC 56, p7) that "*In light of the evidence…..all white hake found in US waters were treated as one stock."* As presented, this logic seems flawed as the evidence presented by the SAW strongly indicates more than one stock. The "treatment" as one stock for management and assessment, however, is a separate matter. Given i) the large majority of fish

are likely from one biological stock unit in US waters, ii) nothing is known on connections to stocks in Canadian waters, and iii) there is no reported method for disaggregating commercial or survey data on white hake, it seems a matter of pragmatism to continue to treat the stock as a single unit for assessment purposes. I would note that with mixing but detectably distinct genetic populations, strong spawning fidelity would seem to be implied and could have implications for spatial management of catches.

In my view the SAW addressed this ToR and the conclusion presented in the executive summary at SAW/SARC 56, p 5, is sound.

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

Previous research recommendations were listed and brief comments on each were provided by the SAW. Of the thirteen previous research recommendations made, the SAW effectively reported that they had either been addressed, were no longer relevant, or could be carried forward in to a new list. The list of new recommendations (including effective carry overs) has eight components. The eight listed recommendations were provided by the SAW without comment or justification. This is not uncommon practice amongst SAW in many regions but is unhelpful in priority setting or funding allocation decisions. If recommendations cannot be ranked by utility, feasibility, and cost it is hard to see value in making them.

Of the eight recommendations made by the SAW, two were for general methodology/software issues and are arguably covered by a current international assessment project using simulated data to test models. The next meeting for that project is due to be held at the NEFSC in summer 2013. Two of the recommendations are for routine work (completion of sample ageing and routine collection of samples); it is hard to see how these constitute research but both are essential to improve the white hake assessment and should be encouraged (within financial constraints and depending on other priorities at the NEFSC). One recommendation is to conduct sensitivity testing of the ASAP model using State survey indices. As noted at ToR 3, it is unclear why this was not done as part of the model development. I would suggest that it should be a standard part of the modeling exercise and am somewhat bemused to see it separately listed as a research recommendation when no discussion on the topic appeared in the SAW report. One recommendation is to explore swept area biomass estimation for white hake. I cannot support this recommendation given its likely complexity and expense and no obvious benefit given other data availability for the stock assessment. The final research recommendations from the SAW relate to estimating CFs from heads of white hake and cooperative research to collect intact fish form commercial gear. While it is in principle possible to develop CFs from heads to whole fish, and recognizing heads have otoliths that can be collected, I suspect the errors in the CFs would be so large as to make this unviable. It seems to me that the two recommendations might usefully be combined in to one that might read "consider how improved commercial age and length samples for white hake (and other species) can best be obtained."

Generally, the SAW-provided research recommendations do not naturally flow from the SAW report. Some do relate to issues raised during review discussion. I prefer to see clear rationale and justification for any research recommendations to allow, as noted above, an idea of utility, feasibility and cost to be derived.

White Hake Conclusions and Recommendations

The main body of the report shall consist of a Background, Description of the Individual Reviewer's Role in the Review Activities, Findings of whether they accept or reject the work that they reviewed, and an explanation of their decisions (strengths, weaknesses of the analyses, etc.) for each ToR, and ***Conclusions and Recommendations*** *in accordance with the ToRs.*

Conclusions

I accept the ASAP base model as a basis for providing advice on status and future catches. This is despite my criticism of the SAW report and the brevity of discussion therein. I do not find the modelling convincing as presented and think more work could usefully be done to improve understanding of uncertainty and to improve credibility. The continued use of the previous model (ASPM), and investigations with that model, has helped to provide confidence that the ASAP model can be used at this stage. Despite the need for more exploration, and issue of BRP development (below), the conclusions that the white hake stock is not overfished in 2011, nor experiencing overfishing, are robust.

BRP development has been problematic. From first principles I have no difficulty with use of F35% for white hake and with suitable analysis would not be surprised to see a lower FX% evaluated as acceptable. I do have difficulty, however, with the basis put forward by the SAW to change from F40% to F35% at this time and consider the primary failures to have been with the SAW report not explaining its interpretation of risk and not reporting key numbers. I understand the reasoning for suggesting a move to F35% - that it is equally risk averse (or prone) as F40% but provides utilization benefits. I agree with the logic in principle if the risks are neutral but am not confident this is the case. I think more work is needed to justify a change. Such work is feasible. I note the short-term utilization benefits of moving to F35% from F40% appear small compared both to the TAC increases that appear possible given projections under the two F options and to probable uncertainties in total catches (though these were not well explored by the SAW).

