53rd Northeast Regional Stock Assessment Review Committee

November 29 – December 2, 2011

Northeast Fisheries Science Center

Woods Hole, MA

SARC 53

Report from Center of Independent Experts (CIE) reviewer

Dr Ewen Bell

9 January 2012

Executive summary.

SARC 53 was convened 29th November 2011 - 2nd December 2011 to review the assessments of Black Sea Bass (*Centropristis striata*) and Gulf of Maine Cod (*Gadus morhua*).

Both proposed stock assessments were moving to the ASAP stock assessment software.

One common feature of the survey series for the two stocks is the introduction of the FSV Henry B. Bigelow following the decommissioning of the FSV Albatross. There were substantial differences in the gear deployed on the two vessels and therefore an assumption of linear catchability would be wholly inappropriate. The Bigelow series is currently too short to use as a stand-alone survey index and parallel catchability experiments were conducted in order to determine catchability coefficients. These coefficients indicate that the Bigelow is 2-6 times better at catching both cod and black sea bass (depending upon the size) although the confidence interval is quite wide. The terminal estimates of stock status will be sensitive to these conversion coefficients and a move to using the Bigelow survey series on its own is encouraged as soon as the time series is considered long enough.

Black Sea Bass

The proposed assessment of Black Sea Bass represented a move from a length-based model (SCALE) to an age based model (ASAP) following an extensive age determination exercise.

The assessment area is very large, running from Cape Hatteras in the south to the Gulf of Maine in the north. Genetic and tagging studies show a cline in population structure with separation in the summer months with an inshore distribution and a mixed phase where fish migrate to a locality in the south and east of the area during the winter. A single -area assessment model with annual time steps was proposed.

Basic screening of the age disaggregated data was not performed as part of the assessment process and subsequent analysis at the meeting demonstrated some concerns regarding the ability of the age-disaggregation to track cohorts through their life span. The various inshore summer surveys show large spatial variations in local recruitment strength and the potential for improved ability in tracking cohorts compared to the area-wide surveys. It is unclear if the ability to follow cohorts through is a result of the merging of sub-stocks into a single unit or low sample size in constructing the age-length keys.

Different model formulations produced significantly different estimates of terminal stock status almost covering the extremes of the historical observations.

The move towards an age-based assessment is welcomed; however in its current form the assessment is rejected as the review team were unanimous in their assertion that the model as presented did not provide a sound scientific basis on which to base management decisions.

As the proposed assessment model was rejected no new Biological Reference Points were estimated. Reverting back to the last accepted assessment model structure (SCALE) has the stock not being overfished and not in an overfished state. The SARC_53 review panel did not in any way review the SCALE assessment and consider that it may suffer from some of the same issues that caused the ASAP model to be rejected.

Gulf of Maine Cod

This assessment makes a drastic revision to the stock status compared with the previous reviewed assessment.

The assessment program has moved from ADAPT to ASAP; however the impact of these changes on the value of the stock assessment is minimal and the move was made for scientific (statistical) reasons. Two factors combined to change the stock status estimate, the weights at age and the size of the 2005 cohort.

The assessment team worked through a number of changes to the assessment method and the data streams in a logical fashion, giving confidence to the end result.

The proposed assessment model was considered a sound scientific basis for management action. Different model formulations point to the stock being at a low level with rising fishing mortality in recent years and therefore the estimate of stock status is considered robust to assessment uncertainty.

The changes in spatial distribution of the stock as indicated by the survey indices and the extent of the fishery should be treated as important and potentially very serious. The contraction of the stock is entirely consistent with the maintenance of high commercial CPUE whilst the assessment indicates falling stocks and increasing fishing mortality. Such phenomena have been observed before and have preceded major stock collapse (Canadian cod stocks).

The proposal to move from a Fmsy proxy of F40% to F35% was rejected as a sound scientific basis for the change was not provided.

Following a small change in how recruitment trajectories are handled at extremely low stock size (linear reduction to the origin in mean recruitment below lowest observed stock size), the projections were accepted and it is noted that the stock is not expected to be able to rebuild by 2014 irrespective of any management action that may be enforced.

Background.

SARC_53 was convened to review the stock assessments performed for Black Sea Bass (BSB) and Gulf of Maine Cod (GOMC). Although these stocks are quite different in their spatial range and ecology, there are a number of common features (substantial recreational catch and proposed use of the ASAP stock assessment programme) that meant they formed a reasonably cohesive pair to review together. In preparation for this review both of these stocks had been the subject of workshops concerning data compilation and stock assessment to which the reviewers had not been party. Both stock assessments were moving away from previous assessment models (SCALE for BSB and VPA for GOMC)

Individual Reviewer's Role in the Review Activities.

I was one of three independent reviewers contracted by the Center for Independent Experts to form a review panel for SARC_53. The Chair acted as a fourth reviewer.

In preparation for the meeting I read through the documentation provided (background documentation and the write ups of the assessments). I also downloaded the ASAP stock assessment programme in order to re-run the assessments as given in the report in order to confirm the results as printed. As part of getting to understand the input data for the model I performed some basic data analyses and shared them with the review panel and subsequently the assessment team.

During the meeting I posed questions to the assessment panel and requested alternative model runs from them.

This report represents my personal view of the assessments and the review process in general.

Conduct of the meeting.

The meeting was held over 4 days (29th November - 2nd December) with the first 3 days held in open session. The morning of the last day was also held in open session to accommodate the fact that the GOMC review had run over time somewhat. The format of the meeting was that each assessment document was presented to the panel, working through the ToRs sequentially, and questions were taken from the panel and the floor as they went. During this process a number of additional analyses would be requested by the panel which the assessment team would prepare off-line and present on the 3rd day of the meeting.

The presentation on GOMC filled the first day and spilled over to occupy the morning of the second day. This had repercussions for the BSB assessment in that the review panel had considerably less time for presentation and discussion.

Additional analyses requested by the group were presented on the 1st and 2nd of December. Given the short period of time available to the BSB assessment team, not all of the additional analyses requested had been prepared.

Comment on the NMFS review process.

There are a number of parallels between the NMFS review process and the ICES Benchmark process, including the "open" nature of the meeting and having a number of experts from beyond the ICES domain to offer independent views on the system. Having sat on both sides of the fence I appreciate the level of work that is put in by not only the reviewers but also the assessment teams under scrutiny.

Given that members of the review panel will have a wealth of experience in the field of stock assessment it is almost inevitable that they will also be involved in the advisory process and are therefore well aware of the implications for management of accepting or rejecting a particular assessment in their own arena of work. The terms of reference supplied to the panel were to determine whether the proposed stock assessment represented a sound basis for scientific advice on which to base management, which left no scope for considering the subsequent consequences. This left the review panel in a slightly uncomfortable position, particularly for the BSB assessment, where rejection of the proposed assessment presumably meant falling back to the previous assessment, which the Panel suspected had several of the undesirable features of the proposed new assessment. While there is a reasonable argument to say that the previous model should be used as a fall-back position as it was deemed acceptable by a similar review panel 3 years ago, there could have been interim developments in the data streams that render the model fit unsuitable. The panel were not requested to review the fit of the previous model with the addition of the most recent data, and it could be that whilst the proposed model was considered unsound it might still represent a better option than use of the previous model.

The pre-meeting documentation was timely in its delivery but somewhat varied in the level of detail and comprehensiveness between the two stocks. Complete standardisation would be the death of scientific endeavour in the assessment, but a set of standard figures and tables and minimum expectation of the text would help readers interested in multiple stocks. A number of tables and figures within the reports were not referred to within the texts. This is presumably the result of legacy workings (i.e. we have always produced these tables!) but it leaves the reader wondering why such tables are still included if their contents are not worthy of comment. I realise the following suggestion would only serve to increase the internal workload of the assessment teams, but some form of internal peer-review might help to tighten up the reports.

With reviewers coming from a range of backgrounds, data will undoubtedly be displayed and presented in unfamiliar formats and the ability for a reviewer to pick up data sets and quickly re-format them would be highly beneficial. In some cases this was problematic because within the written reports data (input and results) were either displayed as figures or in tables but rarely both. Naturally this represents a degree of duplication and redundancy that is not expected in peer-reviewed publications, but makes the job of a reviewer considerably easier than if they are required to set-up and run a model in order to (for example) check the absolute value of residuals. In a similar vein I feel it important that a stock assessment document should provide all the data required to enable a reviewer to replicate an assessment and check the reported outputs. In the case of ASAP I suggest this should be a print out of the input (DAT) file and the report (LOG) file as a minimum.

The presentations during the meeting were lengthy (particularly for GOMC) and followed the documentation almost exactly. This therefore represented quite a lot of duplication for the reviewers (who had already read the report and supporting documentation!) and consequently limited the amount of time available for addressing areas the reviewers had questions or concerns over. One suggestion for future reviews is that members of the review panel could highlight areas where greater detail is required and/or points for discussion so that the physical meeting can concentrate on these areas and uncontroversial areas need not be covered. This might also give the assessment team greater time to address the different requests for additional analyses from the panel to ensure that they are all completed.

