A while ago on an island far, far away . . .

TESA Simulation Modelling workshop short presentation

Daniel Ricard

February 2 2021

Background

Southern Bluefin Tuna

- The Commonwealth Scientific and Industrial Research Organisation (CSIRO) provides science advice to the Australian Fisheries Management Authority (AFMA) for international negotiations at the Commission for the Conservation of Bluefin Tuna (CCSBT)

SESAME

I was involved in this project when I worked at the CSIRO in 2001-2003.

My colleague Ann Preece was very good at formulating acronyms.

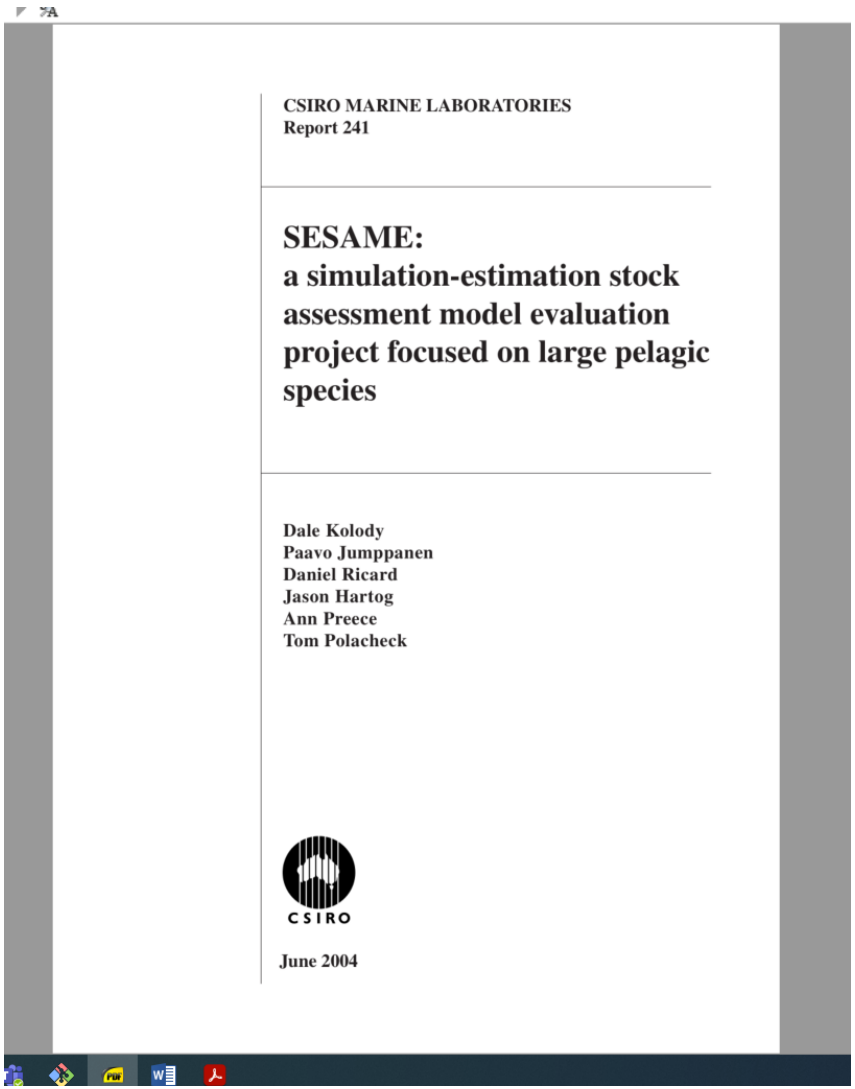
> *SESAME: a simulation-estimation stock assessment model evaluation project focused on large pelagic species*

A large report documented the project.

This work laid strong foundations for the Management Strategies Evaluation (MSE) that subsequently took place for Southern Bluefin Tuna.

That MSE fizzled after the discovery of important unreported catches in the fishery, see Polacheck (2012).