The BRP debate at SAW/SARC 56 does raise some important issues. It is not clear why there are no clear defaults for BRPs in the NEFMC region or why SAW are not issued with clear guidance on how to develop BRPs or depart from defaults. I think there is confusion as to the role of the SAW and managers (*sensu lato*) in the risk management process and this confusion needs to be clarified (see recommendations) in order that good science processes can operate.

Recommendations

Recommendations are highlighted in **bold, red** throughout the sections on specific ToR. These are repeated here for ease of reading. In total there are seven drawn from the report, of which two are effective repeats. Recommendation (7) is made as an additional recommendation (leading to a list of 8 below).

Under Process

1) I **recommend** that SARC considers the use of ftp or cloud-based alternatives. This would considerably help document control before, during and after review meetings.
2) I **recommend** SARC processes be evaluated in relation to equivalent processes elsewhere in the US and in the regional management context. My overall impression of the SAW/SARC 56 process is that there is much that could be done to improve it. I recognize this recommendation may be redundant as I am unsure of discussions on review processes at a joint SSC meeting in 2010 or 2011. Other recommendations here are relevant to this general one.
3) I **recommend** to the NEFSC/SARC organizers, that the standard and substance of SAW reports needs to be strengthened, especially for new assessments; that if reviews are to be of value, the total time, scheduling, and materials available need to be sufficient; and finally, some thought needs to be given to ensuring that presentations are smarter and provide opportunity for in-depth discussion and exploration during meetings.
4) I would **recommend** a fundamental review is needed by the relevant Councils, with input from SSC, with a view to providing guidance to SAW on reference point setting. This is effectively repeated at (8) below.

Under ToR 2

5) I **recommend** that in SAW reports descriptions of all data should be provided, together with descriptions of past and present considerations and conclusions about data utility within the modelling process.

Under ToR 4

6) I **recommend** that residual composition plotting be by cohort to facilitate reading.
7) I **recommend** that for base case runs and important sensitivity runs, clear tables are included in SAW reports that summarize model structure, and internally and externally estimated parameters (e.g. as at Tables A20 for surfclams).

Under ToR 5

8) I **recommend** that the NEFMC consider the issue of risk standards and development of guidance to SAW (see also 4, above).

APPENDIX 1

BIBLIOGRAPHY

Prior to the Workshop, extensive materials were provided *via* a weblink (http://www.nefsc.noaa.gov/SARC/SARC-56-pdfs/). Materials included the SAW reports and draft summaries, previous assessment reports and reviews, and limited background documents. No SAW working documents or review presentations were made available. During the workshop, an internal weblink was provided.

During the workshop a number presentations were given, and some additional materials were provided on request, including further background documents and presentations as well as responses to Panel requests. All files were made available through the weblink. The access was generally adequate.

Further references:

Francis, R.C.C.C. (2011) Data weighting in statistical fisheries stock assessment models. *Can. J. Fish. Aquat. Sci.* 68:1124-1138.

APPENDIX 2

Attachment A: Statement of Work for Dr. Kevin Stokes

56th Stock Assessment Workshop/Stock Assessment Review Committee (SAW/SARC): Benchmark stock assessments for Atlantic surfclam and White hake

Statement of Work (SOW) for CIE Panelists
(including a description of SARC Chairman's duties)

BACKGROUND

The National Marine Fisheries Service's (NMFS) Office of Science and Technology coordinates and manages a contract providing external expertise through the Center for Independent Experts (CIE) to conduct independent peer reviews of NMFS scientific projects. The Statement of Work (SoW) described herein was established by the NMFS Project Contact and Contracting Officer's Representative (COR), and reviewed by CIE for compliance with their policy for providing independent expertise that can provide impartial and independent peer review without conflicts of interest. CIE reviewers are independently selected by the CIE Steering Committee and CIE Coordination Team to conduct the independent peer review of NMFS science in compliance the predetermined Terms of Reference (ToRs) of the peer review. Each CIE reviewer is contracted to deliver an independent peer review report to be approved by the CIE Steering Committee and the report is to be formatted with content requirements as specified in **Annex 1**. This SoW describes the work tasks and deliverables of the CIE reviewer for conducting an independent peer review of the following NMFS project. Further information on the CIE process can be obtained from www.ciereviews.org.

SCOPE

Project Description: 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 is the cornerstone of the Northeast Stock Assessment Workshop (SAW) process, which includes assessment development (SAW Working Groups or ASMFC technical committees), assessment peer review, public presentations, and document publication. This review determines whether the scientific assessments are adequate to serve as a basis for developing fishery management advice. Results provide the scientific basis for fishery management in the northeast region.

