Understanding stock assessment modeling through simulation studies

Athol R. Whitten[1*] Sean C. Anderson[2] Curry Cunningham[1] Felipe Hurtado Ferro[1] Kelli Johnson[1] Roberto Licandeo[1] Carey McGilliard[1] Cole Monnahan[1] Melissa Muradian[1] Kotaro Ono[1] Cody Szuwalksi[1] Juan Valero[4] Katyana Vertpre[1]

[1]School of Aquatic and Fishery Sciences, University of Washington, Box 355020, Seattle, WA 98195-5020, USA [2]Department of Biological Sciences, Simon Fraser University, Burnaby BC, V5A 1S6, Canada [3]Resource Ecology and Fisheries Management Division, Alaska Fisheries Science Center, National Oceanic and Atmospheric Administration, National Marine Fisheries Service, 7600 Sand Point Way NE, Seattle, WA 98115, USA [4]Center for the Advancement of Population Assessment Methodology, Scripps Institution of Oceanography, La Jolla, USA

*Corresponding author

Abstract

Stock assessment simulation studies are critical to evaluating stock assessment methods and understanding their strengths and weaknesses. This paper briefly summarises past stock assessment simulation studies and describes steps necessary to conduct efficient and effective simulations using contemporay fisheries stock assessment methods. This paper also describes key outcomes of a large-scale stock assessment simulation project and discusses important lessons learned from it.

Introduction

Stock assessment simulation is key to evaluating stock assessment methods and understanding their strengths and weaknesses. Simulation is important because it enables modellers to test assessment models on known truths, to examine the consequences of alternative plausible truths, and to match (or mismatch) truths and assessment model assumptions.

Stock assessment studies in practice

Key components of effective simulation studies

An effective stock assessment simulation requires four key components (Fig. 1):

1. A *conditioning model* is used to ground a stock assessment simulation to some plausible reality. A stock assessment simulation may be conditioned on a specific region and/or fish stock or on some generic representation of a fishery based on expert opinion or the published literature.

2. An *operating model* represents a range of conditions that underly the believed true state of nature. An operating model might specify, for example, that natural mortality (M) varies through time with some random walk, or that it is constant through time.

3. *Pseudo-data* is sampled from an operating model with simulated observaton error in a manner that is representative of the fishery of interest.

4. A stock assessment *estimation model* is fit to the pseudo-data. This determines the ability of the methods of interest to estimate the paramters specified in the operating model and to assess the scenario or 'status' of the simulated stock.

Steps 2 to 4 are usually repeated across iterations, adding re-sampled process error and observation error each iteration. Blocks of iterations are repeated across multiple scenarios, with each scenario representing some different combination of conditioning, operating, sampling, and estimation models. This process differs from a management strategy evaluation (MSE) in that MSE 'closes the loop' — introducing management decision rules about how to act on stock assessment output that affect subsequent realistations of operating model years. Stock assessment simulation on the other hand is intended to examine the mechanics and performance of stock assessment models themselves: It (insert insightful benefit of SAS modelling here).

We've arrived at the lessons we suggest in this paper after conducting a series of large-scale stock assessment simulation studies at the University of Washington. These studies are featured in this issue:

Johnson et al. (2013, this issue) evaluated the ability of Stock Synthesis (SS) to estimate key quantities when a known 'true' natural mortality (M) was age-specific or age-invariant, but time-varying. Stock assessment methods included models with age-invariant pre-specified M, age-invariant estimated M, and age-specific estimated M. Ono et al. (2013, this issue) analysed the ability of SS to estimate management metrics for different life-history types (demersal, long-lived pelagic, and short-lived pelagic) when the same quantity and quality of pseudo-data were used to inform assessment models. They also considered whether the frequency and duration of length- and age-composition data, or catch history, affect the bias or precision of estimates of management quantities for different life-history types. Hurtado Ferro et al. (2013, this issue) investigated factors which lead to retrospective patterns in SCAA models. Specifically, they tested how key biological and modelling factors can induce retrospective patterns for various life history types. They explored the potential effects of catch patterns, as well as model miss-specification from time-varying biological parameters, time-varying selectivity and catchability, and their interactions. In those cases where retrospective patterns were observed, they assessed the utility of including time-varying selectivity in the assessment as a means to correct them.

