Simple rules for conducting simulation studies of statistical fisheries stock assessment models

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: tel +1 206. . . ; fax +1 206. . . ; email: wittena@uw.edu

Abstract

Simulation studies are critical to evaluating stock assessment methods and understanding their strengths and weaknesses. This paper briefly summarizes current trends in stock assessment simulation and describes the key components and steps necessary to conduct efficient and effective simulations using contemporary fisheries stock assessment methods. The paper also describes key outcomes of a series of large-scale stock assessment simulation studies and outlines important lessons learned from them.

Keywords: Operating model; Simulation; ...

Introduction

Simulation is key to evaluating stock assessment models and understanding their strengths and weaknesses. It is alway necessarily make simplifying assumptions about fish stocks and fisheries when constructing and fitting stock assessment models, but the results of stock assessments can critically depend on those assumptions. Simulation is important because it enables testing when the truth is known so the consequences of alternative plausible truths can be evaluated along with the consequences regarding uncertainty about those truths. Thus though simulation, stock assessment models and the methods and assumptions that underpin them, can be critiqued, tested, and better understood.

The true dynamics and underlying properties of fish stocks can never be known. Consequently, fisheries models and management actions can only truly be tested through direct action and long periods of monitoring and evaluation. Simulation modeling provides means to investigate how fish stocks might respond to fishing pressure, and how well assessment methods estimate management quantities, without the obvious risks of implementing untested fishing practises, assessment methods, or management procedures. Simulation testing should be a prerequisite for all stock assessment methods before they are applied in practice (Hilborn and Walters, 1987), but this is rarely the case for the complex statistical models in use today. There is a proliferation of new stock assessment methods (e.g. XX), but they have seldom been subject to rigorous simulation testing. With continued increases in computing power however, and the ready availability of pre-packaged stock assessment modeling software, rapid and effective simulation studies are increasingly viable.

The results of simulation studies should, in principle, permit an understanding of the general behavior of methods and models, but the results depend on the structure, assumptions, and chosen parameter values of the models used to generate pseudo-data and estimate quantities of interest (the operating model). Simulation studies must be designed and implemented with careful consideration of the systems of interest and for the desired outcomes. They should address specific questions, and match the scale of observation, data availability, and modeling and management methodologies applied to the fishery or types of fisheries of interest.

Stock assessment simulation studies can be simple: testing direct-estimation methods on simulated data sets specific to the method being examined; intermediate: with a focus on methods that estimate multiple parameters and models that describe full population and fishing fleet dynamics; and extensive: through whole-of-system models that describe all components of a fishery, including the implementation of management strategies and harvest controls. TODO: REFS on MSE for example; Insert text here about how simulation studies of contemporary statistical stock assessment models have become more comment in recent times, and REFS to papers that have done this. Show that the models came a long time before the simulation studies, and so more simulation studies, and the methods and a general approach to doing them, are much needed!

In this paper, we review current trends in simulation studies using contemporary fisheries stock assessment models, and outline the necessary components for conducting simulation studies. We also devise a series of 'simple rules' and share key lessons learned whilst conducting a series of large-scale simulation studies. These studies were designed to answer questions posed by the organizers of the 2013 World Conference on Stock Assessment Methodology. They were conducted using contemporary methods and led to the development of an R package that enables rapid and reproducible simulation studies using readily-available pre-packaged stock assessment software. The package, together with the rules and lessons presented here (see Appendix A for a description of the package(, are intended to encourage and enable further research, testing of new stock assessment techniques, and increased scrutiny of commonly-used and sometimes unchallenged existing methods.

The current state-of-the-art

TODO: Insert brief summary of modeling trends, by providing examples from the literature: Include references to recent papers on stock-assessment simulation such as Piner et al. (2011), Lee et al. (2011), Maunder, Maunder and Deriso

Key components of stock assessment simulations

An effective stock assessment simulation study requires four key components (Fig. 1): TODO: Need references here!

1. A **conditioning model** grounds a stock assessment simulation in some plausible reality. A stock assessment simulation may be conditioned on data relating to a specific region and/or fish stock or on some generic representation of a fishery based on published literature, stock assessment reports, or expert opinion. When pertaining to specific situations, conditioning models can be fit to data in order to obtain parameter values representative of discrete systems. Parameter values can then be modified or certain aspects of a model altered to make simulation processes more general. In a fully-generalized sense, conditioning models need not be fit to data at all, but should at least be representative of plausible stocks or situations.

2. An **operating model** is used to represent the dynamics of a fish stock and/or fishery. It provides the true state of nature for the simulation. An operating model might be used to specify alternative forms of fish life history traits, for example, that fish natural mortality rates vary through time with some random walk, or that they are constant. It might also be used to specify alternative conditions of a fishery, for example, that fish of some particular age are caught by one fishing fleet but not another, i.e. that different fleets exhibit different fishing selectivity.