The purpose of this panel review meeting will be to provide an external peer review of stock assessments for Atlantic surfclam (*Spisula solidissima*)and white hake (*Urophycis tenuis*). Atlantic surfclam is a marine bivalve found along the US east coast. White hake is a demersal

gadoid species found from Newfoundland to Southern New England, and common on muddy bottom throughout the Gulf of Maine. The last peer reviewed benchmark assessment of Atlantic surfclam was in 2009 as part of SARC 49. The last peer reviewed assessment of white hake took place in GARM III in 2008, followed by a more recent data update in early 2012.

OBJECTIVES

The SARC review panel will be composed of three appointed reviewers from the Center of Independent Experts (CIE), and an independent chair from the SSC of the New England or MidAtlantic Fishery Management Council. The SARC panel will write the SARC Summary Report and each CIE reviewer will write an individual independent review report.

Duties of reviewers are explained below in the "**Requirements for CIE Reviewers**", in the "**Charge to the SARC Panel**" and in the "**Statement of Tasks**". The stock assessment Terms of Reference (ToRs) are attached in **Annex 2**. The tentative agenda of the panel review meeting is attached in **Annex 3**. The SARC Summary Report format is described in **Annex 4.**

Requirements for the reviewers: Three reviewers shall conduct an impartial and independent peer review of the Atlantic surfclam and white hake stock assessments, and this review should be in accordance with this SoW and stock assessment ToRs herein. The reviewers shall have working knowledge and recent experience in the application of modern fishery stock assessment models. Expertise should include statistical catch-at-age, state-space and index methods. 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 that includes an appreciation for the varying quality and quantity of data available to support estimation of Biological Reference Points. For surfclams, familiarity with dynamics of sessile species and spatial management is desirable. For white hake, familiarity with gadid fish stocks would be desirable.

PERIOD OF PERFORMANCE

The period of performance begins on the award date, and the contractor shall complete the tasks and deliverables as specified in this statement of work. Each reviewer's duties shall not exceed a maximum of 16 days to complete all work tasks of the peer review described herein.

Not covered by the CIE, the SARC chair's duties should not exceed a maximum of 16 days (i.e., several days prior to the meeting for document review; the SARC meeting in Woods Hole; several days following the open meeting for SARC Summary Report preparation).

PLACE OF PERFORMANCE AND TRAVEL

Each reviewer shall conduct an independent peer review during the panel review meeting scheduled in Woods Hole, Massachusetts during February 19-22, 2013.

STATEMENT OF TASKS

Charge to SARC panel: During the SARC meeting, the panel is to determine and write down whether each stock assessment Term of Reference (ToR) of the SAW (see **Annex 2**) 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 Term of Reference 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 following tasks in accordance with the SoW and Schedule of Milestones and Deliverables herein.

Tasks prior to the meeting: The contractor shall independently select qualified reviewers that do not have conflicts of interest to conduct an independent scientific peer review in accordance with the tasks and ToRs within the SoW. Upon completion of the independent reviewer selection by the contractor's technical team, the contractor shall provide the reviewer information (full name, title, affiliation, country, address, email, and FAX number) to the COR, who will forward this information to the NMFS Project Contact no later than the date specified in the Schedule of Milestones and Deliverables. The contractor shall be responsible for providing the SoW and stock assessment ToRs to each reviewer. The NMFS Project Contact will be responsible for providing the reviewers with the background documents, reports, foreign national security clearance, and other information concerning pertinent meeting arrangements. The NMFS Project Contact will also be responsible for providing the Chair a copy of the SoW in advance of the panel review meeting. Any changes to the SoW or ToRs must be made through the COR prior to the commencement of the peer review.

<u>Foreign National Security Clearance</u>: The reviewers shall participate during a panel review meeting at a government facility, and the NMFS Project Contact will be responsible for obtaining the Foreign National Security Clearance approval for the reviewers who are non-US citizens. For this reason, the reviewers shall provide by FAX (not by email) the requested information (e.g., first and last name, contact information, gender, birth date, country of birth, country of citizenship, country of permanent residence, whether there is dual citizenship,

passport number, country of passport) to the NMFS Project Contact for the purpose of their security clearance, and this information shall be submitted at least 30 days before the peer review in accordance with the NOAA Deemed Export Technology Control Program NAO 207-12 regulations available at the Deemed Exports NAO website: http://deemedexports.noaa.gov/.

Pre-review Background Documents and Working Papers: Approximately two weeks before the peer review, the NMFS Project Contact will send (by electronic mail or make available at an FTP site) to the SARC chair and CIE reviewers the necessary background information and reports (i.e., working papers) for the peer review. In the case where the documents need to be mailed, the NMFS Project Contact will consult with the COR on where to send documents. The reviewers are responsible only for the pre-review documents that are delivered to the contractor in accordance to the SoW scheduled deadlines specified herein. The reviewers shall read all documents deemed as necessary in preparation for the peer review.

