UM Independent System for Peer Reviews
Consultant Report on:

32nd Stock Assessment Review Committee for American Plaice, Scallop, Silver Hake and Maine Haddock Assessments

27 November-1st December 2000, Woods Hole, MA

Paul A. Medley
Sunny View
Jack Hole
Alne, YO61 1RT
UK
paul.medley@virgin.net

Contents

1 OVERVIEW

The following are some personal observations on future improvements in the assessments and are not necessarily shared by the other panel members. This report does not repeat issues raised in the main SARC report, which needs to be consulted for the committee's consensus views on these assessments.

The use of decision rules by the management council is an excellent approach. In general, they make the assessment clearer in terms of results. Although concepts such as fishing mortality are relatively complicated for non-scientists, the consistent approach means managers only have to learn a limited number of fixed technical terms to understand assessment results. Despite this, there appear to be a number of problems with respect to implementation which include, most notably, how the decision rules should be updated and how relative degrees of uncertainty are represented to managers.

The fisheries of the region possess a considerable database. These data have produced a good understanding of the stocks under analysis, but did not necessarily help in quantitative modelling of the stocks. These data may prove useful in developing more complex models for stock assessment, but this would require a move away from standard methods.

Because of the number of stock assessments to be reviewed, there was little time to study assessments in depth. The review was limited to only those methods and results presented. While some additional diagnostics were provided during the meeting, these were limited and essentially no additional analyses could be carried out. Working groups were conducted for these species separately.

2 INDICES OF ABUNDANCE

2.1 Models Used

Models presented were limited to VPA and surplus production models. Better results might be obtained with other models. There are two approaches I feel may improve assessments:

- Building more flexible models. The VPA population model has a rigid structure, which relies on good quality data. Where good catch-at-age data exist, it is probably the best approach available. However, there are always questions over catch and survey data and a more flexible approach (such as statistical catch-at-age models) not only may fit the data better, but also may provide insight into underlying problems. Relaxation of some of the VPA assumptions towards statistical catch-at-age models would still fall within the ADAPT approach.
- Complex models may be used to define hypotheses of what is actually occurring with the stock. In many cases, background research suggested shifts in stock location and different growth rates between areas. These

hypotheses developed in the form of models could be used to simulate data and identify the best assessment approach to use.

The simplest age structure model, the recruitment index model, could be used as a start point to explore alternative approaches for many of the stock assessments. This makes partial use of the age structured survey data, but minimises the assumptions. The survey data would be used to provide an index for only two age classes: recruits and a single plus-group. The model only requires three parameters, but does not rely on a particular stock relationship as surplus production models do. Furthermore, it is relatively easy to expand a recruitment index model with more age classes as information becomes available, and therefore it would seem to be a good choice of model to explore alternative approaches.

The assessments in general made little use of the time series structure of the models. In some cases, the residuals may be autocorrelated (based on a subjective assessment of the plots as ACFs were not provided). This is not a very large problem in itself, but may indicate where the population model was failing to describe the dynamics well. Some statistical techniques may help deal with this problem. For example, the squared differences between Fs estimated in adjacent years can be included in the log-likelihood. This prevents fishing mortalities from changing wildly from year to year and may improve the statistical behaviour of the model while avoiding separable VPA which may be too restrictive. Other time series (e.g. autoregressive) terms may deal with shifting catchability, while not requiring the estimation of independent catchability parameters in each year.

2.2 Estimators of F

In general, more could be made of the available data through improvements in the estimators. Whereas estimates of biomass or population size depend directly on the survey data, estimates of fishing mortality (F) are derived from calculations. The estimates based directly on the surveys depend on the sampling, and while they may not be precise, they are probably relatively unbiased. They might be improved if related to the catch in some way, as was suggested for scallop assessment. However, the estimates may have been further improved by constraining fishing mortality to be positive, and perhaps improving the method's statistical foundation (those given were essentially least-squares estimators). A possible Bayesian approach is set out below.