Black Sea Bass

1) ***Estimate catch from all sources including landings and discards. Characterize the uncertainty in these sources of data. Evaluate available information on discard mortality and, if appropriate, update mortality rates applied to discard components of the catch. Describe the spatial and temporal distribution of fishing effort.***

Completion of this ToR was generally satisfactory.

Whilst it would appear that there are no major portions of catch unaccounted for, the level of detail provided around the estimates of catch from recreational catches was somewhat sparse. Catch data for the recreational fisheries came from the Marine Recreational Fisheries Statistics Survey (MRFSS) and, with the exception of smoothing the outliers in 1982 and 1986 appear to have been taken pretty much at face value. There are precision levels given in table b9 and b19, but there is no discussion surrounding these. It is also unclear what the sampling rate for the recreational fishery is (samples per tonne were provided for the commercial fishery). An indication of the number of trips sampled and the number (and variance) of trips estimated by MRFSS would give the reader an indication of the size and complexity of the issues. Given the apparent sensitivity of these estimates of catch (i.e. anomalous ~10-fold increases in 1982 and 1986), further discussion and/or sensitivity analysis surrounding all these uncertainties would have been welcome.

In terms of describing fishing *effort* (rather than the temporal and spatial distributions of catches) we were provided with relatively little information. The general pattern of inshore fisheries in the summer and offshore (and more southerly) in the winter were well described, but data were not presented in a form whereby it would be possible to look for changes in the spatial pattern of fishing effort of the commercial fleet through time (i.e. annual maps). Likewise we were presented with recreational landings by State aggregated over a 10-year period, but were unable to discern if this pattern had changed through time.

2) ***Present the survey data being used in the assessment (e.g., 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.***

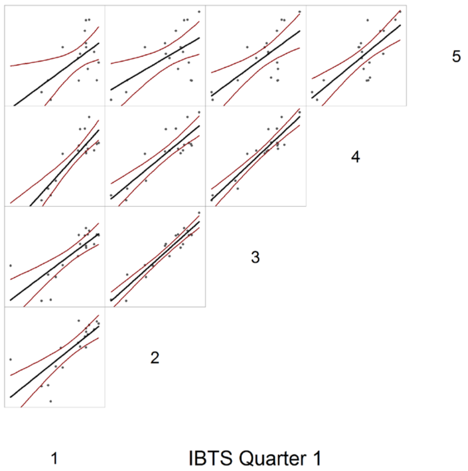
Completion of this ToR was satisfactory.

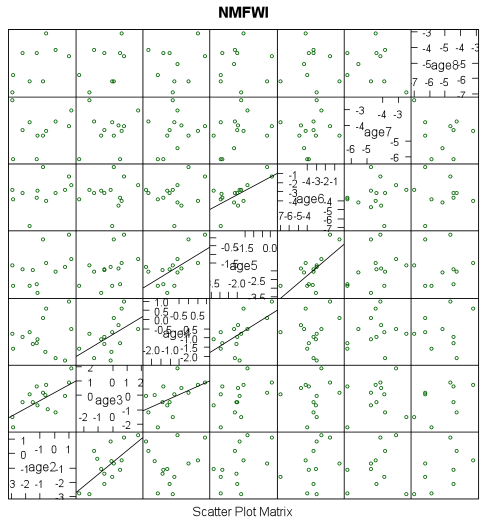
Whilst the construction of the survey and catch data was described in reasonable detail, there appeared to be a lack of screening and analysis of these important data. There is always a degree of extrapolation and interpretation when constructing age based data (i.e. transforming length distributions using age-length keys) and gaining an understanding of the uncertainties generated by this process is particularly crucial when moving to an age-based assessment.

The NOAA scientists were confident that their age-reading of scales was internally consistent, although no analysis of this was presented. The report comments on the number of lengths used to construct the age-length keys and notes that prior to 2008 there were "*an average of 107 and 124 ages in spring and fall, respectively*". Given that this covers a length range of more than 30cm in 1cm increments, the age-length key is quite sparsely defined with often only one or two fish at length towards the tails of the distribution. There has to be a fair degree of uncertainty regarding the age composition coming out of these age-length keys and there was no indication that uncertainty estimates had been made, or that implications for uncertainty in the age-length key on the stock assessment had been explored.

The main purpose of survey data used as an assessment tuning index is to provide an estimate of cohort strength, so the ability of a survey to track cohorts is paramount. Within the ICES community it is standard to plot the survey numbers at age against their numbers in subsequent years (i.e. age 1 year 1 against age 2 year 2, age 3 year 3 etc). From this it is easy to see if large cohorts are consistently picked up through time. As the cohort progresses variations in mortality weaken the relationship, but it would be expected that cohort strength would be significantly tracked for at least 2 years. For example, the North Sea International Bottom Trawl Survey (IBTS) Quarter 1 numbers of whiting display quite good internal consistency, and year-class strength as measured at age 1 is consistently picked up out to age 4. Strong relationships are observed at 1, 2 and 3 year lags. Compare this with the NMFS Winter survey numbers for Black Sea Bass where significant linear relationships are observed only on 1 year lags and out to age 6. There are many reasons as to why the relationship between subsequent observations of a year-class would be weak. Large variations in mortality rates will quickly mask fluctuations in recruitment, alternatively little contrast in recruitment strength will also reduce the strength of this kind of analysis. Errors in the age determination process, low sample numbers and/or merging of sub-population units are alternative explanations. Irrespective of the cause of the weakness in the relationship it means that the power of the survey to provide robust indicators of year-class strength is relatively low.

This would indicate that the survey is a poor indicator of year-class strength and that the age information at the oldest ages is largely uninformative.





Several of the State surveys were only considered in the final model as recruit indices yet the plots show that age distributions had been produced out to at least age 8. It would be of great interest to see if these surveys show any greater internal consistency than the area-wide NMFS surveys.

3) Consider known aspects of seasonal migration and availability of black sea bass, and investigate ways to incorporate these into the stock assessment. Based on the known aspects, evaluate whether more than one management unit should be used for black sea bass from Cape Hatteras north and, if so, propose unit delineations that could be considered by the Mid-Atlantic Fishery Management Council and for use in future stock assessments.

This term of reference was extensively covered.

The split into a northern and a southern component at Cape Hatteras seems logical and is well supported by tagging and genetic studies. The evidence for population structuring at smaller scales is complex. There appears to be increasing differences with distance of local scale (estuarine / inshore state) recruitment events and quite strong site fidelity (during the summer) of the adults, which points to the existence of sub-populations, at least at the generational scale. Genetic data indicate a smearing of this structuring into a North-South cline, but genetic differences require considerable periods of isolation to become significant. The picture is further complicated by the migration to offshore waters in the south of the region during the winter months where these sub-components are caught together. At present there appears to be no way of discerning the parental sub-component of fish caught in the mixed fishery and besides which the number of age and length samples are too low to support stock-splitting particularly in the historical period. Given the inability to split the mixed winter component, the slightly "fuzzy" structuring in the summer component and the sampling rate, there is a reasonable case to argue for a single area assessment.

However, this approach does have significant consequences. Single area models make the "dynamic pool" assumption, that is to say that all individual fish are equally available to the fishery at all times. The dynamic pool assumption is, of course, almost always violated in any stock assessment, although the degree to which this happens varies considerably. Having different levels of fishing effort and selectivity acting on different stock portions (with potentially different size/age structures) during the summer months, the violation in this case may be considerable. As will be described later, the apparent lack of age signals in the data may be a result of integrating catch data from different sub-stocks, and an age-structured assessment requires age signals within the data to provide a robust fit. There are also implications for management as the present assessment and advisory structure is unable to detect and/or highlight the potential for local depletion of stock units.

4) Investigate estimates of natural mortality rate, M, and if possible incorporate the results into TOR-5. Consider including sex- and age-specific rate estimates, if they can be supported by the data.

Completion of this ToR was satisfactory.

A range of plausible candidates for natural mortality estimates were presented, including age specific values. The exceptionally high value of ~1 from the instantaneous rates model does seem implausibly high and is justifiably ignored. The favoured option is to use time-invariant, age specific values coming from the Lorenzen model. Of the other options tabulated, no further analysis was presented on their implications for the fit of the stock assessment model. The text states that different scalings of the Lorenzen values were explored but it is not stated how these scalings were made. It appears that the scaling was performed over the 0-9 age range, irrespective of whether the assumption was for age 9 or age 12 as the maximum, and it is unclear if this was the authors' intention or not. This is largely academic because no further analyses were presented using these values.

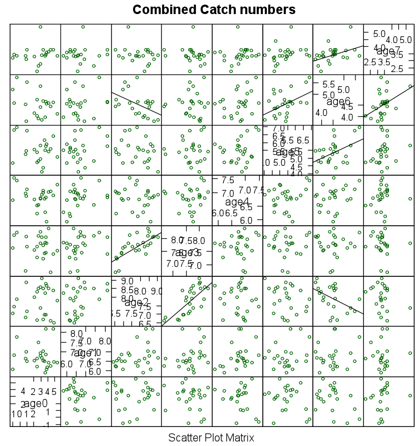
5) Estimate annual fishing mortality, recruitment and appropriate measures of stock biomass (both total and spawning stock) for the time series (integrating results from TOR-4), and estimate their uncertainty. Include a historical retrospective analysis to allow a comparison with most recent assessment results.