Tasks during the panel review meeting: Each reviewer shall conduct the independent peer review in accordance with the SoW and stock assessment ToRs, and shall not serve in any other role unless specified herein. **Modifications to the SoW and ToRs shall not be made during the peer review, and any SoW or ToRs modifications prior to the peer review shall be approved by the COR and contractor.** Each CIE reviewer shall actively participate in a professional and respectful manner as a member of the meeting review panel, and their peer review tasks shall be focused on the stock assessment ToRs as specified herein. The NMFS Project Contact is responsible for any facility arrangements (e.g., conference room for panel review meetings or teleconference arrangements). The NMFS Project Contact is responsible for ensuring that the Chair understands the contractual role of the CIE reviewers as specified herein. The CIE Lead Coordinator can contact the Project Contact to confirm any peer review arrangements, including the meeting facility arrangements.

(SARC chair)
Act as chairperson, where duties include control of the meeting, coordination of presentations and discussions, making sure all stock assessment Terms of Reference of the SAW are reviewed, control of document flow, and facilitation of discussion. For each assessment, review both the Assessment Report and the draft Assessment Summary Report. The draft Assessment Summary Report is reviewed to assure that it is consistent with the outcome of the peer review, particularly statements that address stock status and assessment uncertainty.

During the question and answer periods, provide appropriate feedback to the assessment scientists on the sufficiency of their analyses. It is permissible to discuss the stock assessment and to request additional information if it is needed to clarify or correct an existing analysis and if the information can be produced rather quickly.

(SARC CIE reviewers)

For each stock assessment, participate as a peer reviewer in panel discussions on assessment validity, results, recommendations, and conclusions. 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. From a reviewer's point of view, determine whether each stock assessment Term of Reference of the SAW was completed successfully. Terms of Reference that are completed successfully are likely to serve as a basis for providing scientific advice to management. If a reviewer considers any existing Biological Reference Point or BRP proxy to be inappropriate, the reviewer should try to recommend an alternative, should one exist. Review both the Assessment Report and the draft Assessment Summary Report. The draft Assessment Summary Report is reviewed to assure that it is consistent with the outcome of the peer review, particularly statements that address stock status and assessment uncertainty.

During the question and answer periods, provide appropriate feedback to the assessment scientists on the sufficiency of their analyses. It is permissible to request additional information if it is needed to clarify or correct an existing analysis and if the information can be produced rather quickly.

Tasks after the panel review meeting:

SARC CIE reviewers:
Each CIE reviewer shall prepare an Independent CIE Report (see **Annex 1**). 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 above in the "Charge to SARC panel" statement. If alternative assessment models and model assumptions were presented, evaluate their strengths and weaknesses and then recommend which, if any, scientific approach should be adopted.

If any existing Biological Reference Points (BRP) or their proxies are considered inappropriate, the Independent CIE 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 CIE Report produced by each reviewer.

The Independent CIE 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.

SARC chair:
The SARC chair shall prepare a document summarizing the background of the work to be conducted as part of the SARC process and summarizing whether the process was adequate to complete the stock assessment Terms of Reference of the SAW. If appropriate, the chair will include suggestions on how to improve the process. This document will constitute the introduction to the SARC Summary Report (see **Annex 4**).

SARC chair and CIE reviewers:
The SARC Chair, with the assistance from the CIE reviewers, will prepare the SARC Summary Report. Each CIE 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 (please see **Annex 4** for information on contents) should address whether each stock assessment Term of Reference of the SAW was completed successfully. For each Term of Reference, this report should state why that Term of Reference was or was not completed successfully. The Report should also include recommendations that might improve future assessments.

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.

The contents of the draft SARC Summary Report will be approved by the CIE reviewers by the end of the SARC Summary Report development process. The SARC chair will complete all final editorial and formatting changes prior to approval of the contents of the draft SARC Summary Report by the CIE reviewers. The SARC chair will then submit the approved SARC Summary Report to the NEFSC contact (i.e., SAW Chairman).

DELIVERY

Each reviewer shall complete an independent peer review report in accordance with the SoW. Each reviewer shall complete the independent peer review according to required format and content as described in **Annex 1**. Each reviewer shall complete the independent peer review addressing each stock assessment ToR listed in **Annex 2**.

Specific Tasks for CIE Reviewers: The following chronological list of tasks shall be completed by each CIE reviewer in a timely manner as specified in the **Schedule of Milestones and Deliverables**.