A relatively uninformative prior can be used, except that fishing cannot be negative (a negative F would imply restocking). Assuming survey estimates of population size (P_t) are drawn from a log-normal with a known variance, the probability density function would be:

$$\ln(P_t)-\ln(P_{t+1})-M \sim \mathrm{N}(\mu,\sigma)$$

The posterior would be a truncated normal:

$$F \mid \sigma,P_t,P_{t+1},M \sim \frac{\mathrm{N}(\ln(P_t)-\ln(P_{t+1})-M,\sigma)}{\int_{F=0}^{\infty}\mathrm{N}(\ln(P_t)-\ln(P_{t+1})-M,\sigma)dF}\ ,\ F\geq 0$$

Some small improvement might be gained where the variance was also considered to be unknown, so the likelihood would be dependent on the raw survey data. However, as long as the the data dominates any prior, as might be expected with a survey, the added complication would not benefit the estimation much.

Similarly, a time series approach might be used to combine catch biomass with these estimates if a likelihood for the catch biomass (C_t / B_t) is defined. For example, the mean of some likelihood might be defined as:

$$\overline{F_t} = \frac{C_t}{B_t} - \frac{C_{t-1}}{B_{t-1}} + \overline{F_{t-1}}$$

This would remove any fixed bias in the catch biomass, and smooth the estimate.

A Bayesian approach is more complicated than the averaging method presented. However, it is relatively simple to obtain useful statistics numerically, as the estimates would only require one dimensional integration. It is not clear whether this or other approaches might improve estimates except in the cases where the populations are low, as this approach prevents negative F estimates. Other advantages of different methods, if any, could be identified through simulation.

2.3 Silver Hake Assessment

Special consideration is given to the Bayesian surplus production model at the request of the assessment scientist. The SARC panel did not consider the assessment model reliable for different reasons. The following give my reasons for coming to this conclusion. Note that this is not a criticism of the Bayesian modelling approach, which is recommended wherever it can be applied and was well conducted in this case.

- There is a lack of contrast in the data since 1980. The surplus production model is driven by the high foreign catches in the period 1960-1980. Reported catches may be unreliable most likely under-reported, which would underestimate the potential yield). In addition, oceanographic conditions may have changed making the stock less productive. Considerable information was presented indicating a change in oceanography and stock distribution, but this was not taken into account in the assessment.
- The change in population size is not clearly represented in the abundance indices. Only one of the two indices showed any decline in abundance in response to catches in the period 1960-1980. The decline was small relative to the overall variance of the indices. While the observations were compatible with the model, they did not provide evidence for it, as the indices could be interpreted in other ways. For example, the indices could be affected by the distribution of the stock and poorly related to overall stock size. This is particularly a problem for surplus production models where there is no clear depletion and recovery that can be related to catches and abundance indices.
- There were problems indicated by the change in age structure, which the surplus production model could not address. While an adequate reason

was given as to why older fish were not present in the catches, there was no direct evidence for these hypotheses. The impact of these hypotheses was not tested on the surplus production model, in which they might be expected to change the catchability over time. A discrete catchability change in the time series could not be estimated.

- The structural errors were not tested in the assessment, although they are probably the most important. I believe the different stock status given by the two models presented (VPA and surplus production) is written into their structure, and hence is an assumption. VPA explained the decline in older ages as due to increasing fishing mortality, therefore the stock is overfished. The surplus production model is unstable at low population levels, so maintaining an overfished state for twenty years would be interpreted as very unlikely, particularly with model process error. It therefore favours the low fishing mortality. In reality, there is not enough information to choose between the hypotheses.
- The results between the VPA and surplus production model were diametrically opposed. Advice according to the precautionary approach would tend to put the onus of proof on the model implying the stock is not overfished (i.e. the surplus production model). In this case, the surplus production model suggested catches raised an order of magnitude higher than those currently taken could be sustained. Given the model uncertainties, this extreme advice would not be justified, although, some testing of the model might be (e.g. increasing catches to check whether observed abundance indices follow predicted patterns).

3 GENERAL COMMENTS

3.1 SARC Review System

Overall, the review system works well. The aims of the review are clear and reporting summaries seem easy to follow. If it can be criticised, the output may be oversimplified. However, this was necessary so that the management councils could assimilate the information rapidly and with minimum explanation. Hence, this problem seems to rest with the management councils rather than the SARC.

There was a tendency to use the SARC to review scientific work not strictly leading to stock assessment. While I had no problem with this in principle, in practice there was too little time to review any of the research programmes adequately. This led to vague statements of support for some scientific approaches, but was not entirely satisfactory. It would be worthwhile carefully scoping what is presented and needs to be reviewed so that the time in the meeting can be spent with maximum efficiency.