This term of reference was only partially completed, and ultimately the assessment was rejected unanimously by the panel. The assessors had clearly spent a fair bit of time on this ToR and are clearly trying to advance the assessment of this biologically complicated species, for which they are commended. There were, however, a number of issues that not been fully explored and the resulting assessment was not considered a sound scientific basis on which to base management.

In the assessment documentation provided two weeks prior to the meeting, I had spotted some implementation errors in the model specification. These errors were the use of -999 values (rather than 0) in survey indices for ages that had not been observed in any given year. The inclusion of negative values in a survey index causes the removal of the year (i.e. all ages) from the objective function. The detection of these errors meant that I was anticipating having to request a complete re-run of the assessment model during the meeting.

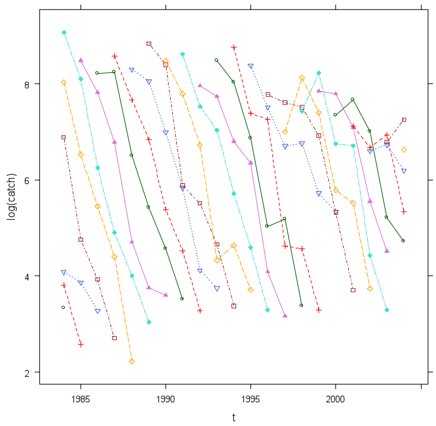
The stock assessment mentioned above was withdrawn the day before the meeting and superseded by a revised version which arrived 1 day into the meeting. Although regrettable, this revision was necessary due to the detection of an error in the way that some survey tuning indices had been constructed. For each tuning series, ASAP requires an aggregate index followed by the age-disaggregated components where available. These indices can be expressed as either biomass or numbers and whilst different indices can be expressed in the different units *within* a survey all units must be the same. In the documentation provided before the meeting some indices had combinations of biomass and numbers within the same index. This revision also allowed the stock assessors to correct the issues brought about by the negative survey values mentioned above, news of which had filtered back to them following discussions with Chris Legault (the author of the assessment software).

As with the age-based survey data, pre-assessment screening of the catch at age data would have been a very useful exercise to determine if there were sufficient informative signals for the catch at age model to utilise.



Again, the scatter plots show that there is little correlation within the catch matrix (i.e. year-classes did not appear to track through the catch data). This is not surprising for the younger year-classes as they are poorly selected for and there was a step-change in selectivity in 1998 following legislation of minimum landing size (MLS). However better tracking of the ages supposed to be fully selected (particularly the 3 and 4 year olds) would be expected.

Prior to fitting complex stock assessment models, visual inspection of the data can provide a useful insight into key features we would expect a model to detect, and, occasionally whether it is worth fitting a model at all. Within the ICES community it is common to plot catch curves in order to screen the catch data, which are simply the log catch numbers at age, tracked by cohort. The following figure shows the catch curves for black sea bass aged 1 and older. The hooking of the line at younger ages is caused by their reduced selection in the fishery. The fact that the trace is fairly linear for ages 3+ indicates full selection and the slope of the line can be interpreted as -1**z* (total mortality).



Several pieces of information can be gleaned from this plot. The change in MLS can clearly be seen as an increase in hooking of the line from 1998 onwards. The sharp steps in some of the lines show that either age-specific mortality rates can be highly variable or that considerable uncertainty exists in the catch at age matrix. The steps do not track through particular fishing years and are therefore unlikely to be changes in annual fishing pressure. The slope of the line is around -1.2 for most years (implying z of 1.2), with a significantly shallower slope (i.e. decrease in F) of around -1 for years 1998 through 2001.

At first it might seem contradictory that the scatter plot of catch numbers at age indicates little coherence whilst the catch curves show reasonable consistency; however, this is not necessarily the case. The relatively consistent slopes of the catch curves demonstrates that within the life of a cohort annual mortality is around 1.2, but there is a large amount of interannual variability caused either by large variations in mortality and/or error within the catch at age matrix. The increases in catch numbers at age observed in a couple of traces indicates that either error is not insignificant or that substantial immigration occurred.

Put together, the plots of internal survey consistency, catch consistency and catch curves all raise doubts concerning the appropriateness of using a single-area age-based assessment for this stock.

In determining the appropriateness (or otherwise) of various model runs, analysis of model residuals is key and plots of these have been provided for all the data sources used. In order to help the reader in interpreting these plots it would be better to only include those residuals which have contributed to the objective function (e.g. Figure B82 has residuals included for years 1968-2010, but there were only age data from 1984). Whilst the time series plots of residuals are necessary when checking for patterns in residuals it is as important to see plots of observed vs fitted values to ensure that the model is correctly specified and there are no internal biases.

With the GOMC stock assessment, a great deal of effort was spent building a "bridge" between the previous stock assessment and the new one, taking the panel through the changes in data raising, then the new model specification and finally adding in the most recent years of data. This process allowed the panel to see where the major differences and advances had been made. The black sea bass documentation did not follow such a measured approach, so the panel was unable to fully comprehend what were the benefits (or otherwise) of the new modelling approach. Admittedly with the GOMC we were moving from one age-based assessment model to another and with the bass we were moving from a length based model to an age based one, but greater efforts at bridge-building would have been beneficial. For instance, the length data could have been "sliced" into ages according to the growth curve used in the last SCALE model enabling a direct comparison of the two models with minimal data-change. The various parameter changes could then have been introduced step-wise to demonstrate their influence.

The assessment team were asked to investigate the effect of fitting the model only with survey data up to age 6 as this is the oldest age at which there is any ability for the surveys to track year-class strength. The rationale behind this is to see what effect the older ages, which have already been shown to be uninformative with respect to year-class strength, are having on the model fit. Unfortunately this was one of the requested additional analyses that was not produced. In the case of the GOMC, limiting the age range of survey ages to those ages that tracked year-class strength caused more severe doming of the fishery selectivity, which had a significant effect upon the assessment results. For black sea bass, where selectivity in the fishery is assumed to be asymptotic at the older ages, this might not be the case but obviously needs formally testing.

Another type of run that was requested (and run ad-hoc on screen during the meeting) was a series of model fits using just one tuning fleet at a time. These "single fleet" runs are commonly performed in ICES assessments and are used to investigate the consistency between surveys in terms of their influence on the assessment. The effect of this was interesting as the single fleet runs surveys all gave higher *f* estimates and lower *SSB* estimates than when the surveys were put together, *f* from the single fleet runs being **at least** twice that produced in the BASE run. The more normal pattern with single fleet runs is to observe a spread of F and SSB estimates and the combined run will fall somewhere central within the spread. The unusual pattern observed here has no easy explanation and warrants further investigation.

A wide range of model formulations were attempted by the assessment team in trying to find an acceptable solution, although the panel were not given a full list of the options investigated. The range of terminal estimates from these model runs essentially encompass the historical range (i.e. terminal SSBs covering the highest and almost the lowest ever seen). While some of these model runs will be more plausible than others, it does point to a great deal of structural uncertainty within the model.

The final formulation has a number of interesting features that warrant further comment. ASAP allows the operator to effectively manually weight the various sources of information by providing estimates of CVs. The assessment team used very low values of CV for the catch data (1% for 1981-2010). The various survey data series had CVs which were either fixed at approximate levels, or annually varying (and appearing to come from calculated values). One reason for using the ASAP model is to move away from the manual weighting of data sources and towards data-lead weighting though incorporation of CVs. The use of fixed, apparently arbitrary values for some series is therefore slightly unfortunate. It should be possible to determine CV estimates for all the different data sources and for the sake of consistency and transparency this should be pursued for future assessments. The particularly low CV value used for the catch data was, apparently, necessary in order to stabilise the model fit. Given that ~50% of the catches come from the recreational fishery which almost certainly has higher uncertainty than the commercial landings, this value seems impossibly low. The fact that this was required to stabilise the model fit is therefore of considerable concern and points to strong differences between the signals coming from the surveys and the catch.

The retrospective runs produced some curious results. In addition to the fairly strong patterns in terminal F and SSB (over-estimating SSB and under-estimating f), the various peels caused patterns in the historical portion of the estimates, pivoting around 1998 and popping out in the 1980s. The reasons for this had not been explored by the assessment team, but it would appear that the assumption of stationarity (within time blocks at least) is being violated to quite some degree and warrants further investigation.

