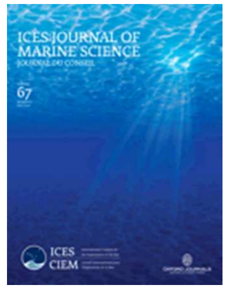

Model averaging to stream-line the stock assessment process

Journal:	*ICES Journal of Marine Science*
Manuscript ID:	ICESJMS-2013-350
Manuscript Types:	Symposium Article
Date Submitted by the Author:	02-Sep-2013
Complete List of Authors:	Millar, Colin; European Commission, Joint Research Center, IPSC, Maritime Affairs Unit Jardim, Ernesto; European Commission, Joint Research Center, IPSC, Maritime Affairs Unit Scott, Finlay; European Commission, Joint Research Center, IPSC, Maritime Affairs Unit Osio, Giacomo; European Commission, Joint Research Center, IPSC, Maritime Affairs Unit Mosqueira, Iago; European Commission, Joint Research Center, IPSC, Maritime Affairs Unit Alzorriz, Nekane; European Commission, Joint Research Center, IPSC, Maritime Affairs Unit
Keyword:	Fisheries management, plausible scenarios, model averaging, stock assessment, model selection, structural uncertainty

Model averaging to stream-line the stock assessment process.

(Food for thought article)

Colin P. Millar*, Ernesto Jardim, Finlay Scott, Giacomo Chato-Osio, Iago Mosqueira, Nekane Alzorriz.

European Commission Joint Research Center. Institute for the Protection and Security of the Citizen.

Via Enrico Fermi 2749, 21027 Ispra (VA), Italy

*Corresponding author: colin.millar@jrc.ec.europa.eu +39 0332 785208

Abstract

The current fish stock assessment process in Europe can be very resource and time intensive. The scientists involved require a very particular set of skills, acquired over their career, drawing from biology, ecology, statistics, mathematical modelling, oceanography, fisheries policy and computing. There is a a particular focus on producing a single 'best' stock assessment model, but as fisheries science advances there are clear needs to address a range of hypotheses and uncertainties, from large scale issues such as climate change to specific ones, such as high observation error on young hake. Key to our discussion is the use of the assessment for all framework to translate hypotheses into models. We propose a change to the current stock assessment procedure, driven by model averaging over a range of plausible

hypotheses, where increased collaboration between the varied disciplines within fisheries science will result in more robust advice.

Introduction

Stock assessment can be defined as the application of quantitative and statistical models to estimate the current and historical status and trends of a fish stock, including the abundance, mortality and productivity (Hilborn and Walters, 1992). The typical stock assessment process involves a team of scientists generating a suite a candidate models, examining their outputs, and then selecting a single model that is considered to be the 'best'. The selection process may be based on a mixture of criteria, for example, the model likelihood, the examination of residuals etc. The outcomes of this assessment process are often point estimates of the stock status, with minimal consideration of the uncertainty of those estimates.

This process can be very resource intensive. A sufficient number of scientists with an appropriately high level of training are required to generate the suite of candidate models, interpret their output and review the conclusions. Additionally, as each model is examined in detail, the process can be very time consuming. The magnitude of the resource problem is likely to increase as demand for advice and data availability increases (Jardim *et al.*, 2013).

A key step in the stock assessment process is that of moving from the initial suite of candidate models to a single 'best' model. This effectively means that the assumptions behind the alternative models are rejected. But these alternative models could provide perfectly valid advice. As Box and Draper (1987) stated “Remember that all models are wrong, the practical question is how wrong do they have to be to not be useful”. Even though only one model may be the most likely, the others may still represent plausible 'states of nature' and contribute to the estimation of our uncertainty about it. Many other authors have raised and discussed issues surrounding basing conclusions on a single model (Hilborn, 1997; Butterworth *et al*. 1996; Hill *et al.*, 2007; Simmonds *et al.,* 2011) the major danger is that by ignoring model selection error leads to too narrow confidence intervals, and generally biased inference (Claeskens and Hjort, 2008). Model averaging provides a way to incorporate this error into a single result by combining the results from several models.

We propose the use of model averaging (Claeskens and Hjort, 2008) within the a4a framework (Millar *et al.*, 2013) to address the two stock assessment issues identified above: 1) the need to choose only a single model and 2) intensive resource use.