3. Simulated data, or **pseudo-data** are sampled from an operating model with simulated observation error in a manner that is representative of a fishing fleet or survey program of interest.

TODO: More on pseudo-data required?

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

TODO: More on estimation models required?

Steps 3 to 4 are usually repeated across iterations, with added re-sampled process error and observation error for each iteration. Blocks of iterations are repeated across multiple scenarios, with each scenario representing some different combination of conditioning, operating, sampling scheme, and estimation method. This process differs from a management strategy evaluations which 'close the loop': such studies explicitly model management decision rules about how to act on stock assessment outcomes that affect subsequent realizations of operating model years. Stock assessment simulation on the other hand is intended to examine the mechanics and performance of stock assessment models themselves.

Simple rules for effective stock assessment simulations

Stock assessment simulation studies should be conducted with three R's in mind: *realism*, *relevance*, and *reproducibility.* Realism means the operating model are reflective of actual fishery situations, and estimation methods that are aligned with, or are natural progressions of, current modeling trends. Relevance means that the simulations should be based on methods which are likely to be used in practice. For example, it is largely pointless to conduct simulations based on equilibrium production models (ref) because they are no longer used in practice. Simulation studies are most likely to affect future stock assessment practice when they are fully reproducible. TODO: Something from paper on Reproducible Research here, and a reference to that

With the three R's as a recurring theme, we offer the following 'simple rules' for conducting effective and efficient stock assessment simulation studies:

Choose widely-used, contemporary assessment models and methods

Simulation studies are most relevant to current research and management if they use and analyze models, methods, and tools that are published, commonly used in practice, and freely available. TODO: List some different options here and say why using these systems is beneficial [improved uptake, ease of review and understanding etc.]

When researchers are interested in multiple modeling frameworks they may choose between (1) conducting a study where both the operating and estimation models are based on the same modeling framework or (2) where they are based on different modeling frameworks. The first option enables better understanding of issues related to model specification, whilst the latter allows one to investigate the impacts of model-choice uncertainty. Frequently, model-choice uncertainty may dwarf other sources of uncertainty in stock assessment situations (REFS).

Condition the operating model carefully

The relevance of a stock assessment simulation study depends on the operating model is conditioned. It is necessary choose between *specific conditioning*, in which an operating model is conditioned on data pertaining to a specific species, region, or even stock, and *generic conditioning*, in which an operating model 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 for that situation 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.

Consider likely or important model misspecifications

Model misspecification refers to a mismatch between the assumptions of the operating model and the assumptions of the estimation method. For example, an operating model might specify time-varying natural mortality, whilst the estimation model estimates a constant natural mortality parameter (e.g. REFS). Some forms of model misspecification such as XX may be deemed to be more likely *a priori* and some types of model misspecification potentially more influential in terms of estimation performance. These forms of model misspecification should be considered and analyzed first.

Effective software development

It is critical that simulation studies are reproducible and that their methods are transparent. Simulation studies written in commonly used and widely supported programming languages are more readily reproducible and accessible than studies conducted by clicking buttons in a graphical user interface. Ideally, code should be made available in a formal package or library structure (e.g. `ss3sim`: Anderson *et al.*, 2013). This makes it easier for future users to understand the code and simulation structure and encourages coders to document their work. Further transparency can be introduced by developing code using a version control system such as Git, which provides a history of all code modifications (e.g. https://github.com/seananderson/ss3sim). Finally, reproducibility is enhanced by controlling model runs through plain-text control files

Simulation frameworks are most useful when they provide flexibility to the designers as well as allowing for future users to develop scenarios and test questions that the designers may not have envisaged. Simulation frameworks are most flexible when code is developed using small discrete functions that can be mixed and utilized for a wide range of implementations. Keeping input and output files in formats that can be read and processed by multiple tools (e.g., comma-separated, or tab-delimited text files) can also add flexibility. Finally, simulation scenarios are best split into case-specific components. For example, an operating scenario might be comprised of one combination of cases that describe trajectories of changes in natural and fishing mortalities and selectivity patterns. If these cases are specified in individual text control files then they can be flexibly combined to create different scenarios, without duplicating case specification.

Check models early and often

Software validation is critical, particularly when the operating model is complicated. We have found deterministic model checking to be very useful. This involves apply the estimation model to data generated deterministically (no [or very little] process or observation error) when the estimation and operating model have the same structure. An important component to model checking is graphical analysis of model outputs. Many complex problems such as XX are unlikely to be detected without graphical model checking [GIVE EXAMPLE][@gelman2002]. To facilitate model checking, such graphics should be rapid and easy to reproduce. We found the visualization packages `manipulate` (RStudio, 2011) and `ggplot2` (Wickham, 2009), for the statistical software R (R Core Team, 2013), to be helpful for this purpose.