It is never a good feeling to reject a stock assessment, particularly where considerable effort has been put into moving to a new assessment basis, but in this instance there are a number of unresolved issues that give sufficient concern to recommend rejection of this assessment. I am unconvinced that there is sufficient signal in the age structured data as prepared for this assessment with which the model can fit to. This may be a function of

- the relatively low numbers used to construct the historical ALKs,
- merging of sub-stock units
- genuine low contrast in the recruitment strength

6) State the existing stock status definitions for "overfished" and "overfishing". Then update or redefine biological reference points (BRPs; point estimates or proxies for B_{MSY}, $B_{THRESHOLD}$, F_{MSY}, and MSY) and provide estimates of their uncertainty. If analytic model-based estimates are unavailable, consider recommending alternative measurable proxies for BRPs. Comment on the appropriateness of existing BRPs and the "new" (i.e., updated, redefined, or alternative) BRPs.

Given that the assessment was rejected there was no basis to address this ToR.

7) Evaluate stock status with respect to the existing model (from the most recent accepted peer reviewed assessment) and with respect to a new model developed for this peer review.
   - ***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 (from black sea bass TOR 6).***

Given that the assessment was rejected there was a limited basis to address this ToR. With respect to the previous SCALE model and BRPs, the stock would not be classed as overfished nor would overfishing be occurring.

8) Develop and apply analytical approaches to conduct single and multi-year stock projections to compute the pdf (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, and definition of BRPs for black sea bass).***
   - ***b. Comment on which projections seem most realistic. Consider major uncertainties in the assessment as well as the 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.

Given that the assessment was rejected there was no basis to address this ToR.

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

Completion of this ToR was satisfactory. The assessment team worked through each of the previous recommendations for further work and stated what had been done on each item. Most of the previous recommendations have been addressed in one form or another. Of those recommendations that have not received any attention, it is noted that these involve large, costly experiments, often with associated sea time .

This ToR was not dwelt on during the meeting due to the constraints of time.

In terms of future work, construction of a spatially structured model with a finer time-scale (6 month steps?) would go a long way to addressing the structuring suspected to occur within this stock, although it is acknowledged that the data requirements for this (in terms of spatially structured Age-Length keys and catch sampling) are considerable.

Given the recent findings regarding the occurrence of sex-change in BSB, particularly that some fish appear to be born male rather than changing sex later in life, research into the implications for measures of spawning potential is recommended. SSB is used as a proxy for the ability of a stock to generate progeny, but with protogynous species, large numbers of the oldest (male) fish are not necessarily a good measure of spawning potential given that a single male may fertilise multiple females.

Gulf of Maine Cod.

1) ***Estimate catch from all sources including landings and discards. Characterize the uncertainty in these sources of data. Evaluate available information on discard mortality and, if appropriate, update mortality rates applied to discard components of the catch.***

Completion of this ToR was generally satisfactory.

It would appear that there are no major portions of catch unaccounted for. There have been a large number of changes to the process of raising the catch numbers since the last assessment and these were largely very well explained.

The change in raising method for discards was well explained although the reasoning for this was not well established. Whilst the adoption of a standard approach to raising discards across species has its merits (i.e. in mixed-fishery analysis), the text leaves the question of the appropriateness of this method for GOMC rather open. The cross-validation exercise using landings (figure A.29) was potentially useful but the figure description was so unclear I'm still not sure which line is which. Whilst the gross trends are similar (high in 1990, lower since 2000) there are very large (up to 100%) differences in the cross-validation exercise in 2007 and 2008. Should these magnitude differences occur in the discard estimates it may have an observable impact upon the assessment. The hind-casting of discard estimates is notoriously difficult and the assessment team gave a detailed explanation of the methods used here. The methodology used to hind-cast discards seems appropriate.

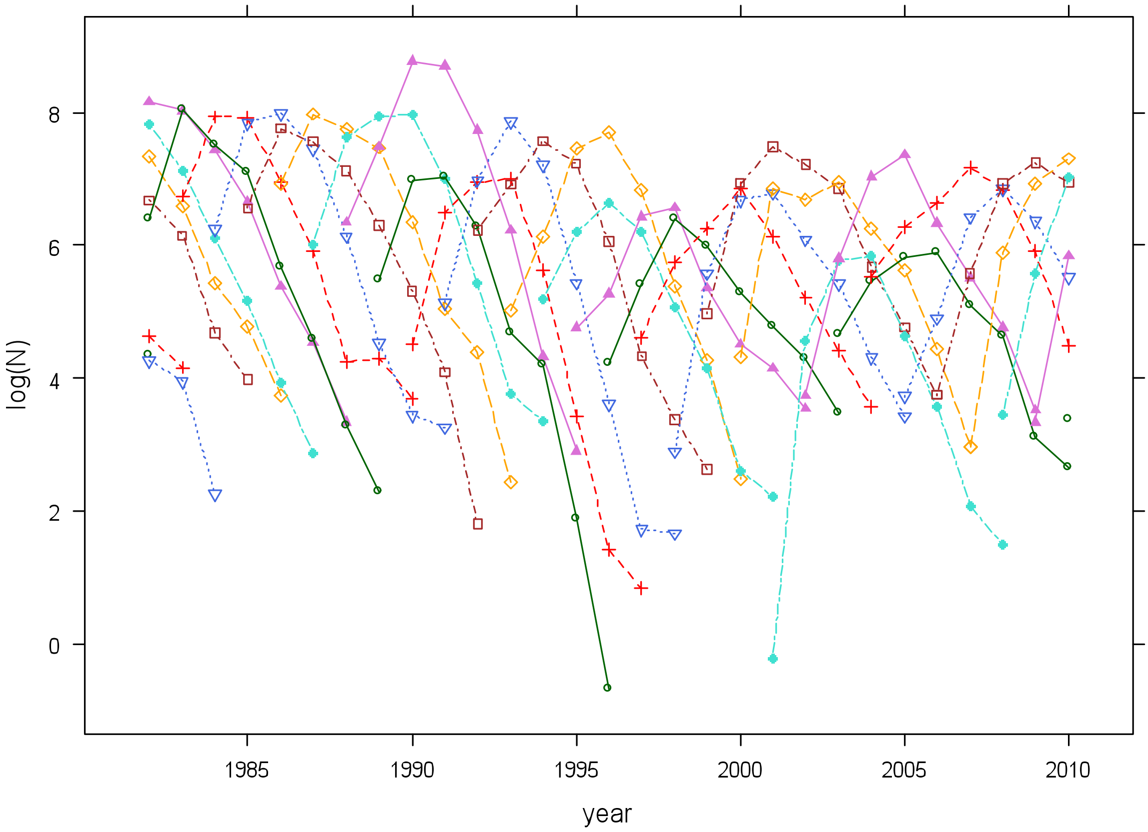
The description of the sampling of recreational catches was useful, but there was no real analysis or discussion regarding the uncertainty on recreational catches. The sampling rate in terms of numbers of lengths per 100t of landings is given, but unless there is 100% coverage in the reporting of recreational landings, the absolute tonnage must result from a model, which presumably has some associated measure of uncertainty. Given the importance of the recreational portion of the catch (up to 50%), the uncertainty around recreational catches may be important for the assessment model.

Although there was discussion of the relative contribution to catches of the different gear groups, there was no discussion on the trends in effort of these groups and whilst fishery-dependent CPUE/LPUE metrics are not used in the assessment, it would aid the reader to see some data regarding fishing effort, particularly when trying to square up assessment results to any changes in management regime. The progressive rise in the importance of the gill net fishery (at the expense of the trawl fishery) was noted.

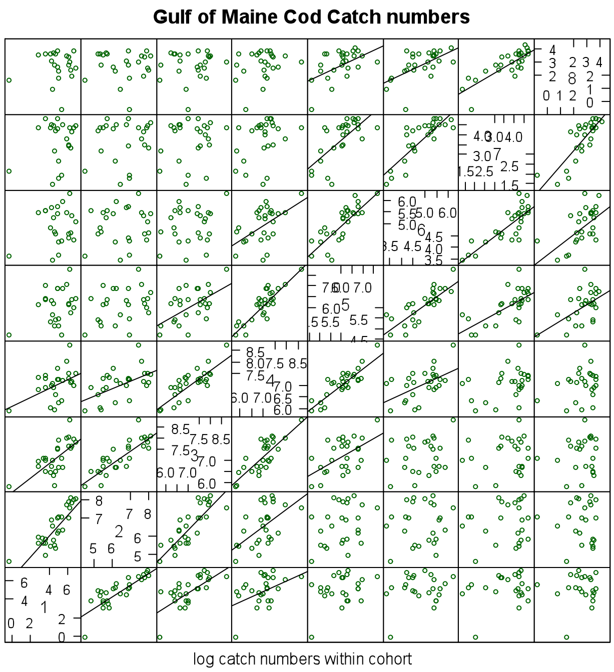
The issue of discard mortality is understandably complex and the assessment team acknowledge the problems in trying to assess post-discard mortality rates. The discard mortality of trawl-caught individuals is undoubtedly very high and while line caught individuals may suffer lower instantaneous mortality it is very difficult to estimate the longer term effect. The assumption of 100% discard mortality used in the assessment is likely to be an over-estimate although the magnitude of this is probably small.