There are several benefits of using a model averaging approach. It removes the need to select a single model which in turn reduces the need to conduct extensive diagnostic model checks. Consequently, the introduction of model averaging allows for a change in stock assessment practices where more time is spent on defining an initial suite of plausible models for each stock. This suite of models must be carefully chosen to represent possible “states of nature”. It may even be possible to define a sufficiently exhaustive set of initial models that would be appropriate to use for a group of stocks, for example, all demersal species in the North Sea.

In this paper we propose a change to the stock assessment process. From a process where model checking and model selection are the focus to one where dreaming up appropriate models is the most important scientific task. We believe this change is not a revolution, merely an evolution of what is already done. We envisage a process where the full gamut of fisheries disciplines: oceanography, genetics, biology, population biology, gear technology, to name a few, can have a direct input into the design of stock assessment models. The way in which these diverse groups communicate with the stock assessment group would be to propose issues that they would like to see addressed. These issues will help form a suite of plausible models which, through model averaging, will be the basis of the stock assessment. Of course, nothing in life is easy and so we discuss practical issues in the final section. First, we describe in more detail the idea of setting up plausible states of nature and techniques for model averaging.

The a4a stock assessment framework

The a4a framework is designed to be a flexible stock assessment modelling tool with an intuitive interface. This is a key part of our model averaging approach, as the success of the idea depends on fisheries scientists being able to easily and efficiently translate ideas into mathematical models of the stock dynamics in the assessment model.

The a4a framework builds a full assessment model from several sub-models: one for fishing mortality, one for survey catchability, one for recruitment and models for the observation variances, allowing the user to set up a wide range of population dynamics and fisheries models.

Most stock assessment models rely on a few basic assumptions: full spatial coverage of the stock, non-fisheries induced mortality (M), a constant stock recruitment relationship, constant survey catchability. These assumptions, depending on the model, can be relaxed or adapted, but in many cases it is not straightforward to incorporate flexibility with respect to these assumptions; largely because most stock assessment models have not been designed as exploratory tools, but rather to provide a single assessment. Modelling frameworks, like a4a (Millar *et al.*, 2013) or SS3 (Methot and Wetzel, 2012), promote the process of exploring different models, opening the possibility of dealing with several distinct models, instead of tweaking small details of a single model.

Plausible states of nature

When dealing with building a model for stock assessment, one is often faced with the fact that more than one hypotheses can be valid, e.g. a regime shift in the North Sea in 1988 changed the temperature which may affected the mortality rate of salmon (*Salmo salar*) via reduced growth (Beaugrand and Reid 2012). Was there a regime shift effect or not? Does the regime shift affect mortality or reproductive success? Other examples of situations where fisheries science could inform a list of plausible scenarios are:

- Changes in phytoplankton and zooplankton coincided with an increase in catches of the western stock of the horse mackerel (*Trachurus trachurus* L.) in the northern North Sea after a northerly expansion from the Bay of Biscay after 1987 (Reid *et al*. 2001).
- Food availability and climate effects can directly affect recruitment of fish stocks as shown for the North Sea cod (*Gadus morhua*) depressed recruitment since the mid 1980s (Beaugrand et al. 2003).
- Cases characterised by spatially variable selectivity tend to confuse standard approaches of interpreting fishing mortality rates in concert with selectivity (Crone et al. 2013). Assuming that differences in spatial and temporal availability of the fish can alter the expected shape of the gear-selectivity curve, there is the need for flexible selectivity patterns. Crone et al. (2013) suggest to include both the contact selectivity and availability in stock assessments to model the combined factors that affect fish vulnerability: log Q = log contact selectivity + log availability

- With reverence to the previous bullet, several authors document shifting stocks, suggesting that migratory and distributional changes have occurred (Nye *et al.*, 2009; Last *et al.*, 2011; Jansen *et al.*, 2012, Poos *et al.*, 2013), and could potentially affect availability to surveys.
- Increasing efficiency of survey vessels is a known issue. For example, one of the two multinational surveys for the North Sea cod was removed from the annual stock assessment due to the suspected time varying catchability (ICES, 2011). Temporal trends in survey or commercial catchability can bias the estimates of stock size and fishing mortality in stock assessment models that do not account for them (Wilber *et al.,* 2010).
- Developing operating models for southern bluefin tuna (Hillary, pers comm.), where uncertainty in assumptions of life history parameters, such as natural mortality and carrying capacity, is incorporated by setting up a grid of plausible input parameter values.

Each of the situations listed above highlights a recognised need by the fisheries/scientific community to address multiple hypotheses about the “states of nature”. Some effects may act across a number of stocks: temperature effects, variations in primary productivity; while others may be more stock specific: gear mesh size increases affecting smaller or thin fish. The next step in the process is combining these issues into a set of plausible states of nature which will essentially define the stock assessment models that should be applied.