1) Conduct necessary pre-review preparations, including the review of background material and reports provided by the NMFS Project Contact in advance of the peer review.
2) Participate during the panel review meeting at the Woods Hole, Massachusetts during February 19-22, 2013.
3) Conduct an independent peer review in accordance with this SoW and the assessment ToRs (listed in **Annex 2**).
4) No later than March 8, 2013, each CIE reviewer shall submit an independent peer review report addressed to the "Center for Independent Experts," and sent to Mr. Manoj Shivlani, CIE Lead Coordinator, via email to shivlanim@bellsouth.net, and CIE Regional Coordinator, via email to Dr. David Die ddie@rsmas.miami.edu. Each CIE report shall be written using the format and content requirements specified in **Annex 1**, and address each assessment ToR in **Annex 2**.

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

January 15, 2013	Contractor sends reviewer contact information to the COR, who then sends this to the NMFS Project Contact
February 5, 2013	NMFS Project Contact will attempt to provide reviewers the pre-review documents
February 19-22, 2013	Each reviewer participates and conducts an independent peer review during the panel review meeting in Woods Hole, MA
February 22, 2013	SARC Chair and CIE reviewers work at drafting reports during meeting at Woods Hole, MA, USA
March 8, 2013	Reviewers submit draft independent peer review reports to the contractor's technical team for independent review
March 8, 2013	Draft of SARC Summary Report, reviewed by all CIE reviewers, due to the SARC Chair *

March 15, 2013	SARC Chair sends Final SARC Summary Report, approved by CIE reviewers, to NEFSC contact (i.e., SAW Chairman)
March 22, 2013	Contractor submits independent peer review reports to the COR who reviews for compliance with the contract requirements
March 29, 2013	The COR distributes the final reports to the NMFS Project Contact and regional Center Director

* The SARC Summary Report will not be submitted, reviewed, or approved by the CIE.

The SAW Chairman will assist the SARC chair prior to, during, and after the meeting in ensuring that documents are distributed in a timely fashion.

NEFSC staff and the SAW Chairman will make the final SARC Summary Report available to the public. Staff and the SAW Chairman will also be responsible for production and publication of the collective Working Group papers, which will serve as a SAW Assessment Report.

Modifications to the Statement of Work: Requests to modify this SoW must be approved by the Contracting Officer at least 15 working days prior to making any permanent substitutions. The Contracting Officer will notify the COR within 10 working days after receipt of all required information of the decision on substitutions. The COR can approve changes to the milestone dates, list of pre-review documents, and ToRs within the SoW as long as the role and ability of the reviewers to complete the deliverable in accordance with the SoW is not adversely impacted. The SoW and ToRs shall not be changed once the peer review has begun.

Acceptance of Deliverables: The deliverables shall be the final peer review report from each reviewer that satisfies the requirements and terms of reference of this SoW. The contract shall be successfully completed upon the acceptance of the contract deliverables by the COR based on three performance standards:

(1) each report shall be completed with the format and content in accordance with **Annex 1**,
(2) each report shall address each stock assessment ToR listed in **Annex 2**,
(3) each report shall be delivered in a timely manner as specified in the schedule of milestones and deliverables.

Upon the acceptance of each independent peer review report by the COR, the reports will be distributed to the NMFS Project Contact and pertinent NMFS science director, at which time the reports will be made publicly available through the government's website.
The contractor shall send the final reports in PDF format to the COR, designated to be William Michaels, via email William.Michaels@noaa.gov

Support Personnel:

William Michaels, Program Manager, COTR
NMFS Office of Science and Technology
1315 East West Hwy, SSMC3, F/ST4, Silver Spring, MD 20910
William.Michaels@noaa.gov Phone: 301-427-8155

Manoj Shivlani, CIE Lead Coordinator
Northern Taiga Ventures, Inc.
10600 SW 131st Court, Miami, FL 33186
shivlanim@bellsouth.net Phone: 305-383-4229

Roger W. Peretti, Executive Vice President
Northern Taiga Ventures, Inc. (NTVI)
22375 Broderick Drive, Suite 215, Sterling, VA 20166
RPerretti@ntvifederal.com Phone: 571-223-7717

Key Personnel:

Dr. James Weinberg, NEFSC SAW Chairman, NMFS Project Contact
Northeast Fisheries Science Center
166 Water Street, Woods Hole, MA 02543
James.Weinberg@noaa.gov (Phone: 508-495-2352) (FAX: 508-495-2230)

Dr. William Karp, NEFSC Science Director
National Marine Fisheries Service, NOAA
Northeast Fisheries Science Center
166 Water St., Woods Hole, MA 02543
william.karp@noaa.gov Phone: 508-495-2233

Annex 1: Format and Contents of Independent Peer Review Report

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 main body of the report shall consist of a Background, Description of the Individual Reviewer's Role in the Review Activities, Findings of whether they accept or reject the work that they reviewed, and an explanation of their decisions (strengths, weaknesses of the analyses, etc.) for each ToR, and Conclusions and Recommendations in accordance with the ToRs. For each assessment reviewed, the report should address whether each ToR of the SAW was completed successfully. For each ToR, the Independent Review Report should state why that ToR was or was not completed successfully. To make this determination, the SARC chair and reviewers should consider whether the work provides a scientifically credible basis for developing fishery management advice.

