SOME RESPONSES TO THE COMMENTS BELOW

Doug Butterworth

Comments are inserted in red text

Review of SCAA papers, by Noel Cadigan.

Some of my specific comments below are not specific to the SCAA models, and are generic to any assessment model for this stock.

On the whole, this SCAA is a highly parameterized model and there is insufficient data to reliably estimate some parameters. It is desirable for an assessment model to include uncertainty about fishery selectivity, natural mortality, and catches; however, there is a narrow region of compromise between model simplicity and model reality that results in good estimates, and I am unconvinced that the SCAA is in that region.

One must take care here in translating frequentist concepts to an approach which is more akin to Bayesian philosophy in its formulation. Though there are indeed many parameters, nearly all are linked to often relatively informative priors (penalties in the frequentist paradigm). The SCAA approach recognizes uncertainty by introducing these priors; in contrast the conventional frequentist assessment paradigm underestimates their contribution to uncertainty by conditioning on fixed values for quantities whose values are not that well known. The primary determinant of the number of parameters is an aim to eliminate residual patterning (see "Aspects of SCAA" document).

The SCAA does not seem to be "converged" in the ADAPT or XSA sense. The sensitivity run with high recruitment residual variability illustrates this. I would not use it to give absolute biomass estimates for management decisions, and I suggest it be used to provide stock trend information. The XSA I suspect is converged and less sensitive in scale to its model configuration. Nonetheless, the absolute biomass levels that XSA converges to may still be biased if M, catches, or selectivity assumptions are wrong. This suggests that XSA should also only be used to provide trend information.

Certainly the trend information is the more reliable, and the CVs for the SCAA estimates of abundance are high. Is the "convergence" attributed to XSA misleading? It is an inevitability of the structure of VPA linked to high asymptotic fishing mortality, which may not apply in circumstances (clearly at play here) of dome-shaped selectivity. Thus does this "convergence" indicate reliability, or rather a false sense of accuracy/precision occasioned by assumptions within the VPA which are not met in practice (see Rebecca and my contribution related to Gulf of Maine cod which shows how a small change in assumptions concerning the relationship between the fishing mortalities on the two oldest age groups can have a marked impact on the scale of the biomass outputs (Butterworth and Rademeyer, 2009))? Should the question rather be turned around to ask: what

particular assumptions in the XSA see it spanning a narrower range of biomasses than do the SCAA results, and is there any strong basis for these assumptions?

SSB should not be used to compare models for this stock. The mature component of GHal is apparently not well sampled by the surveys (or the fishery) and not well understood. The cryptic biomass component is speculative and not of direct interest in any event. I suggest trends in exploitable biomass (sum of biomass at age · selectivity) should be the main stock size metric.

The comment about speculative aspects of the cryptic biomass is well made, and the exploitable biomass is indeed an important metric. But the spawning biomass cannot be "forgotten", even though most spawning fish are not available to the fishery, because these are the basis for future recruitment. Even though the SCAA assessments indicate a preference for higher steepness which suggests recruitment overfishing may not be that great a concern, the data are not very informative on this parameter so that estimated spawning biomass trends should not be overlooked. It should be noted that there are many parallels in this assessment to that for Southern Bluefin Tuna, and the CCSBT Scientific Committee, which includes a semi-permanent high level independent international review panel, continues to use spawning biomass as the benchmark in a fishery which also has strong dome-shaped selectivity with most fishing on immature fish.

Comments on working papers

WP1: Initial Applications of Statistical Catch-at-Age Assessment Methodology to the Greenland Halibut Resource, by Butterworth and Rademeyer

Major

1. In B.2.5 – B.2.7 the authors have not included the "variance penalty" part of the likelihood because it was based on "input" or fixed variance parameters and would not affect estimation of other parameters. However, omitting this part of the likelihood means that total likelihoods for different values of these input variance parameters are not comparable, and the authors have compared them (see Table 1).

This criticism is justified. The rationale for choice of the penalty function (specifically the σ_{Ω} value) lies rather in reducing the extent of patterning in the residuals (see "On Some Aspects of SCAA" document (Butterworth *et al.*, 2009).

2. Are the bubble sizes the same in Figure 7. If so then the fit does not look that much better. The loglikelihoods for these fits are not comparable (see comment 1 above).