In our opinion these cases fall into two categories. One dealing with anything that could be considered to be on a continuum, which would be dealt with by specifying a range of plausible values. For example, it is possible to try a range of breakpoints in a stock recruitment relationship, a range of ages above which fishing mortality is constant, a range of M values, a range of growth scenarios, etc.

The alternative is a situation where different hypotheses form categories rather than a range of options. For example in (Wilber *et al.*, 2010) there are a variety of models proposed for survey catchability with time: step changes, smooth changes, changing catchability with covariates such as spatial patchiness of effort, or spatial extent of fishing effort.

Model averaging

Model averaging is a technique to incorporate model selection uncertainty into inference (Buckland *et al.*, 1996). From the stock assessment perspective, this can be thought of as the incorporation of uncertainty due to different plausible states of nature, as mentioned above. The purpose of this section was not to be prescriptive in the model averaging schemes to consider, but rather to highlight that there are a variety of approaches, frequentist, and Bayesian, simple and complex.

Model averaging can be thought of as a model weighting algorithm where the weights are based on the support for the model in the data (Claeskens and Hjort, 2008). We discuss four model averaging techniques, three chosen as they are relatively easy to implement, and a fourth chosen for its desirable features. The first two methods use deterministic (fixed) weights, while the third and fourth methods are based on the idea that model weights are random and use stochastic simulation to build the distribution of the weights. Here follows a short description of each

1. The first method is a semi-Bayesian approach in which the weights are taken from an estimate of the posterior model probability called "the harmonic mean estimator" (HME). This estimator, introduced by Newton and Raferty (1994) requires samples from the posterior distribution of the parameters of each model. Having estimated the posterior model probability, the samples can then be drawn from each model conditional on the deterministic posterior probability derived from the HME.

2. The second method is a frequentist version of the above, where the deterministic weights are based on the AIC of each model. The distribution comes from assuming that the estimates of the model parameters have a multivariate normal distribution with mean given by the mode of the likelihood, and covariance matrix given by the inverse of the Hessian at the mode.

3. The third method is frequentist and is a true model average estimator, in that the weights are random. These weights are known as smooth AIC weights (Buckland *et al.*, 1997). It requires that the empirical distribution of the data can be simulated. Given a sample from this distribution the model with the lowest AIC is chosen and the maximum likelihood estimates of the parameters are stored. This process is repeated until sufficient sample have been drawn. To simulate from the empirical distribution of the data, the original paper suggests using a bootstrap technique, and so can be thought of as bootstrapping model selection and fitting in one step.

4. The fourth and final method we consider is Bayesian and is known as reversible jump Markov chain Monte Carlo (RJMCMC). This method is a true model average estimator, and this time the random weights are samples from the posterior distribution of the

models. We refer the reader to King et al. (2010) for further details.

These approaches can be developed in relative isolation with respect to the development of plausible models. An efficient division of labour would be to engage statisticians and programmers to develop a scheme for model averaging, this would only need to be done once. Freeing up stock assessment scientists, population biologists, fisheries biologists, oceanographers and others to collaborate on proposing a set of plausible models; this task would be periodically reviewed.

Discussion and Future challenges

Using model averaging allows fisheries scientists to concentrate on fisheries science. The paper suggests a process by which fisheries scientists who are not stock assessment experts in the traditional sense can directly contribute to the stock assessment process. By defining a stock assessment as a set of models, and since model averaging techniques allow a large number of models to be explored, specialised knowledge can be more readily incorporated in the form of models or covariates, because there is less pressure on deciding if this covariate or that model is too big a departure from what came before. In this sense, model averaging may even help to stabalise stock assessment results.

As stated previously, the key to the success of the ideas in this paper, is a flexible and intuitive interface. In a4a (Millar *et al.* 2013, Jardim *et al.*, 2013), this achieved through the use of sub-models for fishing mortality, survey catchability, recruitment and observation variance, specified as linear models using splines. We think this simple at heart stock assessment model is a valuable tool for translating plausible ideas / states of nature into plausible stock assessment models, and could be used to cover many of the examples mentioned in the text.

In terms of implementation, levels of plausible models could be constructed to give some common structure across stocks. This type of coherence is often sought by having an assessment group contain all flat fish, or all demersal fish in an eco-region. From the plausible model perspective, one could set up eco-region level scenarios which are applied to all stocks, followed by further levels: demersal and pelagic, say. In the same spirit as Bentley (2013), this would provide a starting point for any new stock that is assessed, and should improve coherence and consistency, and perhaps even the transparency of the assessment process.

