RESPONSE TO COMMENTS BY NMFS ON REFERENCE POINT ESTIMATION FOR GULF OF MAINE COD BY BUTTERWORTH AND RADEMEYER

D S Butterworth

February 2012

In a letter dated 17 January 2012 to the Executive Director of the New England Fisheries Management Council, scientists from the NMFS NEFSC at Woods Hole offer commentary on reference point estimation analyses for the Gulf of Maine cod in Butterworth and Rademeyer (2011). While some of these points are well made, they have in the main already been addressed by subsequent analyses reported in Butterworth and Rademeyer (2012). Specific responses are provided below. For ease of cross-reference, the original NMFS comments are copied below, with the responses entered in italics at appropriate points.

NMFS Commentary

Biological Reference Points are based on an ASAP run back to 1970 (longer time series than the actual assessment) assuming a Beverton and Holt stock-recruit function.

However, analysis by Butterworth and Rademeyer (2011) demonstrate that if the model is extended into the late 1960s a decline in recruitment at extremely high stock sizes is present (possibly due to cannibalism on juveniles by adult cod, etc...) indicating that a Ricker style stock recruit curve is more appropriate and model estimates indicate that GoM cod is NOT overfished.

Indeed the mechanism of cannibalism mentioned could lead to this effect. It could also be a reflection of multi-species effects that lead to over-estimation of biomass reference points when determined using single species models which ignore these effects. Nevertheless it is important to stress that estimates put forward by Butterworth and Rademeyer are based on direct analysis of the data available using best practice statistical model selection criteria.

The biological reference points approved by the SARC Panel are NOT based on a spawner-recruit relationship, but rather on long-term projections at $F_{40\%}$ (consistent with the methods used to establish biological reference points in the previous assessment).

Advice is required regarding MSY-related reference points. Estimation of such values within an age-structured population model is by construction not possible without associated specification/estimation of a stock-recruitment relationship. Proxies, such as $F_{40\%}$, which implicitly are assuming some stock-recruitment relationship, are often used when the data available to assess a stock do not admit satisfactory direct estimation of a stock recruitment relationship for that stock. However if that relationship can be so estimated in a particular case, use of the best available science dictates that it must necessarily supplant the implicit relationship associated with the proxy.

The methodology proposed by Butterworth and Rademeyer were not supported by the Models Working Group (of which Butterworth was a member) because (a) age composition

data for the fishery are not available prior to 1982 leading to high uncertainty in recruitment estimates; and (b) the stock-recruit function, even when estimated through an extended model time series, is poorly defined. Hence, the biological reference points estimated from such models are uninformative.

First, by way of background, the approach proposed by Butterworth and Rademeyer is not novel, but rather common accepted best practice in many parts of the world. One of the reasons underlying the development of the SCAA assessment approach was to allow the extension of the periods of time covered by age-disaggregated assessments to include years for which catch-at-age information was not available; this in turn can provide better data contrast and hence facilitate the estimation of MSY-related reference points. Indeed the NEFSC survey series are often promoted as the longest and best of their type available for any fishery worldwide; they compare more than favourably with other abundance indices which have been used in extending assessments elsewhere in the world further back in time, where such assessments have been endorsed by multiple peer reviews by leading international scientists in the stock assessment field. The approach, which is also frequently used by some other NMFS Fisheries Science Centers such as the Northwest and the Alaska Centers, was first proposed for the Gulf of Maine cod resource over a decade ago. The final Panel in the 2008 GARM III process did not reject the approach from Gulf of Maine cod, but advanced another approach at that time on the grounds that it provided more conservative estimates. That Panel also selected a near identical approach using a Ricker stock-recruitment relationship, and starting in a year well before proportions-at-age data were available, for the assessment of white hake.

In the October 2011 meeting of the Models Working Group two core reservations were raised about a re-application of this methodology given revised and updated data for the Gulf of Maine cod stock:

- *i) uncertainties arising from the absence of catch-at-age data prior to 1982; in reply to the response that robustness of results to this uncertainty had been demonstrated in analyses presented in 2008, the counter-argument was made that that result had not been confirmed to hold for the revised and updated data; and*
- *ii) the finding of a domed Ricker-like stock recruitment function was heavily dependent on estimates of poor recruitments in the late 1960s when spawning biomass was high, but the 1960s were years where neither the catch nor survey samples had been aged.*

It was not possible in the time available at that meeting to carry out the calculations needed to address these issues. Hence there was agreement to proceed at that time with the assessment starting in 1982. However, the meeting also developed a list of unresolved issues requiring further analysis to move towards settlement of such issues, and Butterworth and Rademeyer (2012) was produced in direct response to that development.

Item i) above is addressed in similar manner to 2008 on pg 6 [see paragraph i) on that page] of Butterworth and Rademeyer (2012), with robustness of results to this concern again being demonstrated. An offer was made there to run more sensitivities to further check this robustness if such were suggested, but no suggestions have been received in the more than two months since this part of the paper was first made available.

Item ii) above is addressed in detail on pg 4 of Butterworth and Rademeyer (2012) where it is shown that information to confirm these poor recruitments of the late 1960s is provided by the survey age data of the early 1970s.

