New England Fishery Management Council
50 WATER STREET | NEWBURYPORT, MASSACHUSETTS 01950 | PHONE 978 465 0492 | FAX 978 465 3116
C. M. "Rip" Cunningham, Jr., *Chairman* | Paul J. Howard, *Executive Director*

To: Paul J. Howard, Executive Director
From: Scientific and Statistical Committee
Date: April 20, 2012

Subject: **SSC review of Butterworth and Rademeyer report**

Background
The Scientific and Statistical Committee (SSC) was asked to:

1) Review the paper, "An Investigation of Differences Amongst SCAA [Statistical Catch at Age] and ASAP [Age Structured Assessment Program] Assessment (including Reference Point) Estimates for Gulf of Maine Cod" (January 2012) by Doug S. Butterworth and Rebecca A. Rademeyer and advise the Council on its applicability for revising management reference points and the approach used to assess Gulf of Maine cod in SAW/SARC-53 based on the following issues addressed in the paper:

    a. Pope vs. Baranov dynamics
    b. Estimation of the starting numbers-at-age vector
    c. The selection of the starting year for the assessment
    d. Allowance for additional variance in fitting to the time series of abundance indices
    e. Fitting abundance indices expressed in terms of mass or of numbers
    f. The form of the term for catch-at-age proportions in the log-likelihood
    g. Domed vs. flat selectivity-at-age for the NEFSC surveys
    h. Estimation of a stock-recruitment curve

In order to meet this term of reference, the SSC considered the following:

1. "An Investigation of Differences Amongst SCAA and ASAP Assessment (including Reference Point) Estimates for Gulf of Maine Cod" by Doug S. Butterworth and Rebecca A. Rademeyer (January 2012)
2. Response to Comments by NMFS on Reference Point Estimation for Gulf of Maine Cod by Butterworth and Rademeyer, D. S. Butterworth, February 2012
3. Applications of Statistical Catch-at-Age Assessment Methodology to Gulf of Maine Cod, Doug S. Butterworth and Rebecca A. Rademeyer (October 2011)
4. Powerpoint presentation summarizing deliberations and decisions made during SAW 53 by Dr. Liz Brooks, NEFSC and chair of the working group.

Acknowledgement
The SSC would like to express our thanks and appreciation to Doug Butterworth, Rebecca Rademeyer, Liz Brooks and Mike Palmer for their work to produce such a rich body of analysis, and their efforts to communicate the outcomes of their work through reports and presentations over the past two SSC meetings.

SSC Conclusions/Discussion
Additional guidance from the Council Chair at the meeting urged the SSC to focus on those issues that would have the most significant implications for management decisions, rather than disputes that are primarily technical in nature. That advice resulted in many issues receiving relatively little discussion, generally when Dr. Butterworth and Dr. Brooks readily agreed that a given topic had little bearing on the outcomes of the assessment. Conversely, TOR 1.c., 1.h., and to a lesser extent 1.g. were identified as having the most significant implications, and therefore were the focus of the majority of the SSC discussions. Consensus statements are provided immediately below each topic, and additional notes are provided for some topics to explain and expand upon the conclusions presented in this report.

a. Pope versus Baranov dynamics
There is little difference in results, and this is not an important issue.

b. Estimation of starting numbers-at-age vector
The SSC agreed with the approach taken in the SAW/SARC 53 assessment that the estimation of abundance at age in the first year is more appropriate than estimating the ratio of biomass to unfished biomass (B1/Bo) and mortality (Z) in the first year.

c. The selection of the starting year for the assessment
Historical productivity should generally be evaluated more thoroughly in future assessments, although the SSC expresses no preference for the exact starting year of the GOM cod assessment. (As with many other topics, this is an issue that requires more information and discussion than was possible to cover in an SSC meeting. It should be considered in a benchmark assessment).

Notes:
This TOR received considerable attention and debate by the SSC. It should be noted that many of the issues related to this TOR are also relevant to TOR 1.h., and vice versa, so advice and comments on the two TORs should be considered together.

There were several issues pertinent to determining the most appropriate starting year for the assessment:

- Statistical decision criteria.
  Butterworth and Rademeyer argue that AIC (Akaike Information Criterion, used to measure model performance) should be the primary determinant, and that overriding an AIC outcome needs to be justified. Other participants in the discussion proposed two additional criteria, CVs of parameter estimates and whether or not the estimate converged to a bound. High CVs are seen as indicative of high uncertainty, and convergence to a bound is seen to be indicative of an unreliable estimate at an extreme constraint. The SSC generally saw merit in using a variety of statistical

metrics to make decisions, but did not precisely define which to use and how. Discussion focused more on the other issues relevant to this TOR.