   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, and especially where there were divergent views.

   c. Reviewers should elaborate on any points raised in the SARC Summary Report that they feel might require further clarification.

   d. Reviewers shall provide a critique of the NMFS review process, including suggestions for improvements of both process and products.

   e. The independent report shall be a stand-alone document for others to understand the proceedings and findings of the meeting, regardless of whether or not others read the SARC Summary Report. The independent report shall be an independent peer review of each ToR, and shall not simply repeat the contents of the summary report.

3. The reviewer 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.

Annex 2: 56th SAW/SARC Stock Assessment Terms of Reference

A. Atlantic surfclam

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

2. Present the survey data being used in the assessment (e.g., regional indices of abundance, recruitment, state surveys, age-length data, relevant cooperative research, etc.). Investigate the utility of commercial LPUE as a measure of relative abundance. Characterize the uncertainty and any bias in these sources of data.

3. Evaluate the current **stock** definition in terms of spatial patterns in biological characteristics, population dynamics, fishery patterns, the new cooperative survey, utility of biological reference points, etc. If appropriate, recommend one or more alternative stock definitions, based on technical grounds. Integrate these results into TOR-4.

4. Estimate annual fishing mortality, recruitment and stock biomass (both total and spawning stock) for the time series (integrating results from TOR-3), and estimate their uncertainty. Include a historical retrospective analysis to allow a comparison with previous assessment results. Review the performance of historical projections with respect to stock size, recruitment, catch and fishing mortality.

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. This should be carried out using the existing stock definition and, if possible, for the recommended "alternative" stock definitions from TOR-3. If analytic model-based estimates are unavailable, consider recommending alternative measurable proxies for BRPs. Comment on the appropriateness of existing BRPs and the "new" (i.e., updated, redefined, or alternative) BRPs.

6. Evaluate stock status with respect to the existing assessment model and with respect to any new assessment model. Determine stock status based on the existing stock definition and, if appropriate and if time permits, for "alternative" stock definitions from TOR-3.
   a. When working with the existing model, update it 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).

7. Develop approaches and apply them to conduct stock projections and to compute the statistical distribution (e.g., probability density function) of the OFL (overfishing level) and candidate ABCs (Acceptable Biological Catch; see Appendix to the SAW TORs).
   d. Provide numerical annual projections (3-5 years). 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).
   e. Comment on which projections seem most realistic. Consider the major uncertainties in the assessment as well as sensitivity of the projections to various assumptions.
   f. Describe this stock's vulnerability (see "Appendix to the SAW TORs") to becoming overfished, and how this could affect the choice of ABC.

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

B. White hake

1. Estimate catch from all sources including landings and discards. Describe the spatial and temporal distribution of fishing effort. Characterize the uncertainty in these sources of data. Analyze and correct for any species mis-identification in these data. Comment on the consistency of the approach to identify the catch of white hake with respect to that used in the red hake assessment.

2. Present the survey data being used in the assessment (e.g., regional indices of abundance, recruitment, state surveys, age-length data, etc.). Investigate the utility of commercial or recreational LPUE as a measure of relative abundance. Characterize the uncertainty and any bias in these sources of data.

3. Evaluate the utility of pooled age-length keys for development of a stock assessment model.

4. Estimate annual fishing mortality, recruitment and stock biomass (both total and spawning stock) for the time series, and estimate their uncertainty. Include a historical retrospective analysis to allow a comparison with previous assessment results. Review the performance of historical projections with respect to stock size, recruitment, catch and fishing mortality.

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. Evaluate stock status with respect to the existing model (from previous peer reviewed accepted assessment) and with respect to a new model developed for this peer review. In both cases, evaluate whether the stock is rebuilt.
   a. If possible update the ASPM with new data and evaluate stock status (overfished and overfishing) with respect to the relevant BRP estimates.
   b. Then use the newly proposed model and evaluate stock status with respect to "new" BRPs and their estimates (from TOR-5).