The assertion of high uncertainty in recruitment estimates prior to 1982 is at variance with results shown in Figs 4(b) and 16(b) of Butterworth and Rademeyer (2012), where the confidence intervals shown for annual recruitment estimates demonstrate that the variances of these estimates prior to 1982 are generally small and only marginally greater than for the years after 1982. Similarly the assertion that the biological reference points estimated are consequently uninformative is not supported by the estimates of the CVs for these quantities shown in Tables 4 and 6 of Butterworth and Rademeyer (2012) – for example for assessments commencing in 1964, the CV's for B_{MSY} are of the order of 10%.

The decision of the Models Working Group not to use data prior to 1982 in the modeling work was supported by the SARC Panel. Additionally, in a sensitivity exercise, the Models Working Group actually used a SR relationship from a model with a 1970 start date to justify the use of $F_{35\%}$ as opposed to $F_{40\%}$. This decision of the Models Working Group was not supported by the SARC Panel.

The specific comment of the SARC Panel on this topic was that they "do not suggest that F40% is necessarily the best proxy to use, rather there has yet to be compelling reasons to abandon it". The primary reasons given by the Panel members for their conclusion need to be examined carefully.

i) Panelist Bell reasonably comments on the apparent inconsistency of the Models Working Group in advocating an assessment starting in 1982, yet arguing for an F35%-based reference point using an assessment starting in 1970: "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?" However there is surprisingly nothing in the Panel's report that indicates that they were made aware of the circumstances that led to the necessarily convoluted approach the meeting took. The meeting worked in a linear fashion, and first agreed to a best assessment (starting in 1982); that assessment gave no strong indication of any trend in recruitment with spawning biomass. However, on the penultimate day of the meeting when the matter of reference points came under discussion, it was realised that when the assessment was started in 1970, there WAS a strong indication of such a trend, which did need to be taken into account in projections (and would fail to be so if advice was based on the assessment starting in 1982 only). At that late stage of the meeting, there was however insufficient time to re-open the matter of the selection of the best assessment (in my view had there then been sufficient time the selection of an assessment starting in 1982 would have needed to be changed). In these circumstances, it would have been quite incorrect, and inconsistent with the best available science, for the Working Group NOT to have attempted nevertheless to take this feature of the results into account in its reference point recommendations. The Panel's comments have to be interpreted in the context that they give no indication, in making their otherwise appropriate comments about (apparent) inconsistency, that they were at all aware of these circumstances (Bell's quoted comment above especially refers).

ii) Panelist Patterson adds that in respect of the stock-recruitment curve fitted by the Working Group to the assessment starting in 1970 that this "did not fit at all well to the data (Figure A 154 of assessment report)". However inspection of this Figure shows immediately that Patterson has been misled by an optical illusion caused by the choice of the horizontal axis scale for this graph. This scale was

evidently and reasonably chosen to allow the implied pristine spawning stock abundance for the resource to be shown, but this had the consequence of compressing the range of spawning biomasses in the assessment to a very narrow window so that a poor fit to the recruitment estimates seems evident. However the correct approach to identify model mis-specification is to inspect a plot of the residuals for the model fit in question. This plot is shown in the following Figure A 155, which gives no indication at all of any mis-specification – this panelist's key assertion is thus incorrect.

iii) Panelist Trzcinski indicates that he has serious concerns about the use of the MSY proxies, and inter alia suggests that "the work by Butterworth and Rademeyer (2008), briefly alluded to on page 41 of the working paper, suggests to me that a deeper discussion about the advantages and disadvantages of incorporating older data and model starting assumptions needs to occur".

In summary then, the comments of two of the panelists either exhibit lack of information about the circumstances associated with the Working Group's decisions or are in error, while the third supports consideration of the very aspects that the further work by Butterworth and Rademeyer (2012) addresses, including in particular using data from before 1982 in the assessment.

With respect to the Butterworth and Rademeyer work using a Ricker-style stock-recruit fit their model had substantial diagnostic problems, most notably a very strong retrospective pattern which can be problematic for determining stock abundance and making catch advice.

The problems of this retrospective pattern were indeed acknowledged in Butterworth and Rademeyer (2011), which is why this issue is addressed further in Butterworth and Rademeyer (2012). Note first Fig. 17 of the latter paper which shows effectively no retrospective pattern in the estimates of B_{MSY} obtained when estimating a Ricker stock-recruitment curve (this is linked to estimation of the relationship in that paper external rather than internal to the assessment). Nevertheless, the authors' preference (see paragraph spanning pgs 12-13 of the latter paper) is to move rather to the Beverton-Holt –adjusted formulation of this relationship, which is both AIC justified and greatly lessens any (inappropriate) dependence of the behaviour of the relationship at larger spawning biomasses on recruitment estimates at low spawning biomass (as led to the retrospective pattern reported in Butterworth and Rademeyer(2011)).

References

Butterworth, D S and Rademeyer, R A. 2011. Applications of Statistical Catch-at-Age assessment methodology to Gulf of Maine cod. Paper presented to October 2011 meeting of NMFS Models Working Group.

Butterworth, D S and Rademeyer, R A. 2012. An Investigation of Differences Amongst SCAA and ASAP Assessment (including Reference Point) Estimates for Gulf of Maine Cod. Paper tabled at the January 2011 meeting of the NEFMC's SSC.