- Trade-offs between information content and uncertainty.
  The SSC generally agreed that an assessment should use as much reliable information as possible. The central issue then becomes how to determine whether particular information is reliable, which relates to how much additional uncertainty is introduced for the additional information provided.

  On the one hand, the longer time series available would give the most information. However, proportions-at-age are not available for commercial catches prior to 1982, and for survey catches prior to 1970. This means that assessments beginning earlier incorporate fewer data and more assumptions. Since these issues arise because the assessment uses an age-based model, one idea offered by the SSC is to use data-poor proxy methods that rely only on catch and/or survey data alongside the fully age-based model to generate and compare additional insights.

  Whether or not to select an earlier starting year is a critical question when earlier years exhibit low recruitment at high biomass, as is the case for some of the model runs by Butterworth and Rademeyer. This can fundamentally change the nature of the stock-recruitment relationship, expected stock dynamics as biomass grows, and estimation of reference points. Butterworth and Rademeyer noted that estimating MSY-based reference points requires the greatest possible contrast and therefore the maximum amount of data in the assessment. However, others noted that reference points are estimated external to the assessment model (i.e., based on analysis of maximum spawning potential), and the need is much less in this case.

- Understanding of historical conditions.
  In addition to the technical and analytical aspects of choosing a starting year, the SSC discussed the extent to which changes in both the ecosystem and fishery through time might affect the decision. The possibility that the ecosystem has undergone a significant regime shift since the 1960s was raised and discussed by the SSC. If that has occurred, then including dynamics from earlier years in an assessment with later years would not be appropriate given that the stock would not be operating under consistent processes as assumed by the model.

  The SSC also had some discussion, spurred by audience comment, about the nature and effects of changes in the fishery. In particular, changes in mesh size through time, other major gear changes, location and timing of fishing, and targeting and use of different size classes (e.g., retention of very small cod as bait historically) might all potentially introduce changes in the system of a scale beyond those the model assumed and could accommodate.

Ultimately, the SSC did not endorse a particular starting year for the assessment. This was due in large part to the questions about major historical changes in the ecosystem and fishery, more so than the analytical issues. The SSC sees potential to resolve the analytical issues by using a variety of approaches and comparing the outcomes and implications of each in a

decision framework. This could be an important component of the more explicit risk evaluation process that will be developed soon.

Furthermore, the SSC sees the questions about historical changes as being paramount to determining how far back in time a single assessment model should go, analytical issues notwithstanding. The SSC meeting was clearly not the appropriate venue to determine whether, when and to what extent a regime shift has occurred, nor was it the appropriate venue to reconstruct the history of fishing gear, behavior, markets, etc. These would be important issues to resolve outside of the assessment process, which would be of tremendous value in making similar decisions in future assessments.

d. Allowance for additional variance in fitting to the time series of abundance indices
Both the ASAP and SCAA models allow for additional variances. The SSC determined this is not a major issue, since it does not significantly change the estimates of spawning stock biomass and recruitment.

e. Fitting abundance indices expressed in terms of mass or of numbers
The relative merits of each approach should be considered further.

Notes:
The SSC considered whether there is a general precedent related to this issue, or whether the problem is unique to this assessment. Dr. Butterworth noted that the default approach is often to use biomass, but to use numbers when diagnostics suggest poor performance of biomass-based indices. Dr. Butterworth's report also reiterates the earlier findings of the working group that the effects of using one approach rather than the other are comparatively small. The SSC therefore spent little time addressing this issue, and reiterated the conclusion of past assessments that it is an issue warranting continued investigation where the implications of the decision are more significant.

f. The form of the term for catch-at-age proportions in the log-likelihood
Further exploration of the two options is needed.

Notes:
Butterworth and Rademeyer noted comparable analytical limitations with either approach, but also that diagnostics suggest neither is fatally flawed. Dr. Brooks noted that the Butterworth and Rademeyer analysis did not directly compare their approach with ASAP, which limited the ability of the SSC to fully consider the issue relative to the final assessment. Nevertheless, given the limitations noted by Butterworth and Rademeyer, this is another area requiring continued investigation, but also an issue the SSC spent little time addressing in the interest of time.

g. Domed vs. flat selectivity-at-age for the NEFSC surveys
More work is needed to determine the most likely nature of the selectivity function, although the decision appears to have comparatively little effect on the outcome of the GOM cod assessment.