The typical sizes of the residuals is measured by the σ_{comCAA} parameter, which decreases with the introduction of greater selectivity variability to provide an improved fit (see Table 1), though comparability is indeed problematic for other reasons.

3. Figure 8 suggests that the SCAA does not have the convergence property that XSA usually has. We should not automatically accept that the converged part of an XSA is an accurate and precise estimate of stock size. The SCAA seems more useful for estimating trends in stock size rather than absolute stock size.

See responses above. Note also that the conflict between the overall survey index trends and the survey proportions at age data is playing a role here.

4. The catch age composition likelihood component seems ad hoc (B21). I don't see how it achieves the objective: "undue importance is not attached to data based upon a few samples only". There could be many samples that all consistently show that an age was not present in the catch, and yet B21 would give this information low weight. I can't recommend a specific likelihood alternative, but some are provided in Hrafnkelsson and Stefánsson (2004).

The ultimate test for what is in essence a relative weighting issue is whether the resultant residuals are reasonably homoscedastic, and those obtained are not too violently in conflict with this. Note that where proportions are small, ages are lumped into plus and minus groups.

5. Clearly assumptions about fishery and survey selectivity have great impact on biomass estimates, particularly SSB and 10+ biomass. I am not an expert for this stock and I find it difficult to understand how the fishery can differentially select older ages like some of the results in Figure 9 suggest. Unless there is strong evidence to the contrary, I think it is more pragmatic to assume selectivity is constant for older ages (e.g. 10+), similar to the XSA. **In my opinion this is best practice**.

Remember that selectivity in these models is a proxy for availability as well as a reflection of gear effects. At the gear level only, there is the mechanism of larger fish being faster swimmers and hence more easily able to avoid the net. In terms of distribution, and hence availability, different ages of fish can be located in different places, and the fishery acts to optimize value, which may result in avoidance of areas where the larger fish are present but with lower overall densities. For halibut, there is a relationship between size and depth, and both survey and catch selectivities are clearly dome shaped – why should this dome shape suddenly switch to being flat as age increases? Though asymptotically flat selectivity may be **common** practice for most North Atlantic assessments, I would strongly contest that this should be seen as **best** practice. The most common defense offered for it is "being precautionary" as fully dome shaped rather than flat selectivity generally leads to larger biomass estimates. There are many cases where the dome shape must be operative, for example for South African hake and for Patagonian toothfish, subsequent initiation of longline and pot fishing operations respectively revealed large fish which asymptotically flat selectivity assessments would have indicated as fished out long beforehand. There are also other cases where AIC considerations show very strongly that domed behaviour should be accorded far higher weighting on a statistical grounds than flat selectivity which reflects model misspecification.

6. I expect estimates of SSB for this stock to be more uncertain than normal, given that the fishery and the surveys do not catch them. A more useful "currency" to describe stock trends is exploitable biomass. I think assessments of this stock should put SSB in the "background" more, and focus on exploitable biomass.

See comments above.

7. There are a number of fixed parameters (i.e. h, θ, φ) and the sensitivity of key results to the values assumed for these parameters should be explored.

Indeed, and this is reported in further work. The impact of θ and φ tends to be transient, so does not greatly affect estimates of current status.

8. M=0.2 does not seem reasonable for all ages 1-14+. One would expect M to decrease with age. If this is the case, then not accounting for this will introduce spurious signals in selectivity and catchability patterns.

Certainly the consequences of different age-dependent *M* formulations should be explored as sensitivity tests, but note that arguments can also be made for **increasing** natural mortality at older ages through senescence.

Minor

1. I presume that someone (e.g. assessment lead) has checked the inputs in Appendix A. There are some anomalous maturity values for the 1980 cohort (e.g. age 11 in 1991, Table A4).

We simply used data provided to us. This clearly is a necessary task for those responsible for it.

2. Not clear if R_{start} in B12 is a estimated parameter or not. Is it R_o in B4?

Apologies for the confusion introduced – they are the same and the value follows from that estimated for *K* and other input or estimated parameters. Equation B12 used the alternative terminology to emphasise initialization.

3. I would expect that CPUE indices could also have autocorrelated errors, but the model seems to only consider autocorrelation for survey indices.

Indeed, and this is pursued in subsequent analyses.

4. B.4 seems ad hoc, but I appreciate that it must be difficult to reliably estimate all the parameters in this model. Perhaps it would have been better to use a flexible parametric model for S_a (e.g. double logistic).

