SOME THOUGHTS ON THE SISAM/WCFSA ASSESSMENT METHOD EVALUATION PROCESS

Doug Butterworth

Background

1) WCSAM intends (possibly just the first step in) a comparison of different assessment methods by applying different methods to a number of different stocks with different characteristics. For a particular stock, this would be the method currently applied, and some other methods (typically two perhaps), where these alternative methods will likely differ from stock to stock.

2) The proposal of stocks for this process has been encouraged from both inside and outside ICES. Although there is no obligation to proceed with all +-14 stocks currently under consideration, unless there are strong reasons to exclude some of these, there should be an attempt to at least to achieve 1) above for each stock eventually included, and some form of 3) below for at least some of them.

3) Because 1) would provide only limited comparisons, simulated pseudo-data sets conditioned either exactly or closely on the data and assessment output for each assessment of each stock should also be created, so that each assessment method can be applied to all these sets.

4) Two forms (not mutually exclusive) of analyses of such simulated datasets have been proposed:

i) a fairly limited (at this stage) comparison of key management-related assessment outputs (e.g. current spawning biomass and replacement yield) to inform on an assessment (estimation) method's robustness; and

ii) an in-depth examination using simulated datasets of a few "grand questions" (currently under discussion, and including, e.g. to inform on the utility of catch-at-age data in different circumstances) - quite possibly this would be conducted at this initial stage for only, say, two of the stocks, each chosen because of their potential to inform on a specific grand question.

5) The ICES Methods WG meeting provides an opportunity to aid the development of this initiative, particularly given that the meeting will be having general discussions on the most appropriate ways to set up simulations of this type.

6) Given the limited time available until WCSAM, decisions regarding generation of these simulated datasets need to be finalised very shortly.

Some suggested ways forward

6) By the time the Methods WG meeting starts, there must be available for each stock currently proposed for consideration:

i) agreement by the scientist currently responsible for assessment of that resource to participate in the process;

ii) documentation of the data inputs to that assessment; and

iii) a specification of the current assessment method together with the results which it provides.

7) The Methods WG meeting should finalise choices on:

a) the stocks to be included in the exercise;
b) the grand questions to be investigated and the associated stocks to be considered for each of these;
c) performance measures for broader comparisons for the full set of stocks; and
d) the specifications for simulated datasets required.

Approaches to generate the simulated datasets

8) POPSIM has been proposed as the basis to generate some/all of the pseudo-data sets required. This has several potential advantages: e.g. pre-existing software compatible in I/O terms with a number of assessment packages, some flexibility, and funding available for a project which would dovetail with this initiative.

9) There is a debate concerning the range of error types to be considered for inclusion in generating the pseudo-datasets (which might differ for the grand question investigations and the broader comparison studies). The essence of this involves whether or not "process errors" are to be included (where process errors in this context are effects, such as varying stock-recruitment residuals, which modify the dynamics from that estimated by the assessment concerned, whereas in contrast observation errors impact only data such as survey and CPUE indices of abundance and catch-at-age data, leaving the dynamics unchanged).

a) On the one hand process error inclusion enables a more comprehensive investigation.
b) On the other hand, specification and conduct of the process of pseudo dataset generation, and specification of statistics to compare performance, then become potentially considerably more complex - e.g. on what are these errors conditioned - invariant past catch series across the datasets? - how can that be done without, e.g., generation of joint Bayesian posteriors from each of the assessment methods considered? Furthermore, automated assessment procedures to estimate e.g. S/R relationships may run into convergence problems.

10) From the information circulated, POPSIM seems at present to be based on an annual time-step and length-based dynamics linked to the von Bertalanffy growth equation. While this should certainly be appropriate for some of the stocks (with their associated datasets) currently on the list for consideration, a number of questions (not necessarily needing immediate answers) arise:

a) Is POPSIM structured so as to be able to deal with all these stocks (e.g. for the short-lived SA anchovy, shorter time steps are required and the ageing information ais not in the form of an age-length key)?
b) Probably there needs to be consideration of the information planned to be available by the time of the Methods WG meeting to determine for which of these stocks the current version of POPSIM may not be suitable - does POPSIM then get generalised, or would pseudo dataset generation for such stocks require separate coding initiatives (by whom - the scientist currently responsible for that assessment)?
c) If the current assessment of a stock is closely but not exactly compatible with POPSIM's underlying structure (e.g. age-based rather than length-based

dynamics), how are the error variance components within the POPSIM structure to be tuned to replicate the error variances evident from the fit of the assessment model to the actual data?

d) What process is to be used for assessments implemented under the Bayesian paradigm?