Crosssvalidation Meeting, Tokyo 22-27 January 2018

Laurence Kell , Toshihide Kitakado , Rishi Sharma , Ai Kimoto

January 29, 2018

Abstract

The meeting in Tokyo from the 22nd to the 27th of January covered aspects of OM design that were related to test whether the assessment model used to set the OM has sufficient predictive power or not. We set up an objective cross-validation technique to demonstrate how this could be used for Eastern Atlantic Bluefin Tuna (EABFT) Assessment base case scenarios for the 2017 Assessment. In addition, we developed an outline for a series of papers that would come out from this work with the following; i) dealing primarily with a current review of practice used in tRFMO's for OM Model setup, and the limitations with the current approaches used, ii) developing objective cross-validation approaches to assess model performance/predictive power for a EABFT current assessment, and iii) a simulation study demonstrating best practices,iv) optimizing some of the grid designs to test Management Procedures, and v) showcasing a case study with Atlantic Albacore. Some of these papers will be submitted for the special issues on MSE's in Canadian Journal of Fisheries and Aquatic Sciences in April of 2018, and other manuscripts will be completed by September, 2018.

1   Introduction

The tuna Regional Fisheries Management Organisations (tRFMOs) are increasingly using Management Strategy Evaluation (MSE) to develop robust management advice. This requires the conditioning of Operating Models (OMs) that simulate resource dynamics and monitoring data in order to test the strategies used to set management regulations. This is done by simulating candidate Management Procedures (MPs) as feedback controllers and then choosing the MP that best meets management objectives. Where an MP is the combination of pre-defined data, together with an algorithm to which such data are input to provide a value for a control measure such as a Total Allowable Catch (TAC). Once the OMs and management objectives have been chosen then the "best" MP is an emergent property, the task is to find it. Therefore the procedures for proposing, validating, accepting and weighting OM scenarios is fundamental to the MSE process.

There are many alternative ways to condition OMs which are used to represent the resource dynamics. One way is to use a stock assessment, this implies that the assessment is able to describe nature almost perfectly, however, if this is true why bother with MSE? Therefore, in this project we review current practice adopted when conditioning OMs in the tRFMOs, we then look at objective ways to chose and validate OM scenarios (validation procedures for EBFT and a simulation study on different life history types, a special case being tuna-like), optimizing grid design through partial factorial designs and multivariate techniques, and provide a worked example based on North Atlantic albacore.

The meeting covered various aspects of OM Design,primarily dealing with model setup, model validation and model predictability. The primary topics covered are shown by subheading below.

2   Review of current practice in tRFMO's

The range of OM's examined across tRFMOs were primarily based on the assessment models. In some cases these were developed for peculiarities of the species (IO SKJ/AO BFT), and may have explicit spatial structure. Issues of eliminating unrealistic scenarios need to be standardized, and should be clearly documented so one tRFMO can learn from another. Getting agreement on the scenarios examined should be discussed from the onset. In addition, robustness trials should be agreed on from the onset. Grid based OM's dealing with structural uncertainty has primarily been the basis of most development work so far, though more processes dealing with sampling

and time series approaches that account for non-stationarity of ecological processes should be examined in most cases. Data weighting and issues of which models are more plausible is an area for further work, as some management bodies (IWC) maybe a lot further along than others. Spatial issues in OM and multi-stock structures are additional areas where further work can be undertaken, and observation error models accounting for sampling biases are important to consider. The current approach using the assessment models as the basis for OM design is a good starting point, though further processes (observation error and time series processes) should be added as additional processes that should be accounted for in OM designs.

3 Grid Design in Operating Models: Work Smarter, Not Harder

MSE's often use assessment models with complex structural uncertainty designs examining multi-way interactions. The paper examines different approaches to evaluate multi-way interactions and optimize grid design to capture the uncertainty in the system being modeled. The objective is not to make a perfect assessment model, but rather to capture the true dynamics of the system that encapsulates the underlying uncertainty so a reference set along with a robustness set of models could be examined to optimize a control rule that could be empirical or model based in nature.

4 Cross-validation I:An Objective Procedure for the Validation of Operating Models Conditioned on Stock Assessment Datasets

In fish stock assessment model validation is often based on a naïve adaptation of Pearson residuals, i.e. the difference between observations and posterior means, even if this approach is flawed (Thygesen, 2017). A reason for this is because statistics based on residuals from model fits are not always a good guide to how well that model will predict, since a high R^2 or low root mean square error (RMSE) can be obtained by over-fitting. For example, in a simple polynomial regression better fits to the data can be obtained by adding higher order terms but the predictions from the model on new data will usually get worse as higher order terms are added. This problem is compounded by often having to compare scenarios for alternative datasets and model structures with different data requirements, and so AIC cannot be used to compare models. In addition there are also a range of potential problems to identify when examining residuals, e.g. bias, drift, skewness, missing variables, and heteroscedasticity. When inspecting residual patterns, however, there is a danger of hypothesis fishing (Wasserstein and Lazar, 2016) and so it is good practice to reserve part of the data for validation. This ensures that the significance of a pattern in the data is not tested on the same data set which suggested the pattern.