Experience has taught us to rather err on the side of over-parametrization, reducing later only if there is clear patterning (see our recently published paper in ICES JMS on the assessment of Gulf of Maine cod, where we found inappropriate underparametrization of the selectivity function could lead to spuriously high estimates of precision (Butterworth and Rademeyer, 2008)).

WP2: Further Applications of Statistical Catch-at-Age Assessment Methodology to the 2J3K-O Greenland Halibut Resource, by Butterworth and Rademeyer

Major

1. At the beginning of the Results and Discussion section the authors state "The best fit to the data obtained in the initial analyses of Butterworth and Rademeyer (2009a) was their variant 4". I am unsure if this is a valid conclusion because there seems to be a problem with the loglikelihood (i.e. omitted variance penalties).

Correct, as acknowledged above (see further commentary there).

2. Similar to my comments for WP1, I think exploitable biomass is the better metric to use when comparing different formulations.

See comments above.

3. The retrospective runs in Figure 6 illustrate that the SCAA does not converge like an XSA, but XSA convergence may be misleading. However, the retrospective trends seem quite stable. This could be quantified by showing $B^{sp}{}_{2003} / K$ in Table 2.
4. The Ω_R=0.35 (Fig. 8) demonstrates the uncoverged property of the SCAA. With this specification recruitments are estimated more freely back in time, and the scale of stock size estimates becomes confounded with survey catchabilities. The model basically only provided relative estimates of stock size.

See related comments above.

5. I was surprised that in Table 5 $B^{sp}{}_{2008} / K$ is lower for the single index runs than for B2. A guess about why this happens is that the age composition information from the surveys and catches are more influential in the single index runs, and the age composition information may somehow suggest a more pessimistic current stock status. This may be part of the reason why XSA, which treats catch and their age compositions as exact, is more pessimistic than SCAA.

I'd broadly agree. This arises from the CAA – overall survey index conflict, where the former favours lower and the latter higher abundance. With all the surveys taken into

account, survey information gets greater weight relative to CAA, and moves the estimated abundance above that for any of the surveys considered in isolation.

6. Runs with "large" σ_c and σ_m are worrying to me. They imply a much different and more pessimistic stock status. The range of M's in the bottom panel of Figure 10 do not look overly extreme and are perhaps as plausible as M=0.2. Insufficient results are presented to allow a full evaluation of these runs, but on the surface they suggest the stock could be at very low levels, even though Can RV Fall survey has been increasing in recent years. This is all puzzling to me and I feel it should be sorted out. This type of sensitivity also applies to XSA, ADAPT, etc.

I suspect what underlies this is that for recent years where the younger cohorts have not been sampled that often, admitting this level of variability allows for inappropriately exact matches to data.

7. Figure 12 illustrates the problems with going back to far in time with a model. Although tables of results are not presented, I suspect that MSY quantities may be sensitive to the assumed values for ϕ and θ.

MSY is not greatly affected, ranging from 42.6 to 46.1 thousand tons, compared to 43.2 thousand tons for B2.

Minor

1. Need to clarify what the B^{sp}_{2008} and B^{sp}_{2008}/K rows mean in Table 2, particularly for the retrospective runs.

This is clarified in the existing Table 2 caption.

2. On Pg. 2, I did not understand the text "For the commercial selectivity-at-age, with $\sigma_\Omega = 2$, for baseline B2 σ_Ω (output) is 0.58".

The first refers to the penalty function (or prior), and the second to the standard deviations of these deviations in the actual model fit.

3. Figure 10 is too busy. I was interested if M decreased with age but I could not really tell from this figure.
4. It seems some of the runs in Figure 11 are mislabeled.

Apologies, variants "10 – sigM=0.2" and "11 – sigM=0.4" have been swoped.

References

Butterworth DS, Ianelli J, Parma A and Punt AE. 2009. On some aspects of SCAA.

Butterworth DS and Radameyer RA. 2008. Statistical catch-at-age analysis vs. ADAPT-VPA: the case of Gulf of Maine cod. ICES Journal of Marine Science, 65.

Butterworth DS and Rademeyer RA. On the precision of absolute abundance as estimated by VPA.

Hrafnkelsson, B, and Stefánsson, G. 2004. A model for categorical length data from groundfish surveys. Can. J. Fish. Aquat. Sci. 61: 1135–1142.