SESAME



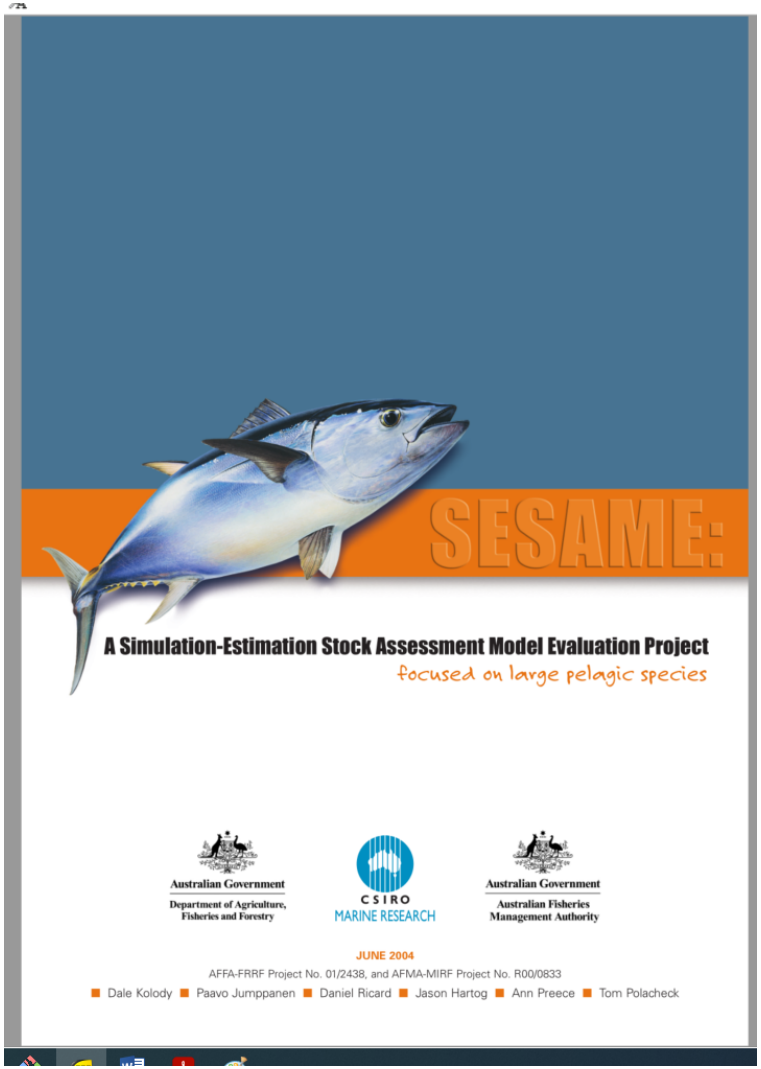
CSIRO MARINE LABORATORIES
Report 241

SESAME:
a simulation-estimation stock assessment model evaluation project focused on large pelagic species

Dale Kolody
Paavo Jumppanen
Daniel Ricard
Jason Hartog
Ann Preece
Tom Polacheck

CSIRO

June 2004



SESAME

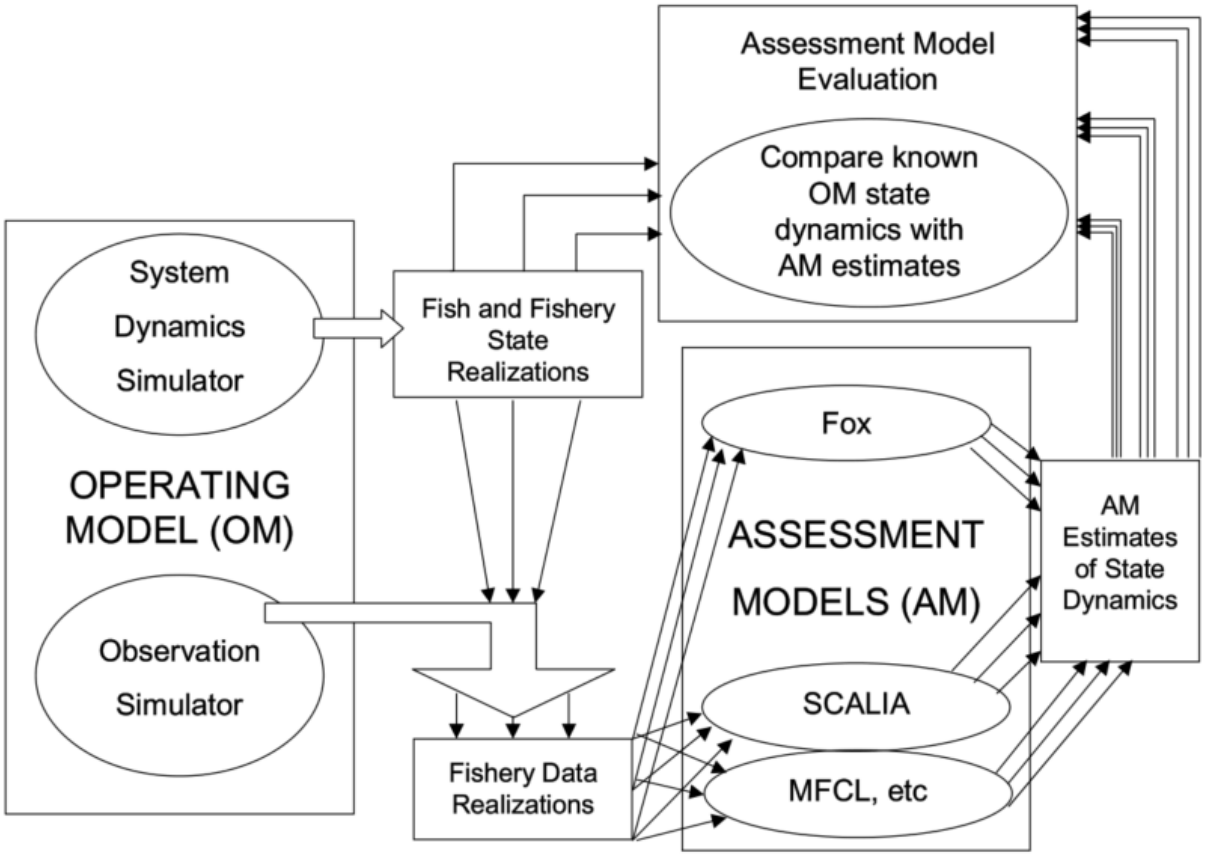


Fig. 1. **Outline of simulation-estimation methodology for stock assessment model evaluation.**