In this paper, we review a series of lessons that we learned through these studies about conducting large-scale, relevant, and rapid stock assessment simulation studies…

1. Choose widely-used and current assessment models

Stock assessment simulation studies are most relevant to research and management if they focus on the models and tools that are used in practice.
In our studies we chose to focus on SS, which is a widely used integrated-assessment modelling framework, now used in all assessments on the West Coast of the United States (REF). In other regions commonly used model frameworks will be different.

When researchers are interested in multiple modelling frameworks they may choose between (1) conducting a study where both the OM and EM are based on the same modelling framework to better understand issues related to model specification or (2) conducting a study where the OM and EM are based on different modelling frameworks to investigate the impact of model-choice uncertainty. Frequently model-choice uncertainty may dwarf other sources of uncertainty (REF).

2. Condition the operating model carefully

The relevance of a stock assessment simulation study depends on the system that the OM is conditioned to reflect. Researchers must choose between specific conditioning, in which the OM might be conditioned for specific species, region, or even stock and (2) generic conditioning in which the OM is conditioned on a general representation of a system. Both types of conditioning have their place in stock assessment simulation but each has its pros and cons.
Specific conditioning may result in findings that are relevant to a specific species, region, or stock and may therefore be more likely accepted than generically conditioned models. Generic conditioning, however, may result in findings that are applicable across a wider range of fisheries, but are less applicable to any one stock

3. Consider likely or important model misspecifications

Model misspecification refers to a mismatch between the truth (OM) and our assessment of that truth (EM). For example, an OM might specify time-varying M, whereas the EM might assume a constant M. Researchers might deem some forms of model misspecification more likely *a priori* and some types of model misspecification potentially more influential for stock assessment. It is these forms of model misspecification that researchers should focus on first.

4. Make your framework reproducible and transparent

To produce credible science, it is critical that stock assessment simulation studies are reproducible and transparent. To achieve these properties, simulation study can be written in code as opposed to by clicking on buttons in a graphical user interface (GUI). Further, researchers might consider writing their code in a formal package or library structure (e.g. Anderson et al. 2013 R package). This makes it easier for future users to understand the code and simulation structure and encourages the code authors to document their work. Researchers can introduce further transparency by developing their code in a version control system such as Git, which provides a history of all code modifications (e.g. https://github.com/seananderson/ss3sim). Finally, by controlling model runs through plaintext control files, researchers and other users can easily understand and reproduce the settings that created their simulations.

5. Make your framework flexible

A stock assessments simulation framework is most useful if it is flexible for the current users as well as for future users who may develop scientific questions with the framework that the original users never considered. One way researchers can make their framework flexible is to build their code around smaller functions that can be mixed as needed. Another key element is keeping the input and output files in formats that can be read and process multiple tools (e.g. comma-separated, or tab-delimited text files). Finally, researchers may consider splitting scenarios into different cases so that cases can need mixed and matched into scenarios. For example, a scenario might be comprised of a combination of cases for the M trajectory, F trajectory, and selectivity patterns.
If these cases are specified in individual text control files then they can be flexibly combined without duplicating case specification.

6. Check your models early and often

The complexity of conducting stock assessment simulations means that the chance of making mistakes is high, and our ability to make sense of complicated model output can be limited. Deterministic model checking is therefore vital. To check a model deterministically we can reduce or eliminate process and observation error and check for bias between the OM and EM models. This might mean running the stock assessment simulation with minimal stock-recruit deviations and minimal observation error on survey indices.

An important component to model checking is graphical model checking. Many complex problems are unlikely to be detected without graphical model checking (REFs, maybe Gelman). To facilitate model checking these graphics should be rapid and easy to produce. We found the visualization packages `manipulator`, `shiny`, and `ggplot2` to be helpful for this purpose.

7. Keep simulation runtime minimal

To discover problems with your simulation and to understand how your simulations are performing you need to run them repeatedly under a variety of conditions. In addition to using a fast computer and writing code carefully, you can minimize runtime by reducing the number of scenarios and iterations. To reduce scenarios, a researcher might consider creating a base-case model and investigating deviations from that model instead of a full factorial design. To reduce iterations, a researcher can inspect test runs with an increasing number of iterations to determine the minimum number of iterations for study conclusions to converge.

Other thoughts

keep folder structure as simple as possible; keep all output; write code so the simulations can be distributed across cores, computers, and researchers

Discussion and conclusions

- perhaps suggest the kinds of questions that this approach could answer and other steps forward
- ...?

References