There was no mention made of basic screening of the age disaggregated data. The majority of age-based stock assessments in the ICES arena will include a catch-curve analysis and plots of correlations in the catch at age matrix. Whilst these do not form part of the formal analytical assessment, they are highly informative in understanding the input data, spotting anomalies and identifying features that the assessment model should pick up.

The following figures show the catch-curves (log-transformed catch numbers at age, grouped by cohort) for the GOMC data, ages 1-8. The hooking of the line at younger ages is due to increasing selectivity, and then a reasonably linear phase follows as the stock is fully selected. The curves are reasonably smooth, indicating that there are no major issues with age determination. The fully selected slope appears to become shallower after the late 1990s compared with the earlier period, implying a reduction in fishing mortality, whilst the increased in hooking of the line implies that we should see a decrease in selection on the younger ages in the later period. There are changes in the slope at the older ages although the direction of change is variable, some cohorts appearing to increase their mortality rate, some decreasing substantially. Under the assumption of separability used in the ASAP model we would therefore expect to see a high degree of uncertainty in selection at the older ages, a factor likely to be influential on the estimate of SSB.



There is good coherence within the catch at age matrix and therefore there are no major concerns with regards to the process of age determination.



Changes to the Length-Weight relationship used in the assessment were well documented and their use is justified. The use of seasonal changes in condition is considerably more appropriate than assuming constant condition. Naturally there is some uncertainty regarding their applicability to the non-observed period (resulting from any changes in spawning season and/or seasonal food availability), but the assumption of stationarity in the previous annual weight-length relationship assumes no changes in the seasonality of the fishery (which is far more likely).

Changes were also made to the maturity ogives used in the assessment and, although their impact was not explored in a specific sensitivity run, given the minor adjustments made, their contribution to the overall change in stock status will be minimal.

2) Present the survey data being used in the assessment (e.g., 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.

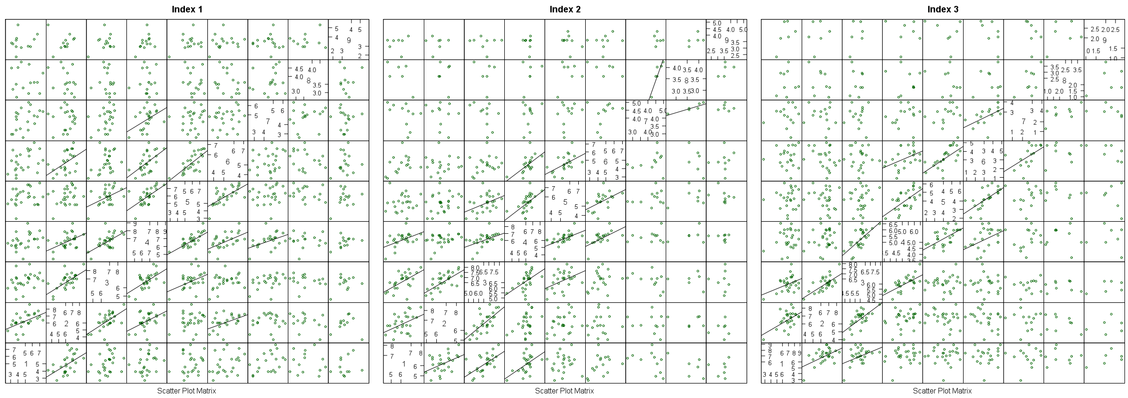
Completion of this ToR was satisfactory.

The single biggest change in the survey series is the retirement of the FSV Albatross and introduction of the FSV Henry B. Bigelow. As there are only three years of data from the Bigelow it has been necessary to apply conversion factors to the Bigelow data in order to create a continuous survey time series. The data collected during the comparative tow exercises were used to derive length-dependent conversion coefficients. It was noted that the statistical methodology used to create the conversion coefficients for GOMC (segmented regression) was more simplistic than used for the BSB coefficients (GAM). The explanation given for this was that the cod analysis was done first and the technique was further developed before the BSB analysis. It is possible that there may be some adjustment to the conversion coefficients should the new statistical method be used on GOMC, but these adjustments are likely to be small and the uncertainty bounds on these estimates are anyway quite large. As the survey series are converted to swept area estimates of abundance, the estimate of q coming from the model can be interpreted as an estimate of absolute catchability, which for the Albatross series was around 0.9. Given that the Bigelow is more efficient at catching fish, this puts the estimate for the Bigelow at around 2 and raises the possibility that the scaling of the model is significantly out (a catchability estimate of 1 means that 100% fish available to the gear are caught). Given that the catches are fixed, the only way to increase the number of fish within the model (and thus address the issue of $q>1$) would be to significantly increase the natural mortality estimates, although the data do not support the idea of substantially higher estimates. Converting scientific surveys to absolute indices of abundance are fraught with problems and this approach is very rarely taken within the ICES community. In addition to the assumptions already made regarding a linear relationship of survey abundance to stock abundance, using the index as absolute means that the ground surveyed is directly in proportion to the ground available to the fish and that abundance on a ground type is uniform. These assumptions are much harder to justify and therefore surveys are usually assumed to be relative measures of abundance. The $q>1$ estimate should not be overplayed and once the time series of the FSV Bigelow is of sufficient length, it should be used as an independent series without converting to Albatross units.

There was a mismatch in the survey data printed in the report and those in the model input. The values in the report have been transformed to a swept area estimate compared with the numbers provided in the model input, which are numbers per tow. Whilst this transformation has no influence on the model results, it makes the job of the reviewer more difficult in checking data consistency.

There were no analyses presented regarding the suitability of the surveys for use as tuning data. There now follows some simple data-screening techniques that could be used to inform assessment set-up before throwing the data into complex models.

Scatter-plots of logged survey indices in which cohorts are plotted at lags of increasing years show that the three age-based surveys have reasonable ability to track cohorts out to about age 6 but have no significant power beyond this.



Given the problems with the 2005 year-class the assessment team had gone back into the raw survey data to discover if there were any clues in the original data that might have given a warning signal that the very large year-class might have been overestimated. This showed that the signal was entirely driven by one station in the NEFSC-spring survey. Routine analysis of survey data, including mapping by age-class, is highly recommended as this would flag up such issues in advance of assessments being performed and give time to investigate their validity and impact. Mapping of the overall survey indices clearly demonstrated a reduction in the spatial range of GOMC and loss of the spawning component along the Maine coast.

Standardised survey numbers at age by year (i.e. divided by the mean and plotted on a log-scale) show an absence of strong year effects in the surveys. The same data plotted by year-class are analogous to the scatter plot above, but it is possible to see if the ability to follow cohorts changes with time, and there are indications that the NEFSC-spring survey is less consistent in recent years.

The revisions to the process of estimating weight at age are justified and well documented.

3) ***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, catch recruitment and fishing mortality.***

Completion of this ToR was comprehensive.

The assessment team did an excellent job of demonstrating how the various changes to input data and model structure affected the assessment. There have been substantial revisions to the stock trajectory since the last assessment, and the methodical approach taken by the assessment team gave the review panel a great deal of confidence in the reasons behind these changes.

Although there were variations in model performance and absolute estimates of terminal stock size between the various model formulations tested, all model runs using the latest data imply that the stock is towards the bottom of the observed historical range and that fishing mortality has increased in recent years.

The move from the ADAPT to ASAP was made for reasonable methodological reasons and the impact of the different model was minimal. One of the stated reasons for the move to ASAP was in order to allow for data-driven weighting of information streams (rather than manual weighting). It is therefore surprising that the CVs input for the catches were set at an apparently arbitrary at 5% rather than deriving them from the data. Having said that, the sensitivity analysis on difference values of catch CV requested by the panel showed that the 5% CV assumption had little impact upon the model fit and would have no impact upon the stock status in relation to BRPs.

The biggest change in stock perception comes from the revised estimate of the 2005 year class, which survey indices had indicated to be very large as age 1 and 2 but subsequently failed to materialise in the fishery and disappeared from the surveys at age 3 (as discussed under ToR 2).

The use of commercial CPUE indices is a general concern in many stock assessments, particularly where management measures are impacting upon fisher behaviour and/or where concentrations of the stock are concerned (hyperstability). The decision to drop the commercial tuning index from the assessment is therefore supported.

Model diagnostics were generally well documented, although there are a couple of notable exceptions. Plots of observed vs expected data are required to determine if there is any bias or model mis-specification. These were requested and subsequently provided during the meeting. One advantage of the ASAP model is the automatic generation of uncertainty levels for all estimated parameters, so it is surprising that these are not included when plotting the catch and survey selection patterns.

There is some evidence of a residual pattern in the catch at age since 2005 with positive residuals on ages 1-3 and negative on 4 and older although the size of these residuals is low compared with historical values. Future assessments should perhaps consider moving to three selectivity periods.

The progressive move towards the use of gill net gear (rather than trawl gear) means that a gradual change in selection pattern might be expected thus placing strain on the assumption of separability within the model. However, major changes in legislation governing mesh size and minimum landing size have evidently had a far greater effect than this more subtle change as evidenced by the requirement for different separable periods coinciding with the timing of legislation.