Do's - things that worked

- have a strong and diverse team
- have clearly defined set of operating model scenarios
- have a clearly defined set of assessment models to apply to the simulated data
- have clearly defined set of performance measures
- have an agreed-upon format and a shared source for data inputs to models
- have an agreed-upon format and a shared source for the outputs of assessment models
- use shell scripts and R as the Swiss Army knives and Duct tape

4.5 DATABASE OVERVIEW

We found a relational database to be an extremely powerful tool for organizing the operating model state realization summary statistics and assessment model estimates. Every assessment model fitting to each data realization produced a uniquely identifiable file of results that were uploaded to an ODBC-compliant database. In addition to stock assessment model estimates, these files potentially contained flags that were useful for diagnosing function minimization problems (e.g. final gradients of the objective function with respect to parameters).

8 performance indicators for 16 assessment models from one OM run

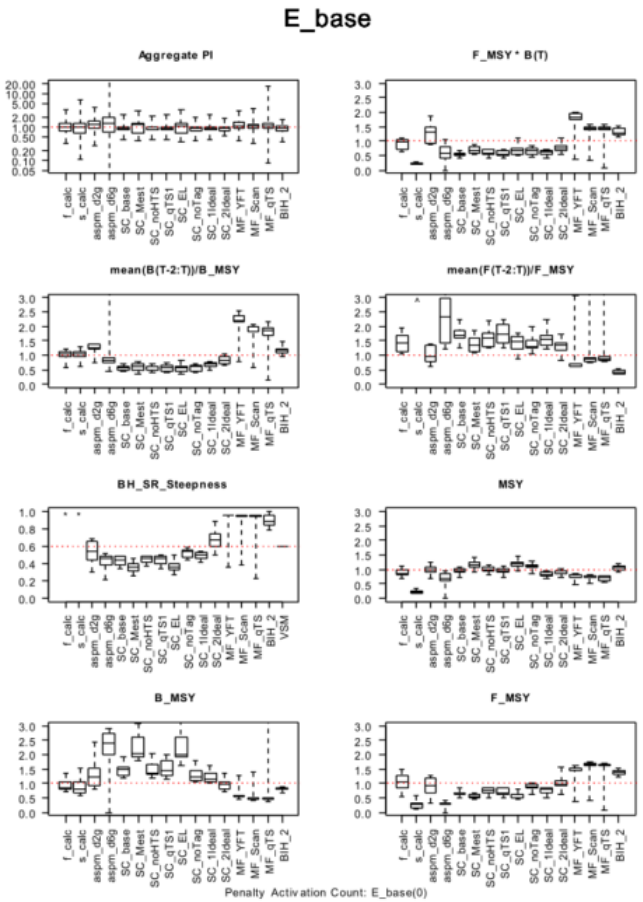


Fig. 3a. Boxplots of MPD performance indicators resulting from the application of a range of assessment models to 10 data realizations simulated from the E_base SBT operating model. Individual values represent the ratios of (AM estimated)/(OM actual) in all cases except BH_SR_Steepness which shows the actual values (the Aggregate PI is an arithmetic mean of 28 ratios). OMs are defined in Table 1, AMs in Table 2 and PIs in Table 8.

Dont's - things that could be improved

- Use Word to write the report, this is a completely unnecessary hassle
- Rely on an adhoc operating model, not because VSM was not a good operating model, just because it has been done before by now

  VSM (which is an abbreviation of virtual stock model) is a stand-alone Windows based command line application which simulates multi-species fisheries. - Appendix I SESAME Report

- use (enforce) source control software, nowadays this would be git
- plan analyses and program software so that all analytical steps fall like dominoes

4.6 QUALITY CONTROL

We endeavoured to test that operating and assessment models were implemented and documented correctly, but inevitably, coding errors occur in complicated software. The complexity of the simulations also led to mis-specifications and interpretation errors among participants, that sometimes were not recognized until well into the project when the technical documentation was being completed. When the errors were large, all affected results were re-run; relatively minor errors are documented and perhaps apparent only as peculiar model specifications. While we cannot be sure that all the errors were identified, the multitude of comparisons among independently coded population dynamics models gives us a reasonable degree of confidence that the gross features of most models behaved roughly as intended in the majority of cases.

Thank you

when talking about your well-designed methods, always replace “easy” or “simple” with “elegant”