3.2 Decision Rules and Reference Points System

Decision rules must have a clear meaning and there must be a clear demarcation between management decision making and scientific advice. This was, overall, achieved. However, there was some confusion over legal implications of the decision rules and therefore the advice given. Legal issues

should be separate from scientific advice where possible, as the legal and scientific positions on a stock may not correspond, and this needs to be brought to the attention of managers and legislators.

Changing reference points appears to be a problem, as it was claimed frequent changes would undermine management confidence in assessments. Although it might be argued this communicates genuine uncertainty in the assessment, it makes it difficult for management to assess results. The unwillingness to change reference points and assessment methods generated understandable frustration among the assessment scientists.

The same assessment method used to define the stock status should be used to re-estimate the reference points. Incompatibility between methods increases errors. Very often, parameter estimates are correlated, so their absolute values are not known with certainty. However, when they are used to calculate other statistics, these correlations may cancel and the results may be much more precise. For example, the estimate of B_{MSY} and current biomass will be correlated if they are derived using a surplus production model.

One way to reduce changes is to define reference points as fixed dimensionless values. For example, it is common practice to define SSB as proportions of the unexploited, which can be estimated more precisely than either current or unexploited SSB separately. Performance variables can often be defined where the target or limit is consistently 1.00 (or 100%).

The decision rule system bears comparison with the US West Coast approach (STAR). The STAR stock assessment scientists were required to produce decision tables where the management council had not specified the decision to be made, but it had to be inferred from the management objectives. In contrast, the SARC panel had a very clear decision rule to be applied, but was not required to produce decision tables. The best of both approaches should be combined, with clear decision rules and decision tables required. For the US East Coast, this would require explicitly dealing with risks in the decision rules, which is not currently done.

3.3 Software

The model diagnostics that were presented were limited to those provided by the available software. While these were, on the whole, adequate in identifying whether the model fitted the data, they were not adequate in identifying possible improvements in models. Diagnostics such as parameter correlation matrices, observed and predicted data (i.e. survey indices) with linear regressions and R^2 values should be standard. Other statistics can be obtained indicating influential data points and residual autocorrelations. For models with abundance indices by age, examination of residuals along cohorts is often a better indicator of problems in the model than combining across cohorts into years.

4 TERMS OF REFERENCE

4.1 General

The Stock Assessment Review Committee (SARC) is a formal, one-week long meeting of a group of stock assessment experts who serve as a peer-review panel for several tabled stock assessments. It is part of the overall Northeast Stock Assessment Workshop (SAW) process that also includes peer assessment development (SAW Working Groups), public presentations, and document publication within a cycle that lasts six months. The panel consists of some 12-15 assessment scientists which include 4 scientists from the NEFSC; a scientist from the Northeast Regional office, staff from the NEFMC, MAFMC, and ASMFC with additional panelists from state fisheries agencies, academia (US and Canada), and other federal research institutions (US and Canada).

Designee will serve as a panelist on the 32nd Stock Assessment Review Committee panel. The panel will convene at the NEFSC in Woods Hole the week of 27 November (27 November - 1 December, 2000) and review assessments for sea scallop, silver hake, Gulf of Maine haddock and American plaice.

Specific

(1) Prior to the meeting: become familiar with the working papers produced by the SAW Working Groups (total number not final; there will be at least one per stock);

(2) During the meeting: participate, as a peer, in panel discussions on assessment validity, results, recommendations, and conclusions. Participate in the formulation of the draft SARC Advisory Report;

(3) Review the final Draft Advisory Report and Consensus Summary Report.

(4) No later than January 8, 2001, submit a written report of findings, analysis, and conclusions. The report should be addressed to the “UM Independent System for Peer Reviews, “ and sent to David Die, UM/RSMAS, 4600 Rickenbacker Causeway, Miami, FL 33149 (or via email to ddie@rsmas.miami.edu).

A Workshop Participant’s duties will occupy a total of 7-10 workdays; a day or two prior to the meeting for document review; the week long meeting; and a day or two following the meeting to ensure that the final documents are consistent with the SARC’S recommendations and advice.

No consensus opinion between two CIE reviewers is sought.