In order to further explore the influence each of the surveys has on the model fit, a sequence of model runs in which only one survey at a time is included ("Single-fleet" runs) were requested. This approach to model testing is standard in ICES assessments. The two NEFSC surveys (spring and fall) gave very similar results in terms of stock trajectory and terminal stock status, whilst the run just using the MAspring survey had a lower terminal mean F and fitted a more severe doming of the selection pattern, considerably inflating the stock size at older ages. This is not surprising given that the MA survey is on inshore grounds where older age classes are generally absent. The final, combined run places the stock trajectory between these single fleet runs closer to the NEFSC survey runs.

Additional model runs were conducted in which the age range on surveys was truncated to include only those ages considered to be able to follow year-class strength (6 rather than 9). This had a significant effect upon the model results, causing more drastic doming of the fishery selection pattern at the oldest ages as the model fits solely to the catch data. Clearly there is some tension between the two data sources regarding mortality at the oldest ages where the sample numbers are lowest (and therefore subject to greater uncertainty).

The new model formulation appears to have a lower retrospective problem compared to the previous ADAPT model but a consistent bias of over-estimating SSB and under-estimating F remains.

4) Perform a sensitivity analysis which examines the impact of allocation of catch to stock areas on model performance (TOR-3).

Completion of this ToR was satisfactory and the impacts of re-allocation of stock areas was convincingly demonstrated to be minimal.

5) If time permits, consider the small-scale distribution of cod (e.g., spawning sites, resource distribution, fishing effort) in the Gulf of Maine and advise on its management implications.

Completion of this ToR was satisfactory given that it was time dependent and the assessment team were able to begin exploring some of these issues.

The changes in spatial distribution stock as indicated by the survey indices and the extent of the fishery should be treated as important and potentially very serious. The contraction of the stock is entirely consistent with the maintenance of high commercial CPUE whilst the assessment indicates falling stocks and increasing fishing mortality. Such phenomena have been observed before and have preceded major stock collapse (Canadian cod stocks).

6) State the existing stock status definitions for "overfished" and "overfishing". Then update or redefine biological reference points (BRPs; point estimates or proxies for BMSY, BTHRESHOLD, FMSY , and MSY) and provide estimates of their uncertainty. If analytic model-based estimates are unavailable, consider recommending alternative measurable proxies for BRPs. Comment on the appropriateness of existing BRPs and the "new" (i.e., updated, redefined, or alternative) BRPs.

The assessment documentation proposed moving the proxy for F_{msy} from $F_{40\%spr}$ to $F_{35\%spr}$. The justification for doing so was based upon a stock-recruit model fitted from an ASAP run in which the time-line of the model was extended back to 1970 (rather than the 1982 model used in the assessment). From this a value of Fmsy was calculated which corresponds to F35%spr. This convoluted approach is somewhat inconsistent. If the 1970 model is considered appropriate to reliably estimate stock size and recruitment in the earlier period, why was it not used as the basis for estimating the current stock size? It is unclear if the discards were estimated back in this historical part of the assessment, which then raises questions about the resulting recruitment estimates. Even if this assessment time frame were acceptable it is puzzling as to why the direct Fmsy estimate was not taken rather than converting it into terms of %spr. There was no sound justification for the use of F35% over F40%, indeed the effect upon MSY is minimal (although the effect on Bmsy is significant). Given that no scientific justification could be given for this proposed change, the panel unanimously rejected it in favour of retaining the F40% proxy for Fmsy.

7) Evaluate stock status with respect to the existing model (from the most recent accepted peer reviewed assessment) and with respect to a new model developed for this peer review.

   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 (from Cod TOR 6).***

Completion of this ToR was satisfactory.

Regardless of which assessment model was used, or which MSY proxy was selected, the stock remains Overfished and Overfishing is occurring. Given such strong signals from the stock assessments the advice in relation to the OFLs is considered robust to assessment uncertainties.

8) ***Develop and apply analytical approaches to conduct single and multi-year stock projections to compute the pdf (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, and definition of BRPs for black sea bass).***
   b. ***Comment on which projections seem most realistic. Consider major uncertainties in the assessment as well as the 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.***

Completion of this ToR was satisfactory following amendments requested by the review panel.

The forecast simulations had drawn from the full range of historical recruitments. Whilst this may be appropriate for mid- to long-term projections of this stock, given the current low stock size, the expectation of the full range of recruitments is potentially over-optimistic. From a basic biological perspective, very low densities of adults means that the potential for large recruitments must be reduced and the stock is now in a region in which there are few observations on which to draw recruitment expectations. In modelling terms we are therefore in the realms of extrapolation and it is not clear how to manage recruitment expectations at such stock levels (linear decline? depensation?). A linear model taking mean recruitment down to the origin is as arbitrary as any other model but is simple to effect and is neither pessimistic nor optimistic. Following a change to the recruitment model in line with these suggestions, the results of the projections were accepted and it was noted that the stock is unable to rebuild by 2014 irrespective of any fishery management actions.

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

Completion of this ToR was satisfactory. The assessment team worked through each of the previous recommendations for further work and stated what had been done on each item. Most of the previous recommendations have been addressed in one form or another.

Appendix 1: Bibliography of materials provided for review

Black Sea Bass

Cook. R. 2006. Report by the Chair of the 43rd North East Regional Stock Assessment Review Committee (SARC). 73 p.

McCartney, M. and M. L. Burton. 2011. Population genetic structure of black sea bass (Centropristis striata) on the eastern U. S. coast, with an analysis of mixing between stocks north and south of Cape Hatteras, North Carolina. Manuscript from the Southeast Data Analysis and Review Process.

Miller, T. J., R. Muller, R. O'Boyle and A. A. Rosenberg. 2009. Report by the Peer Review Panel for the Northeast Data Poor Stocks Working Group. NOAA NEFSC Reference Document 09-XX

Moser, J. and G. R. Shepherd. 2009. Seasonal distribution and movement of black sea bass (*Centropristis striata*) in the Northwest Atlantic as determined from a mark-recapture experiment. J. Northwest Atl. Fish. Sci. 40: 17-28

Northeast Data Poor Stocks Working Group Report. 2008. Part A. Skate species complex, Deep sea red crab, Atlantic wolfish, scup and black sea bass. NOAA NEFSC Reference Document 09-02A. 86 p.

Shepherd, G. R. and M. Terceiro. 1994. The summer flounder, scup and black sea bass fisheries of the Mid-Atlantic Bight and Southern New England Waters. NOAA Technical Report NMFS 122. 18p

Southern Demersal Working Group. 2011. Stock Assessment Workshop (SAW 53). B. Black Sea Bass. Corrected BSB Assessment (Dated 11/29/11). NEFSC. 148p.

Gulf of Maine Cod

Anonymous. 2008. GARM III Executive Summary. 30p. In R. O'Boyle et al (Eds.) GARM III. Document Summary

Mayo, R., G. Shepherd, L. O'Brien, L. Col and M. Traver. 2008. Appendix F. Gulf of Maine cod Assessment. Pages 228 -582 In R. O' Boyle et al (Eds). GARM III. Document Summary.

Northern Demersal Working Group. 2011. Stock Assessment Workshop (SAW 53) A. Gulf of Maine Atlantic cod (*Gadus morhua*) stock assessment updated through 2010. NEFSC. 296p.

O'Boyle, R., V. Crecco, L. Van-Eeckhaute, D. Kahn, C. Needle, B. Rothschild, S. Smith, and J. H. Volstad. 2008. Report of the Groundfish Assessment Review Meeting (GARM III). Part 1. Data Methods. DOCUMENT SUMMARY 64 p.

O'Boyle, R. , J. De Oliveira, S. Gavaris, J. Ianelli, Y. Jioa, C. M. Jones and P. Medley. 2008. Report of the Groundfish Assessment Review Meeting (GARM III). Part 2. Assessment Methodology (Models). DOCUMENT SUMMARY. 70p.

O'Boyle, R. M. Bell, S. Gavaris, V. Haist, S. Reeves and G. Thompson. 2008. Report of the Groundfish Assessment Review Meeting (GARM III). Part 3 Biological Reference Points. DOCUMENT SUMMARY. 88p.

Working Group in Re-Evaluation of Biological Reference Points for New England Groundfish. 2002. Re-Evaluation of Biological Reference Points for New England Groundfish. NOAA Northeast Fisheries Science Center Reference Document 02-04. 107p.

Appendix 2: A copy of the CIE Statement of Work

Attachment A: Statement of Work for Dr. Ewen Bell (CEFAS)

External Independent Peer Review by the Center for Independent Experts

53rd Stock Assessment Workshop/Stock Assessment Review Committee (SAW/SARC): Black sea bass and Gulf of Maine cod.

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

Scope of Work and CIE Process: 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 Technical Representative (COTR), 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 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.