7. Develop approaches and apply them to conduct stock projections and to compute the statistical distribution (e.g., the probability density function) of the OFL (overfishing level) and candidate ABCs (Acceptable Biological Catch; see Appendix to the SAW TORs).
   d. Provide numerical annual projections (3-5 years). 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).
   e. Comment on which projections seem most realistic. Consider the major uncertainties in the assessment as well as sensitivity of the projections to various assumptions.
   f. Describe this stock's vulnerability (see "Appendix to the SAW TORs") to becoming overfished, and how this could affect the choice of ABC.

8. Evaluate the validity of the current **stock** definition, taking into account what is known about migration among stock areas. Make a recommendation about whether there is a need to modify the current stock definition for future stock assessments.

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

Annex 2 (cont.):

Appendix to the Assessment TORs:

Explanation of "Acceptable Biological Catch" (DOC Natl. Standard Guidelines, Fed. Reg., vol. 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)

Explanation of "Vulnerability" (DOC Natl. Standard Guidelines, Fed. Reg., vol. 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 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)

Rules of Engagement among members of a SAW Assessment Working Group:

> Anyone participating in SAW assessment working group meetings that will be running or presenting results from an assessment model is expected to supply the source code, a compiled executable, an input file with the proposed configuration, and a detailed model description in advance of the model meeting. Source code for NOAA Toolbox programs is available on request. These measures allow transparency and a fair evaluation of differences that emerge between models.

Annex 3: DRAFT Meeting Agenda

[Note: The final SARC 56 agenda is still in preparation. The meeting will start at 10am on Feb. 19 and end late in the day on Friday, Feb. 22, 2013. Reviewers must attend the entire meeting. A draft agenda follows:]

56th Stock Assessment Workshop/Stock Assessment Review Committee (SAW/SARC): Benchmark stock assessments for Atlantic surfclam and white hake

Feb. 19-22, 2013

Stephen H. Clark Conference Room – Northeast Fisheries Science Center
Woods Hole, Massachusetts

DRAFT AGENDA* (version: 7 Jan. 2013)

TOPIC	PRESENTER(S)	SARC LEADER	RAPPORTEUR
Tuesday, Feb. 19			
10 – 10:30 AM			
Welcome	**James Weinberg**, SAW Chair		
Introduction	**Edward Houde**, SARC Chair		
Agenda			
Conduct of Meeting			
10:30 – 3:15	Assessment Presentation (A. Atlantic Surfclam)		
	TBD	TBD	TBD
3:15 –	SARC Discussion w/ Presenters (A. Atlantic Surfclam)		
	Edward Houde, SARC Chair		TBD
Wednesday, Feb. 20			
9 –	Assessment Presentation (B. White Hake)		
	TBD	TBD	TBD
1:30 –	SARC Discussion w/presenters (B. White Hake)		
	Edward Houde, SARC Chair		TBD
4	Revisit with presenters (A. Atlantic Surfclam)		
	Edward Houde, SARC Chair		TBD

6:45 PM **(**Social Gathering –**)**

Thursday, Feb. 21

8:30 –	Revisit with presenter (B. White hake) **Edward Houde**, SARC Chair	TBD
10:30	Review/edit Assessment Summary Report (B. White Hake) **Edward Houde**, SARC Chair	TBD
3:00	Review/edit Assessment Summary Report (A. Surfclam) **Edward Houde**, SARC Chair	TBD

Friday, Feb. 22

9:00 AM – 5:00 PM SARC Report writing. (closed meeting)

*All times are approximate, and may be changed at the discretion of the SARC chair. The meeting is open to the public, except where noted.

Annex 4: Contents of SARC Summary Report

1. The main body of the report shall consist of an introduction prepared by the SARC chair that will include the background, 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 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 CIE reviewers should consider whether the work provides a scientifically credible basis for developing fishery management advice. Scientific 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 the CIE 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 (BRP) or BRP proxies are considered inappropriate, include recommendations and justification for alternatives. If such alternatives cannot be identified, then indicate that the existing BRPs or BRP proxies are the best available at this time.

3. The report shall also include the bibliography of all materials provided during the SAW, and relevant papers cited in the SARC Summary Report, along with a copy of the CIE Statement of Work.