Keep simulation runtime minimal

In order to discover problems with simulations and understand how they are performing, model iterations should be run repeatedly under a variety of conditions. In addition to using a fast computer and writing code carefully, runtime can be minimized 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.

TODO other points: write code so the simulations can be distributed across cores, computers, and researchers. Use a remote repository service such as GitHub (https://github.com) to ease collaboration among researchers and sharing with others.

Applying the paradigm

In our studies we used Stock Synthesis (SS), a free and widely-used integrated assessment modelling framework (Methot and Wetzel, 2013). Johnson et al. (this issue) evaluated the ability of SS to estimate key quantities of management interest for three life histories (cod- sardine- and flatfish-like) and three fishing mortality scenarios when natural mortality was age-invariant, but time-varying. Assessment configurations evaluated includes those in which (age-invariant) natural mortality pre-specified or estimated. Ono et al. (this issue) analyzed the ability of SS to estimate management metrics for the three life-history types given a range of assumptions regarding data quality and quantity. They also considered whether the frequency and duration of length- and age-composition data collection, or catch history, affected the bias or precision of estimates of management quantities. Hurtado Ferro et al. (this issue) investigated factors which lead to retrospective patterns in fisheries stock assessment models. Specifically, they tested how key biological and modeling factors can induce retrospective patterns for various life history types. They explored the potential effects of catch patterns, as well as model miss-specification due to time-varying biological parameters, time-varying selectivity and catchability, and their interactions. They assessed the utility of including time-varying selectivity in the assessment as a means to correct them in those cases where retrospective patterns were observed,.

If a simulation study should be designed with the three ‘Rs’ in mind, then its success can be measured using the same criteria. Our studies were relevant, realistic, and reproducible. Combining a series of studies in one large project that covered multiple life history types, and a series of plausible biological and fishing scenarios, enabled us to determine some broadly-applicable principles of contemporary stock assessment models.

Lessons learned

If a simulation study should be designed with the three ‘Rs’ in mind, then its success can me measured by the same criteria. Our studies were relevant: EXPLAIN; realistic: EXPLAIN; and reproducible EXPLAIN. Combining a series of studies in one large project that covered multiple life history types, and a series of plausible biological and fishing scenarios, enabled us to determine some broadly-applicable principles of contemporary stock assessment models.

TODO: Insert key lessons from each of the studies, and summarise the ones that were applicable across all studies

Discussion and conclusions

What we showed and what was covered. Where it fits into current research.

Simulation modeling studies should be most useful in answering the following types of questions (general discussion of the strengths and weaknesses of this approach, i.e., when is it most suitable).

Use of pre-packaged stock assessment software can be problematic when that software contains legacy methods are not appropriate for a particular purpose. Frequent use of legacy methods is common when there is widespread use of packaged fisheries stock assessment software (Martell and Ianelli, 2012). The application of ready-made models fails to question the original assumptions of the conceptual modeller or software programmer, and in effect, serves to reinforce their original ideas and impressions of fisheries and fish population dynamics (Longhurst, 2010). It is thus important that the information and data available to modellers is routinely scrutinised, and that common methods and models are tested and evaluated as data are updated. Simulation studies are key to addressing such problems.

Benefits or otherwise of using the same modeling system to create the operating and assessment model.

Difference between simulation modeling for research purposes, and modeling to test, compare, and scrutinise alternative methods. Both are useful but should be designed differently and have different measures of 'success'.

Simulation studies can be very useful in data limited situations, for example, in understanding recreational fisheries.

Recommended next steps: Following studies and key findings from the conference for example

Acknowledgements

...

References

Anderson, S. C., Cunningham, C., Hurtado-Ferro, F., Johnson, K. F., Licandeo, R., McGilliard, C., and Monnahan, C. *et al.* 2013. ss3sim: An R package for generalized stock-assessment simulation with Stock Synthesis. https://github.com/seananderson/ss3sim.

Hilborn, R., and Walters, C. J. 1987. A General Model for Simulation of Stock and Fleet Dynamics in Spatially Heterogeneous Fisheries. Canadian Journal of Fisheries and Aquatic Sciences, 44: 1366–1369.

Lee, H.-H., Maunder, M. N., Piner, K. R., and Methot, R. D. 2011. Estimating natural mortality within a fisheries stock assessment model: An evaluation using simulation analysis based on twelve stock assessments. Fisheries Research, 109: 89–94.

Longhurst, A. R. 2010. Mismanagement of marine fisheries. Cambridge University Press, Cambridge.

Piner, K. R., Lee, H.-H., Maunder, M. N., and Methot, R. D. 2011. A simulation-based method to determine model misspecification: examples using natural mortality and population dynamics models. Marine and Coastal Fisheries, 3: 336–343.

R Core Team. 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

RStudio. 2011. manipulate: Interactive Plots for RStudio.

Wickham, H. 2009. ggplot2: elegant graphics for data analysis. Springer New York. http://had.co.nz/ggplot2/book.