Project Description: The purpose of this meeting will be to provide an external peer review of stock assessments for black sea bass (*Centropristis striata*) and Gulf of Maine Atlantic cod (*Gadus morhua*). Black sea bass occupy reefs, wrecks and shell bed habitats. They may attain lengths up to 60 cm with maximum age of 10-12 years. Black sea bass change sex from female to male between ages 2 to 5. Black sea bass are jointly managed by the Atlantic States Marine Fisheries Commission (ASMFC) and the Mid-Atlantic Fishery Management Council. The last peer reviewed assessment of black sea bass was in 2008 as part of the Data Poor Stocks Working Group, with annual updates since then. The Atlantic cod is a demersal gadoid species found on both sides of the North Atlantic. Cod may attain lengths up to 130 cm with maximum age in excess of 20 years. Commercial and recreational fisheries for cod are managed by the New England Fishery Management Council. The last peer reviewed assessment of Gulf of Maine cod was in 2008 as part of the GARM III. Results of the 2011 peer review will form the scientific basis for fishery management in the northeast region.

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), which are carried out by the SAW Working Groups, 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.**

The SARC 53 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 Mid-Atlantic Fishery Management Council. The SARC panel will write the SARC Summary Report and each CIE reviewer will write an individual independent review report.

Requirements for CIE Reviewers: Three CIE reviewers shall conduct an impartial and independent peer review of the stock assessments that are provided, and this review should be in accordance with this SoW and stock assessment ToRs herein. CIE reviewers shall have working knowledge and recent experience in fish stock assessments. For sea bass, knowledge of complex life histories and their implications for Biological Reference Points is desirable. For GOM cod, familiarity with forward projecting models and estimation is desirable.

In general, CIE reviewers for SARCs shall have working knowledge and recent experience in the application of modern fishery stock assessment models. Expertise shall include statistical catch-at-age, state-space and index methods. Reviewers shall also have experience in evaluating measures of model fit, identification, uncertainty, and forecasting. Reviewers shall 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 BRPs.

Each CIE reviewer's duties shall not exceed a maximum of 15 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 15 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).

Location of Peer Review: Each CIE reviewer shall conduct an independent peer review during the panel review meeting scheduled in Woods Hole, Massachusetts during November 29 – December 2, 2011.

Charge to SARC panel: During the SARC meeting, the panel is to determine and write down whether each stock assessment Term of Reference 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. 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 Biological Reference Points (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.

Statement of Tasks:

1. Prior to the meeting

(SARC chair and CIE reviewers)
Review the reports produced by the Working Groups and read background reports.

Each CIE reviewer shall complete the following tasks in accordance with the SoW and Schedule of Milestones and Deliverables herein:

Upon completion of the CIE reviewer selection by the CIE Steering Committee, the CIE shall provide the CIE reviewer information (full name, title, affiliation, country, address, email, and FAX number) to the COTR, who forwards this information to the NMFS Project Contact no later the date specified in the Schedule of Milestones and Deliverables. The CIE is responsible for providing the SoW and stock assessment ToRs to the CIE reviewers. The NMFS Project Contact is responsible for providing the CIE reviewers with the background documents, reports, foreign national security clearance, and other information concerning pertinent meeting arrangements. The NMFS Project Contact is also 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 COTR prior to the commencement of the peer review.

Foreign National Security Clearance: When CIE reviewers participate during a panel review meeting at a government facility, the NMFS Project Contact is responsible for obtaining the Foreign National Security Clearance approval for CIE reviewers who are non-US citizens. For this reason, the CIE reviewers shall provide by FAX the requested information (e.g., first and last name, contact information, gender, birth date, passport number, country of passport, travel dates, country of citizenship, country of current residence, and home country) to the NMFS Project Contact for the purpose of their security clearance, and this information shall be submitted at least 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: 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 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 CIE Lead Coordinator on where to send documents. CIE reviewers are responsible only for the pre-review documents that are delivered to the reviewer in accordance to the SoW scheduled deadlines specified herein. The CIE reviewers shall read all documents in preparation for the peer review.

2. During the Open meeting

Panel Review Meeting: Each CIE 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 COTR and CIE Lead Coordinator.** 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 discussion, 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.

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. 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.

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.

3. After the Open 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 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).

Contract Deliverables - Independent CIE Peer Review Reports: Each CIE reviewer shall complete an independent peer review report in accordance with the SoW. Each CIE reviewer shall complete the independent peer review according to required format and content as described in **Annex 1**. Each CIE 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 November 29 – December 2, 2011.
3) Conduct an independent peer review in accordance with this SoW and the assessment ToRs (listed in **Annex 2**).
4) No later than December 16, 2011, 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 to David Sampson, CIE Regional Coordinator, via email to david.sampson@oregonstate.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: CIE shall complete the tasks and deliverables described in this SoW in accordance with the following schedule.

24 October 2011	CIE sends reviewer contact information to the COTR, who then sends this to the NMFS Project Contact
15 November 2011	NMFS Project Contact will attempt to provide CIE Reviewers the pre-review documents by this date
Nov. 29 – Dec. 2 2011	Each reviewer participates and conducts an independent peer review during the panel review meeting in Woods Hole, MA
1-2 December 2011	SARC Chair and CIE reviewers work at drafting reports during meeting at Woods Hole, MA, USA
16 December 2011	CIE reviewers submit draft CIE independent peer review reports to the CIE Lead Coordinator and CIE Regional Coordinator
19 December 2011	Draft of SARC Summary Report, reviewed by all CIE reviewers, due to the SARC Chair *
23 December 2011	SARC Chair sends Final SARC Summary Report, approved by CIE reviewers, to NEFSC contact (i.e., SAW Chairman)
30 December 2011	CIE submits CIE independent peer review reports to the COTR
6 January 2012	The COTR distributes the final CIE 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 COTR within 10 working days after receipt of all required information of the decision on substitutions. The COTR 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 CIE 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: Upon review and acceptance of the CIE independent peer review reports by the CIE Lead Coordinator, Regional Coordinator, and Steering Committee, these reports shall be sent to the COTR for final approval as contract deliverables based on compliance with the SoW and ToRs. As specified in the Schedule of Milestones and Deliverables, the CIE shall send via e-mail the contract deliverables (CIE independent peer review reports) to the COTR (William Michaels, via William.Michaels@noaa.gov).

Applicable Performance Standards: The contract is successfully completed when the COTR provides final approval of the contract deliverables. The acceptance of the contract deliverables shall be based on three performance standards:
(1) each CIE report shall be completed with the format and content in accordance with **Annex 1**,
(2) each CIE report shall address each stock assessment ToR listed in **Annex 2**,
(3) the CIE reports shall be delivered in a timely manner as specified in the schedule of milestones and deliverables.

Distribution of Approved Deliverables: Upon acceptance by the COTR, the CIE Lead Coordinator shall send via e-mail the final CIE reports in *.PDF format to the COTR. The COTR will distribute the CIE reports to the NMFS Project Contact and Center Director.

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: 301-713-2363 ext 136

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:

NMFS Project Contact:

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

Mr. Frank Almeida, Acting NEFSC Science Director
National Marine Fisheries Service, NOAA
Northeast Fisheries Science Center
166 Water St., Woods Hole, MA 02543
frank.almeida@noaa.gov Phone: 508-495-2233

Annex 1: Format and Contents of CIE Independent Peer Review Report

1. The CIE independent 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 reviewer 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 Term of Reference of the SAW was completed successfully. For each Term of Reference, the Independent Review 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.

   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 CIE 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 CIE independent report shall be an independent peer review of each ToRs, 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 the CIE Statement of Work
   Appendix 3: Panel Membership or other pertinent information from the panel review meeting.

Annex 2: Stock Assessment Terms of Reference for SAW/SARC53
(to be carried out by SAW Working Groups) (file vers.: 5/20/11)

A. Black sea bass

1. Estimate catch from all sources including landings and discards. Characterize the uncertainty in these sources of data. Evaluate available information on discard mortality and, if appropriate, update mortality rates applied to discard components of the catch. Describe the spatial and temporal distribution of fishing effort.
2. Present the survey data being used in the assessment (e.g., 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. Consider known aspects of seasonal migration and availability of black sea bass, and investigate ways to incorporate these into the stock assessment. Based on the known aspects, evaluate whether more than one management unit should be used for black sea bass from Cape Hatteras north and, if so, propose unit delineations that could be considered by the Mid-Atlantic Fishery Management Council and for use in future stock assessments.

4. Investigate estimates of natural mortality rate, M, and if possible incorporate the results into TOR-5. Consider including sex- and age-specific rate estimates, if they can be supported by the data.

5. Estimate annual fishing mortality, recruitment and appropriate measures of stock biomass (both total and spawning stock) for the time series (integrating results from TOR-4), and estimate their uncertainty. Include a historical retrospective analysis to allow a comparison with most recent assessment results.
6. State the existing stock status definitions for "overfished" and "overfishing". Then update or redefine biological reference points (BRPs; point estimates or proxies for B_{MSY}, $B_{THRESHOLD}$, F_{MSY}, and MSY) and provide estimates of their uncertainty. If analytic model-based estimates are unavailable, consider recommending alternative measurable proxies for BRPs. Comment on the appropriateness of existing BRPs and the "new" (i.e., updated, redefined, or alternative) BRPs.