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

APPENDIX 3
PERTINENT INFORMATION FROM THE REVIEW

1) Participants List

Participant Last Name	Participant First Name	Role	Affiliation
Adams	Charles		NEFSC
Alspach	Tom		Sea Watch
Blaylock	Jessica		NEFSC
Brooks	Liz		NEFSC
Chute	Toni		NEFSC
Coakley	Jessica		MAFMC
Cryer	Martin	CIE	MPI, New Zealand
Curti	Kiersten		NEFSC
Dameron	Tom		Surfclam/Quahog Advisory
Deroba	Jon		NEFSC
Gabriel	Wendy		NEFSC
Gerencer	Bill		M.F. Foley Company, Inc.
Hart	Dvora		NEFSC
Hendrickson	Lisa		NEFSC
Hennen	Dan	Surfclam analyst	NEFSC
Hoff	Tom		Wallace & Assoc.
Hogan	Fiona		NEFMC
Houde	Ed	Chair	UMCES-CBL
Jacobson	Larry	Surfclam SAW Chair	NEFSC
Kretsch	Alexa		SMAST
Legault	Chris		NEFSC
McCay	Bonnie		Rutgers U
Miller	Alicia		NEFSC
Munroe	Daphne		Haskin Shellfish Lab, Rutgers U.
Nieland	Julie		NEFSC
Nitschke	Paul		NEFSC
O'Brien	Loretta		NEFSC
Odell	Jackie		NSC
Palmer	Mike		NEFSC
Potts	Doug		NEFSC
Powell	Eric		GCRL-USM

Rago	Paul		NEFSC
Robillard	Eric		NMFS/NERO
Serchuk	Fred		NEFSC
Shepherd	Gary	White hake SAW Chair	NEFSC
Smith	Michael	CIE	CEFAS
Sosebee	Kathy	White hake analyst	NEFSC
Stokes	Kevin	CIE	Stokes.net.nz, LTD
Terceiro	Mark		NEFSC
Traver	Michele		NEFSC
Wallace	Dave		Wallace & Assoc., Inc.
Weinberg	James	NEFSC SAW Chair	NEFSC
Wigley	Susan		NEFSC
Wood	Tony		NEFSC

2) Paper on white hake BRP receive 27th march 2013

Clarification of Risk Analyses for Selection of Biological Reference Points for White Hake
SARC 56
Prepared by Population Dynamics Branch
February 26, 2013

During the course of the White Hake Working Group model meeting, calculations were made to compare the risks associated with using F35% vs. F40% as a biological reference point. As described in the Working Group report, the criterion considered was the probability of spawning stock biomass at MSY decreasing below 20% of virgin spawning stock biomass (SSB0) under three different stock-recruitment assumptions: (1) steepness (h) =0.7, (2) steepness=0.8, and (3) an empirical cumulative distribution function that declines linearly to the origin if spawning biomass is projected to drop below the lowest estimated SSB value from the assessment model, aka the 'hockey stick' model. The Working Group reached consensus that an acceptable level of risk was that no more than 5% of the projected SSB distribution at the MSY proxy should fall below 0.2*SSB0. F values associated with a 5% probability were calculated for each recruitment model and are presented in Table 1. At h=0.7, F=0.22 has 5.3% probability, while at h=0.8, F=0.25 also has a 5.3% probability. Due to time constraints, the Working Group interpolated the F=0.24 (F35%) for either steepness to be just over 5%.

During the course of the SAW assessment review, the CIE reviewers requested that the lead analyst provide the actual probability at the F=0.24 (F35%) value to compare the equivalence between the proposed F35% and F40% currently used for management. In doing so, it was discovered that the probability under a steepness of 0.7 and F=0.24 was actually 9.7%[1], rather than the assumed 5%. Although the original information presented in the Working Group meeting was correct, there was not a linear relationship between steepness and risk, so the 10% risk was unexpected. The CIE reviewers also requested that a projection be conducted for h=0.6 to further explore the slope of this nonlinear relationship and to better understand the associated risks. Since the Working Group had established 5% as the threshold for risk in comparing the F35% and F40%, and the value for F=0.24 exceeded that level, the CIE determined that the two options were not equivalent in terms of risk as the Working Group proposed. Consequently, the SARC CIE reviewers decided the reasoning to move to F35% was not sufficient and recommended the biological reference point be maintained at the default value of F40%.

[1] Estimate of 10.2% presented during the SAW review resulted from the Working Group ASAP model (run 59). The final ASAP run (60) resulted in the estimate of 9.7%.

Table 1. Summary of biological reference points and risks of falling below 20% of SSB0 for alternative assumed values of steepness and an empirical stock recruitment function. Levels of F for each steepness necessary to increase the risk to 5% of falling below 0.2* SSB0 are presented.

steepness	SSB0	0.2*SSB0	SSBmsy	percent below 0.2*SSB0 F35%=0.24	F40%=0.20	F that results in ~5% draw:
0.6	139,200	27840	51300	26.1	7.2	0.19
0.7	128,100	25620	42960	9.7	2.0	0.22
0.8	119,200	23840	36940	4.1	0.7	0.24-0.25
emp.cdf	81,700	16340	28450 (F_{35}) or 32400 (F_{40})	0.0	0.0	0.35-0.36