Model averaging avoids the pitfalls of using a single model: too narrow confidence (credible) intervals; over optimistic tests of significance; and generally biased results (Claeskens and Hjort, 2008). Dealing with uncertainty in general (model and parameter) in fisheries advice is developing, but more discussion needs to be had about the translation of (model and parameter) uncertainty in advice through to the implementation of policy. We do not address these issues, instead we refer to Hill at al. (2007) who give a thorough discussion of model uncertainty (as well as a good historical perspective) in terms of ecosystem management, their examples and recommendations apply equally well to single species stock assessment in our context

Bimodal distributions may result from the use of model averaging in the face of competing hypotheses. In these situations, taking the posterior model-averaged mean is clearly not a good summary description. A better description would be the marginal density or at least the highest posterior density interval. It is the management procedures that require a point estimate that stand out as inadequate in such situations, rather than the idea of model averaging. Different models may make different predictions for good reasons. Is it better to select one scenario based on human reasoning rather than average a plausible range of scenarios based on data?

Model averaging is not easy. Simpler methods tend to be approximate, doing it correctly is difficult. The AIC approach is straightforward, but relies on a normal approximation to the distribution of model parameter estimates, which could be seriously inadequate. The HME method uses samples from the parameter (posterior) distribution requiring the use of an MCMC algorithm (automatic methods are available for this). However, the HME is notorious for having infinite variance and has been thoroughly discouraged by some leading statisticians (Neal, 2008). The Smooth AIC method is a true model averaging procedure, but unfortunately is only possible if one can sample from the empirical distribution of the data. In stock assessment it very difficult to take bootstrap samples (note it is not advisable to sample from the residuals because the use of a model in the procedure could bias the model selection). RJMCMC is the only usable, non-approximate, method of the four discussed, but requires some skill in setting up the various proposal distributions required.

When Bayesian model averaging is being used priors on models need to be specified. There are various approaches to this, and perhaps a sensible approach is to set priors on groups of models where all the models in each group will be related the same issue. In an advisory body, guidelines for practical issues like this could be developed alongside the model averaging procedure itself.

An evolution of the stock assessment process.

We are not suggesting a radical change to the way fisheries are managed. We are not suggesting a move away from single species assessments, nor are we suggesting a change to the advice. We are lobbying for a change in the typical stock assessment procedure. The most exciting foreseen benefit is that this approach aims to bring improved coherence within an advisory body. There are many ICES working groups that attract the best scientists in their field (oceanography, benthic ecology, gear technologists, experts on the history of surveys). If we can develop ways to accumulate ideas from from a range of sources into set of stock assessment models, then we have the basis of a very cross discipline, scientifically defensible, and powerful stock assessment.

References

Beaugrand, G., Brander, K.M., Lindley, J.A., Souissi, S., Reid, P.C. Plankton effect on cod recruitment in the North Sea (2003) Nature, 426 (6967), pp. 661-664.

Beaugrand, G. and Reid, P. C. 2012. Relationships between North Atlantic salmon, plankton, and hydroclimatic change in the Northeast Atlantic – ICES Journal of Marine Science, 69: 1549–1562.

Box, G. E. P., and Draper, N. R., (1987), Empirical Model Building and Response Surfaces, John Wiley & Sons, New York, NY.

Breiman L. (2001) Statistical modeling: the two cultures. Statistical Science 16(3), 199-231.

Buckland, S. T., Burnham, K. P., & Augustin, N. H. (1997). Model selection: an integral part of inference. Biometrics, 603-618.

Butterworth, D. S., Punt, A. E., & Smith, A. D. M. (1996). On plausible hypotheses and their weighting, with implications for selection between variants of the Revised Management Procedure. REPORT-INTERNATIONAL WHALING COMMISSION, 46, 637-642.

Claeskens, G., & Hjort, N. L. (2008). Model selection and model averaging(Vol. 330). Cambridge: Cambridge University Press.

Crone, P. R., M. N. Maunder, J. L. Valero, J. D. McDaniel, and B. X. Semmens (Editors). Selectivity: theory, estimation, and application in fishery stock assessment models. Workshop Series Report 1. Center for the Advancement of Population Assessment Methodology (CAPAM). NOAA/IATTC/SIO, 8901 La Jolla Shores Dr., La Jolla, CA 92037. 46 p.

Hilborn, R., & Walters, C. J. (1992). Quantitative fisheries stock assessment: choice, dynamics and uncertainty. Reviews in Fish Biology and Fisheries, 2(2), 177-178.