7. Evaluate stock status with respect to the existing model (from the most recent accepted peer reviewed assessment) and with respect to a new model developed for this peer review.
    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 (from black sea bass TOR 6).

8. Develop and apply analytical approaches to conduct single and multi-year stock projections to compute the pdf (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, and definition of BRPs for black sea bass).

b. Comment on which projections seem most realistic. Consider major uncertainties in the assessment as well as the 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.

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

B. Cod (Gulf of Maine Stock)

1. Estimate catch from all sources including landings and discards. Characterize the uncertainty in these sources of data. Evaluate available information on discard mortality and, if appropriate, update mortality rates applied to discard components of the catch.
2. Present the survey data being used in the assessment (e.g., 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. 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, catch recruitment and fishing mortality.
4. Perform a sensitivity analysis which examines the impact of allocation of catch to stock areas on model performance (TOR-3).
5. If time permits, consider the small-scale distribution of cod (e.g., spawning sites, resource distribution, fishing effort) in the Gulf of Maine and advise on its management implications.
6. State the existing stock status definitions for "overfished" and "overfishing". Then update or redefine biological reference points (BRPs; point estimates or proxies for B_{MSY}, $B_{THRESHOLD}$, F_{MSY} , and MSY) and provide estimates of their uncertainty. If analytic model-based estimates are unavailable, consider recommending alternative measurable proxies for BRPs. Comment on the appropriateness of existing BRPs and the "new" (i.e., updated, redefined, or alternative) BRPs.

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

8. Develop and apply analytical approaches to conduct single and multi-year stock projections to compute the pdf (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.

9. Review, evaluate and report on the status of the SARC and Working Group research recommendations listed in recent SARC reviewed assessments 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 2 (cont)
Appendix to the Assessment TORs (cont.):

ABC Control Rule Methods Proposed by the Mid-Atlantic Fishery Management Council:

A multi-level approach will be used for setting an ABC for each Mid-Atlantic stock, based on the overall level of scientific uncertainty associated with its assessment. The stock assessment will be required to provide estimates of the maximum fishing mortality threshold (MFMT) and future biomass, the probability distributions of these estimates, the probability distribution of the overfishing limit (OFL; level of catch that would achieve MFMT given the current or future biomass), and a description of factors considered and methods used to estimate their distributions. The multi-level approach defines four levels of overall assessment uncertainty defined by characteristics of the stock assessment and determination by the SSC that the uncertainty in the probability distribution of OFL adequately represents best available science. The procedure used to determine ABCs is different in each level of the methods framework. The SSC will determine to which level the assessment for a particular stock belongs when setting single or multi-year ABC specifications and a description of the justification for assignment to a level will be provided with the ABC recommendation. The ABC receommdations should be more precautionary as an assessment moves from level 1 to level 4. Recommendations for ABC may be made for up to 3 years for all of the managed resources except spiny dogfish which may be specified for up to 5 years. The rationale for assigning an assessment to a level will be reviewed each time an ABC determination is made.

Levels of stock assessments, characteristics, and procedures for determining ABCs are defined as follows:

Level 1: Level 1 represents the highest level to which an assessment can be assigned. Assignment of a stock to this level implies that all important sources of uncertainty are fully and formally captured in the stock assessment model and the probability distribution of the OFL calculated within the assessment provides an adequate description of uncertainty of OFL. Accordingly, the OFL distribution will be estimated directly from the stock assessment. In addition, for a stock assessment to be assigned to Level 1, the SSC must determine that the OFL probability distribution represents best available science. Examples of attributes of the stock assessment that would lead to inclusion in Level 1 are:

- Assessment model structure and any treatment of the data prior to inclusion in the model includes appropriate and necessary details of the biology of the stock, the fisheries that exploit the stock, and the data collection methods;
- Estimation of stock status and reference points integrated in the same framework such that the OFL calculations promulgate all uncertainties (stock status and reference points) throughout estimation and forecasting;
- Assessment estimates relevant quantities including F_{MSY}[1], OFL, biomass reference points, stock status, and their respective uncertainties; and
- No substantial retrospective patterns in the estimates of fishing mortality (F), biomass (B), and recruitment (R) are present in the stock assessment estimates.

The important part of Level 1 is that the precision estimated using a purely statistical routine will define the OFL probability distribution. Thus, all of the important sources of uncertainty

[1] With justification, F_{MSY} may be replaced with an alternative maximum fishing mortality threshold to define the OFL.

are formally captured in the stock assessment model. When a Level 1 assessment is achieved, the assessment results are likely unbiased and fully consider uncertainty in the precision of estimates. Under Level 1, the ABC will be determined solely on the basis of an acceptable probability of overfishing (P*), determined by the Council's risk policy (see alternatives in section 5.2.2), and the probability distribution of the OFL.

Level 2: Level 2 indicates that an assessment has greater uncertainty than Level 1. Specifically, the estimation of the probability distribution of the OFL directly from the stock assessment model fails to include some important sources of uncertainty, necessitating expert judgment during the preparation of the stock assessment, and the OFL probability distribution is deemed best available science by the SSC. Examples of attributes of the stock assessment that would lead to inclusion in Level 2 are:

- Key features of the biology of the stock, the fisheries that exploit it, or the data collection methods are missing from the stock assessment;
- Assessment estimates relevant quantities, including reference points (which may be proxies) and stock status, together with their respective uncertainties, but the uncertainty is not fully promulgated through the model or some important sources may be lacking;
- Estimates of the precision of biomass, fishing mortality rates, and their respective reference points are provided in the stock assessment; and
- Accuracy of the MFMT and future biomass is estimated in the stock assessment by using *ad hoc* methods.

In this level, ABC will be determined by using the Council's risk policy (see alternatives in section 5.2.2), as with a Level 1 assessment, but with the OFL probability distribution based on the specified distribution in the stock assessment.

Level 3: Attributes of a stock assessment that would lead to inclusion in Level 3 are the same as Level 2, except that

- The assessment does not contain estimates of the probability distribution of the OFL or the probability distribution provided does not, in the opinion of the SSC, adequately reflect uncertainty in the OFL estimate.

Assessments in this level are judged to over- or underestimate the accuracy of the OFL. The SSC will adjust the distribution of the OFL and develop an ABC recommendation by applying the Council's risk policy (see alternatives in section 5.2.2) to the modified OFL probability distribution. The SSC will develop a set of default levels of uncertainty in the OFL probability distribution for this level based on literature review and a planned evaluation of ABC control rules. A control rule of 75 percent of F_{MSY} may be applied as a default if an OFL distribution cannot be developed.

Level 4: Stock assessments in Level 4 are deemed to have reliable estimates of trends in abundance and catch, but absolute abundance, fishing mortality rates, and reference points are suspect or absent. Additionally, there are limited circumstances that may not fit the standard approaches to specification of reference points and management measures set forth in these guidelines (i.e., ABC determination). In these circumstances, the SSC may propose alternative approaches for satisfying the NS1 requirements of the Magnuson-Stevens Act than those set forth in the NS1 guidelines. In particular, stocks in this level do not have point estimates of the OFL or probability distributions of the OFL that are considered best available science. In most cases, stock assessments that fail peer review or are deemed highly uncertain by the SSC will be assigned to this level. Examples of potential attributes for inclusion in this category are:

- Assessment approach is missing essential features of the biology of the stock, characteristics of data collection, and the fisheries that exploit it;
- Stock status and reference points are estimated, but are not considered reliable;
- Assessment may estimate some relevant quantities including biomass, fishing mortality or relative abundance, but only trends are deemed reliable;
- Large retrospective patterns usually present; and
- Uncertainty may or may not be considered, but estimates of uncertainty are probably substantially underestimated.

In this level, a simple control rule will be used based on biomass and catch history and the Council's risk policy.

The SSC will determine, based on the assessment level to which a stock is classified, the specifics of the control rule to specify ABC that would be expected to attain the probability of overfishing specified in the Council's risk policy. The SSC may deviate from the above control rule methods framework or level criteria and recommend an ABC that differs from the result of the ABC control rule calculation, but must provide justification for doing so.

(~:\sarc\sarc53…\TORs\DRAFT SAW-SARC-53_[date].doc)

(END OF ANNEX 2)

Annex 3: Draft Agenda
53rd Northeast Regional Stock Assessment Workshop (SAW 53)
Stock Assessment Review Committee (SARC) Meeting

November 29 – December 2, 2011

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

(A meeting agenda will be provided approximately 2 months before the meeting. Reviewers must attend the entire meeting.)

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 any 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: Panel Membership or other pertinent information from the panel review meeting.

- Thomas J. Miller (Chair, MAFMC SSC & University of Maryland Center for Environmental Science, Solomons, MD)
- Ewen Bell (CIE Reviewer, CEFAS, Lowestoft, Suffolk, UK
- Kenneth Patterson (CIE Reviewer, Brussels, Belgium)
- Kurtis Trzcinski (CIE Reveiwer, DFO, Bedford, NS, Canada)