Notes:

This issue has been a source of considerable debate in multiple assessments, and is another issue that warrants continued investigation. The SSC supported the issue remaining a research priority. A key consideration when adopting a domed selectivity function is the effects of cryptic biomass (i.e., biomass of older fish that is not well sampled) on our understanding of stock dynamics and status. It was noted, however, that the relative degree of doming between commercial and survey gears is a less important issue than the degree of doming within any given data series.

h. Estimation of stock-recruitment curve

Stock-recruitment relationships and alternative modeling approaches should be explored further.

Notes:

Many of the issues relevant to determining whether and how to estimate a stock-recruitment relationship are similar to those related to selection of a starting year for the assessment. These are discussed above under TOR 1.c. In particular, the trade-offs between the additional information provided by extending the assessment further back in time relative to the additional uncertainties introduced by the data limitations of those early years is critical to determining whether and how to estimate a stock-recruitment curve. Butterworth and Rademeyer estimate generally low recruitment at high biomass in the earliest starting years they considered. Including those values creates a dome-shaped curve, suggesting strong density-dependent processes are operating at high biomass. That doming lowers the estimate of B_{MSY}, and therefore changes determination of stock status, because not only is productivity maximized at much lower biomass, but it also declines sharply at very high biomass.

Dr. Butterworth noted that, although the biomass estimates for the earliest years have the highest uncertainty, it is not an unacceptable high level of uncertainty. However, Dr. Brooks noted that the outcome of a dome-shaped stock-recruitment curve, and its resultant implications for reference points and status determination, is driven by biomass estimates for the earliest years being located just beyond the "cloud" of points for later years. The uncertainty in those estimates means that the actual value could be shifted toward the left and right into the cloud, or toward the right and further away. Interestingly, either shift would have the effect of reducing the extent of doming, and therefore the outcome seems highly sensitive to the exact location of five data points.

The SSC suggested that debates about the shape of a stock-recruitment curve should consider not only statistical considerations, but also consideration of first principles. Hypotheses about the ecological dynamics underlying a stock-recruitment function of any shape should be identified and evaluated to determine the likelihood of any competing shape of the function (e.g., would cannibalism be the reason for reduced recruitment at high biomass?). Developing those hypotheses should also include characteristics of the fishery. For example, one SSC member questioned whether small-mesh fishing in nearshore nursery grounds in the 1960s and 1970s increased mortality of juvenile fish relative to later years, thus fundamentally violating the assumption of constant selectivity. This could increase the estimated size of year classes in the early years of the assessment and raises the question of whether recruitment does in fact decline at high biomass.

Ultimately, the SSC agreed that this is an issue warranting continued investigation, as recommended by multiple assessment reports. In fact, the SSC itself has wrestled with these issues in past deliberations, and has been unable to reach a clear consensus. That outcome is not surprising given the numerous technical issues related to this question. To help better structure future debates on this topic, the SSC proposed three critical questions that need to be addressed in sequence:

1) Is there evidence for a stock-recruitment relationship?
2) What is its most likely form?
3) What factors should be considered and approaches should be employed in estimation? (e.g., how to treat high variances that significantly affect the outcomes?)

If the answer to question #1 is "no", then discussion should shift to alternative approaches and the implications of those.

Options to consider when addressing question #3 might include use of smoothers to model shape. Standard methods to characterize variance and test stability of the S-R relationship such as splitting the time series into a training set and a testing set, cross-validation, jackknifing, or bootstrapping, should be considered. These methods will provide better insight into reliability of Bmsy estimates.

Furthermore, the SSC reiterated its conclusion from the discussion of TOR 1.c. that there could be tremendous value in moving toward a more robust decision framework that considers the outcomes, implications and limitations of multiple modeling approaches simultaneously to better evaluate risk.

Management strategy evaluation (MSE) could be a powerful tool in that process, and is one that is underutilized in fisheries management in New England. Like the GOM cod assessment report and most others, the Butterworth and Rademeyer report provides the sorts of tables summarizing outcomes of multiple modeling approaches that would be part of a broader decision framework, although the SSC noted that they cannot be used in that capacity at present as the numerous individual results require much greater review than was possible with the time the SSC had at hand. Dr. Butterworth agreed, and noted that their report represents exploratory modeling of alternative approaches. A true decision table would require additional sensitivity analyses to be conducted and presented alongside the outcomes.