Cross validation evaluates the predictive error of a model by testing it on a set of data not used in fitting. There is often insufficient data, however, in stock assessment datasets to allow some of it to be kept back for testing. A more sophisticated way to create test datasets is, like the jackkife, to leave out one (or more) observation at a time. Cross validation then allows prediction residuals to be calculated, i.e. the difference between fitted and predicted values where the later is calculated from the out-of-sample predictions.

Prediction residuals can either be for historical or future observations. In the later case for example a one-step forward prediction is where data points are made available to the model one measurement at a time, and the model is evaluated by its ability to predict the next data point. This is the general principle of frequentist statistics (Dawid, 1984).

In this study we show how prediction residuals can be used to validate stock assessment scenarios, using as an example East Atlantic and Mediterranean bluefin. Model validation examines if the model family should be modified or extended, and is complementary to model selection and hypothesis testing. Model selection searches for the most suitable model within a family, whilst hypothesis testing examines if the model structure can be reduced.

The provision of fisheries management advice requires fitting a model to data to assess stock status, the prediction of the response of the stock to management, and checking that predictions are consistent with reality. The accuracy and precision of the prediction depend on the quality the model, the information in the data and the prediction horizon (i.e. how far ahead we wish to

predict). Often small tweaks to the input data or assumptions can result in substantial differences to advice. The aim of this study therefore is to develop a procedure for the validation of stock assessment model scenarios comprising alternative model structures and datasets. Where validation examines if a model family should be modified or extended, and is complementary to model selection and hypothesis testing. Model selection searches for the most suitable model within a family, whilst hypothesis testing examines if the model structure can be reduced. We do this for a case study based on the 2017 East Atlantic and Mediterranean bluefin stock assessment data (ICCAT, 2018) with scenarios for for a range of datasets, model structures and the values of fixed parameters. Prediction errors are then calculated using leave-one-out, one-step and hindcast procedures and compared to model errors.

The aim of this study is to develop an objective procedure based on cross-validation for comparing Operating Model and Stock Assessment scenarios, comprising alternative datasets and model structures.

5 Cross-validation II:Simulation to assess predictive power in Assessment Models

The paper aims at evaluating performance by crossvalidation approaches (e.g. JK and HC) for selecting stock assessment models with different data and complexity through simulation. The rationale is that these approaches can be used for situations where standard model selection criteria cannot be applied. Age-structured models conditioned on real data were used for generating simulation data to fit several stock assessment models such as surplus production models, delay-difference models, statistical catch-at-age models (and stock synthesis models). The models were then compared by Jack-knifed or hindcasted metrics against observed abundance indices, mean weight of catch etc. Simulation results showed that the Xval approaches work well for enhancing contrast in fitting and predictability among stock assessment models. Results can be used for conditioning and screening of operating models used in the management strategy evaluation.

6 North Atlantic albacore

A case study using guidelines specific to the observation error model will be presented including best practice principles learned from the items 1-4 above.

7 Timeline for the project

Five manuscripts will be developed out of this project that will be beneficial to practitioners in the assessment and MSE arena. They are the following:

1) The Review will be completed by March, 2018 and submitted to the special issue of Canadian Journal (CJFAS) by April 1st, 2018.

2) The Eastern Bluefin case study will be completed and tested on the two base case assessments; the VPA and the integrated Stock synthesis model. Procedures were developed at the Tokyo meeting and run for some analysis. The results will be compiled, code archived by March, 2018 and manuscripts completed by mid-June, 2018 for the tuna RFMOs MSE meeting.

3) The simulation study cross-validation procedure will be developed between March 2018 and June 2018, with a manuscript prepared and submitted by October, 2018.

4) The OM optimal grid design paper will be tested by May, 2018 with a manuscript submitted by September, 2018.

5) The North Atlantic Albacore case study will be developed and submitted for a manuscript by April, 2018 for the special issue of Canadian Journal on MSE.

8 Online Documentation

The overall project is documented here with links to the other projects, code and other resources (https://github.com/laurieKell/xval/wiki).

9 Possible future meetings

This meeting was very productive as it was in-person, if funds are available, additional meeting could be held in September, 2018, to complete all the components of this project. It will be also beneficial to have the members attend across tuna-RFMOs MSE June meeting to present some of the results.

10 Acknowledgements

We would like to thank ABNJ for making provisions to have an in-person meeting to brainstorm and develop ideas for this project between the 22nd and 27th of January, 2018.