Hilborn, R. (1997). Uncertainty, risk and the precautionary principle. In American Fisheries Society Symposium (Vol. 20, pp. 100-106).

Hill, S. L., Watters, G. M., Punt, A. E., McAllister, M. K., Quéré, C. L., & Turner, J. (2007). Model uncertainty in the ecosystem approach to fisheries.Fish and Fisheries, 8(4), 315-336.

ICES. 2011. Report of the Workshop on the Analysis of the Benchmark of Cod in Subarea IV (North Sea), Division VIId (Eastern Channel) and Division IIIa (Skagerrak) (WKCOD 2011), 7 –9 February 2011, Copenhagen, Denmark. ICES Document CM 2011/ACOM: 51. 94 pp.

Jansen T, Kristensen K, Payne M, Edwards M, Schrum C, et al. (2012) Long-Term Retrospective Analysis of Mackerel Spawning in the North Sea: A New Time Series and Modeling Approach to CPR Data. PLoS ONE 7(6): e38758. doi:10.1371/journal.pone.0038758

Jardim, E., Millar, C.P., Scott, F., … . Can stock assessment be as simple as a linear model. ICES Journal of Marine Science, this issue (submitted).

King, R., Morgan, B., Gimenez, O., & Brooks, S. (2010). Bayesian analysis for population ecology. CRC Press.

Last, P. R., White, W. T., Gledhill, D. C., Hobday, A. J., Brown, R., Edgar, G. J., and Pecl, G.

281 (2011). Long-term shifts in abundance and distribution of a temperate fish fauna: a response to climate change and fishing practices. Global Ecology and Biogeography, 20: 58– 72.

283 Methot Jr, R. D., & Wetzel, C. R. (2012). Stock synthesis: A biological and statistical framework for fish stock assessment and fishery management. Fisheries research.

285 Millar C.P, Jardim, E.J., Osio, G.C., Mosquiera, I. (2013). Assessment for all (a4a): a flexible and robust stock assessment framework. PLOSOne (in revision).

287 Neal, R. (2008, August 17). The Harmonic Mean of the Likelihood: Worst Monte Carlo Method Ever. Retrieved from http://radfordneal.wordpress.com/2008/08/17/the-harmonic-mean-of-the-likelihood-worst-monte-carlo-method-ever/

290 Newton, M. A., & Raftery, A. E. (1994). Approximate Bayesian inference with the weighted likelihood bootstrap. Journal of the Royal Statistical Society. Series B (Methodological), 3-48.

292 Nye, J. A., Link, J. S., Hare, J. A., and Overholtz, W. (2009). Changing spatial distribution of fish stocks in relation to climate and population size on the Northeast United States continental shelf. Marine Ecology Progress Series, 393: 111– 129.

295 Olsen, Esben Moland and Ottersen, Geir and Llope, Marcos and Chan, Kung-Sik and Beaugrand, Grégory and Stenseth, Nils Chr. (2010). Spawning stock and recruitment in North Sea cod shaped by food and climate. Proceedings of the Royal Society B: Biological Sciences

298 Patterson, K. R. (1999). Evaluating uncertainty in harvest control law catches using Bayesian Markov chain Monte Carlo virtual population analysis with adaptive rejection sampling and including structural uncertainty. Canadian Journal of Fisheries and Aquatic Sciences, 56(2), 208-221.

Philip C. Reid, Maria de Fatima Borges, Einar Svendsen, (2001). A regime shift in the North Sea circa 1988 linked to changes in the North Sea horse mackerel fishery, Fisheries Research, Volume 50, Issues 1–2, February 2001, Pages 163-171, ISSN 0165-7836,

Poos, J. J., Aarts, G., Vandemaele, S., Willems, W., Bolle, L. J., & van Helmond, A. T. M. (2013). Estimating spatial and temporal variability of juvenile North Sea plaice from opportunistic data. Journal of Sea Research, 75, 118-128.

Simmonds, E. J., Campbell, A., Skagen, D., Roel, B. A., & Kelly, C. (2011). Development of a stock–recruit model for simulating stock dynamics for uncertain situations: the example of Northeast Atlantic mackerel (Scomber scombrus). ICES Journal of Marine Science: Journal du Conseil, 68(5), 848-859.

Wilberg M J., J T. Thorson, B C. Linton, and J Berkson (2010). Incorporating Time-Varying Catchability into Population Dynamic Stock Assessment Models Reviews in Fisheries Science, 18(1):7–24